EFFECTS OF INBREEDING ON HUMAN QUANTITATIVE TRAITS AND COMPLEX COMMON DISEASES OF LATE-ONSET

Igor Rudan

Ph.D. Thesis
The University of Edinburgh
2005

To Harry and Rosa, my Scottish friends

CONTENTS

ABSTRACT viii

FOREWORD ix

DECLARATIONS xi

NOTES xii

GLOSSARY xiii

1. INTRODUCTION 1

1.1. CURRENT APPROACHES TO STUDYING GENETIC BASIS OF HUMAN TRAITS AND DISEASES 2

1.1.1. Introduction 2
1.1.2. Advances in Genome Research Enhanced Development of Genetic Epidemiology 6
1.1.3. Polymorphic Markers and Linkage Disequilibrium Enabled Search for Genes Responsible for Human Diseases 10
1.1.4. Study Designs of Genetic Epidemiology: Extended Pedigree Studies, Affected Sib-Pair Studies and Genetic Association Studies 14
1.1.5. Recent Understanding of Genetic Determinants of a Spectrum of Human Diseases 20
1.1.5.1. *Monogenic (Mendelian) Diseases* 20
1.1.5.2. *Oligogenic Diseases* 21
1.1.5.3. *Complex Polygenic Diseases of Late Onset* 23
1.1.6. Current Understanding of Genetic Architecture of Common Complex Diseases 28
1.1.7. Leading Current Approaches to Identify Common Complex Disease Genes 34
1.1.8. Translating Genomic Knowledge into Public Health Benefits 37

1.2. STUDYING THE EFFECTS OF INBREEDING ON HUMAN TRAITS AND DISEASES 41

1.2.1. Introduction 41
1.2.2. Global Prevalence and Distribution of Inbreeding 44
1.2.3. Effects of Inbreeding on Pre-Reproductive Traits and Diseases 53
1.2.3.1. *Early Mortality* 53
1.2.3.2. *Early Morbidity* 53
1.2.3.3. *Cognitive Performance* 55
1.2.3.4. *Other Effects* 55
1.2.4. Effects of Inbreeding on Post-Reproductive Traits and Diseases 57
1.2.4.1. *Quantitative Traits* 57
1.2.4.2. *Complex Late-Onset Diseases* 60
1.2.5. Inbreeding Studies can Improve Understanding of Genetic Architecture of Late-Onset Complex Traits: Theoretical Considerations 63

2. EXAMINEES AND METHODS 67

2.1. STUDY POPULATION: THE CROATIAN ISLAND ISOLATE RESOURCE 68

2.2. INVESTIGATION OF THE EFFECTS OF INBREEDING ON BLOOD PRESSURE 74

2.2.1. Sample Selection 74
2.2.2. Computation of Individual Inbreeding Coefficients in 25 Villages 74
2.2.3. Blood Pressure Measurements 75
2.2.4. Statistical Analyses 79

2.3. INVESTIGATION OF THE EFFECTS OF INBREEDING ON CORTICAL INDEX AND OSTEOPOROSIS 81

2.3.1. Sample Selection 81
2.3.2. Computation of Individual Inbreeding Coefficients in 14 Villages 83
2.3.3. Cortical Index Measurements and the Definition of Osteoporosis 83
2.3.4. Statistical Analyses 87

2.4. INVESTIGATION OF THE EFFECTS OF INBREEDING ON LEARNING DISABILITY 89

2.4.1. Sample Selection 89
2.4.2. Estimation of Inbreeding Coefficients in 10 Villages 89

2.4.3. Estimation of Learning Disability Prevalence 90
2.4.4. Statistical Analyses 90

2.5. INVESTIGATION OF THE EFFECTS OF INBREEDING ON 10 COMMON COMPLEX DISEASES OF LATE ONSET (CORONARY HEART DISEASE, STROKE, CANCER, SCHIZOPHRENIA, UNI/BIPOLAR DEPRESSION, ASTHMA, TYPE II DIABETES, GOUT, PEPTIC ULCER, EPILEPSY) 92

2.5.1. Sample Selection 92
2.5.2. Computation of Individual Inbreeding Coefficients in 14 Villages 93
2.5.3. Defining Disease Status and Estimating Prevalence 96
2.5.4. Statistical Analyses 96

2.6. INVESTIGATION OF THE EFFECTS OF HETEROSIS ON MATING SUCCESS, FITNESS, HEIGHT, BLOOD PRESSURE, CHOLESTEROL AND TRIGLYCERIDE MEASUREMENTS 98

2.6.1. Sample Selection 98
2.6.2. Defining Inbred, Endogamous, Admixed and Outbred Individuals in the Study Sample 102
2.6.3. Data Collection and Trait Measurements 103
2.6.4. Statistical Analyses 104

2.7. ETHICAL ISSUES 106

3. RESULTS 107

3.1. EFFECTS OF INBREEDING ON BLOOD PRESSURE 108

3.1.1. Ecologic Study at the Village (Population) Level 108
3.1.2. Regression Analysis at the Individual Level 108

3.2. EFFECTS OF INBREEDING ON CORTICAL INDEX AND OSTEOPOROSIS 113

3.3. EFFECTS OF INBREEDING ON LEARNING DISABILITY 117

3.4. EFFECTS OF INBREEDING ON 10 COMMON COMPLEX DISEASES OF LATE ONSET (CORONARY HEART DISEASE, STROKE, CANCER, SCHIZOPHRENIA, UNI/BIPOLAR DEPRESSION, ASTHMA, TYPE II

DIABETES, GOUT, PEPTIC ULCER, EPILEPSY) 121

3.4.1. Ecologic Study at the Village (population) level 121
3.4.2. Cohort Study at the Individual level 123

3.5. EFFECTS OF INBREEDING, ENDOGAMY, ADMIXTURE AND OUTBREEDING ON FITNESS, MATING SUCCESS AND QUANTITATIVE TRAITS (HEIGHT, BLOOD PRESSURE, CHOLESTEROL AND TRIGLYCERIDES) 126

4. DISCUSSION 130

4.1. AIMS AND SCOPE 131

4.2. PROPOSED THEORETICAL MECHANISMS OF INBREEDING EFFECTS ON HUMAN LATE-ONSET TRAITS AND COMMON DISEASES 134

4.2.1. The Deleterious Effects of Inbreeding Depression on Quantitative (Endo)Phenotypes that Confer Risk for Late-Onset Diseases May Be Multiplicative 134
4.2.2. Effects of Inbreeding on Rare Variants with Large Effect in Post-Reproductive Adults ("Invisible Mendelian Diseases" of Late Onset) 135
4.2.3. Autozygosity in Many Rare Recessive Variants of Small Effect Could Result in Epistatic Effects that Could Jointly Impair the Capacity to Compensate Against Environmental Risks 138
4.2.4. Heterozygote Advantage in Loci Under Balancing Selection is Expected to be Reduced by Inbreeding 139
4.2.5. Empirical Evidence of Inbreeding Effects in Humans and Other Organisms 140

4.3. SUMMARY OF THE INVESTIGATED EFFECTS OF INBREEDING ON HUMAN QUANTITATIVE TRAITS AND LEARNING DISABILITY 142

4.3.1. Blood Pressure 142
4.3.2. Cortical Index 147
4.3.3. Learning Disability 149

4.4. SUMMARY OF THE INVESTIGATED EFFECTS OF INBREEDING ON HUMAN LATE-ONSET COMPLEX DISEASES 150

4.5. SUMMARY OF THE INVESTIGATED EFFECTS OF ADMIXTURE AND OUTBREEDING ON HUMAN HEALTH AND DISEASE 156

4.6. CRITICAL ASSESSMENT OF THE RELIABILITY OF THE FINDINGS 160

4.7. WIDER IMPLICATIONS OF CONDUCTED RESEARCH 165

4.6.1. Inbreeding and Global Burden of Disease 165
4.6.2. The Outbreeding Theory 166

4.8. CONCLUSION 169

ACKNOWLEDGEMENTS 172

5. REFERENCES 173

APPENDIX 1
(Contains ethical approvals from relevant Ethics Committees in Scotland and Croatia, information booklet presented to the population to obtain informed consent, and the consent form for participation in the research project).

APPENDIX 2
(Contains a full version of questionnaire used in the field study).

APPENDIX 3
(Contains copies of 6 research papers published in international journals based on the results of this PhD Thesis).

THE UNIVERSITY OF EDINBURGH

ABSTRACT OF THESIS

Regulation 3.1.14 of the Postgraduate Assessment Regulations for Research Degrees refers These regulations are available via:- http://www.aaps.ed.ac.uk/regulations/exam.htm

Name of Candidate: IGOR RUDAN

Address : DEPARTMENT OF PUBLIC HEALTH SCIENCES
TEVIOT PLACE, EDINBURGH
SCOTLAND, UK

Postal Code: EH8 9AG

Degree: Ph.D.

Title of Thesis: EFFECTS OF INBREEDING ON HUMAN QUANTITATIVE TRAITS AND COMMON COMPLEX DISEASES OF LATE ONSET

No. of words in the main text of Thesis: 45,000

Studying the effects of inbreeding in human populations could provide insights into the genetic architecture of medically relevant quantitative traits and common complex diseases of late onset. In a historic example of 2,761 examinees from isolated village populations of the islands of Brac, Hvar and Korcula, Croatia, collected through field work undertaken in the 1970's and 1980's, individual inbreeding coefficients were computed based on genealogical records. Inbreeding showed a strong positive effect on blood pressure and negative on cortical index. The 14 villages were revisited in 2000 to assess the prevalence of learning disability and of common complex diseases of late onset. A cohort study and an ecological study, after appropriate standardization, both showed that inbreeding increased the prevalence of coronary heart disease, stroke, psychiatric disorders, cancer, gout, asthma, glaucoma and peptic ulcer, but not type II diabetes. A strong effect on the prevalence of learning disability was also noted in 10 villages. In a follow-up study on 1,001 examinees from 10 other villages sampled on neighbouring islands in 2002, positive effects of outbreeding on fitness, height, blood pressure, cholesterol and triglyceride values were detected. The possible explanations for the observed effects include: (i) The joint effect of inbreeding depression on all polygenic quantitative phenotypes that confer risk for late-onset diseases is predicted to be multiplicative rather than additive. (ii) The "genetic load" of rare "Mendelian" variants with large deleterious effects in post-reproductive adults is unknown, but could be much greater than expected as these variants were invisible to selection through human history. (iii) Deleterious effects resulting from autozygosity in hundreds of affected rare recessive variants of small effect under common disease/rare variant (CD/RV) hypothesis could result in epistatic effects that could jointly impair the capacity to compensate against environmental risks. (iv) Heterozygote advantage in loci under balancing selection could be reduced by inbreeding. Consanguinity is common in many populations and the possible effects of inbreeding depression on disease burden and reduced life expectancy should be further investigated.

FOREWORD

This thesis is a result of my involvement in a large research programme, which includes several groups of researchers in both Scotland in Croatia. Although I acknowledged their continuing help and support (see Acknowledgement on page 163), in this foreword I explain in detail which elements of this PhD Thesis were entirely my own work, and which resulted from the work of the entire group involved in the joint research programme.

The idea for the thesis, to investigate the effects of inbreeding on late-onset quantitative traits and complex common diseases in Croatian islands isolates, is my own. I began to develop it in 1997 and I published the first concepts, objectives, strategies and results as a single author in the years of 1999 and 2001, based on the datasets that I collected myself, originating from the Croatian Cancer Registry and from the Island of Lastovo, Croatia.

Since May 2000, I closely collaborated with Professor Harry Campbell from the University of Edinburgh, who is my supervisor on this thesis and who substantially helped me to further develop those ideas over the course of time. In this thesis, we agreed to test our hypotheses on the effects of inbreeding on late-onset traits and diseases in two datasets:

(i) A historic, 20-year old sample from 25 villages of 3 different Croatian islands (Brac, Hvar and Korcula) where inbreeding is prevalent; the data on blood pressure, cortical index and socio-economic variables were collected by the team of the Institute of Anthropological Research in Zagreb, Croatia, in late 1970s and early 1980s. That field research was undertaken under the leadership of Professor Pavao Rudan, who is the

Director of the Institute. The two variables of interest were measured by Dr Nina Smolej-Narancic, who kindly provided them for our further analyses.

This historic dataset was appended with individual inbreeding coefficients computed from the genealogical records, which I reconstructed myself between 1997 and 2000, and with data on complex chronic diseases in the individuals from the historic sample who were still alive in 2000, which I extracted myself in collaboration with the local general practitioners.

(ii) A newly collected dataset derived from contemporary populations of 10 villages from 4 different Croatian islands (Rab, Vis, Lastovo and Mljet), where inbreeding is also prevalent; Professor Campbell and I jointly raised funding for the field work leading to the development of this second database from the Wellcome Trust, the Croatian Ministry of Science and Technology, the British Council, the Royal Society UK, the Medical Research Council UK and the National Institutes of Health (USA). I led the field work to collect the new dataset in 2002 and 2003.

The analysis of the collected data is my own work, supervised by my mentor Professor Harry Campbell. The only exception is the statistical analyses of association between individual inbreeding coefficients and blood pressure and its implications on genetic architecture of the trait. That analysis was performed by Dr Andrew Carothers from the Human Genetics Unit of the Medical Research Council in Edinburgh, to increase the quality of the presentation of the results when a related paper was submitted for publication, and I have repeated those analyses under his supervision when writing this thesis.

Professor Alan Wright has substantially assisted me in interpreting the data and realizing their significance in wider context, and I used the elements from our discussions to write the two closing sections of the discussion of this thesis.

DECLARATIONS

I hereby declare that:

- I composed the thesis entirely on my own;
- The work presented in the thesis is my own, and for the elements of the thesis that were a result my work within a group of scientists, I declare that I have made a substantial contribution to the work, and I clearly indicated this contribution in Foreword.
- This work has not been submitted for any other degree or professional qualification except as specified.

Edinburgh, June 6th, 2005.

Igor Rudan

NOTES

- Some of the material included in a thesis has been published before the submission of the thesis. This was done with the approval of the supervisor.

- In all cases where this was appropriate, the publisher's formal permissions and the permissions of all the joint authors were obtained for the photocopies of published papers attached in **Appendix 3**.

GLOSSARY

Admixture	The mixture of two or more genetically distinct populations
Autozygosity	Homozygosity in which both alleles at a locus in an individual are a copy of a single allele inherited from an ancestor who was related to both parents of the individual
Candidate region	Linked region of the genome (see "Linkage analyses") that is more likely to contain a causal genetic variant influencing a studied phenotype of interest
Complex trait (disease)	A measured phenotype, such as disease status or a quantitative trait, which is influenced by many environmental and genetic factors, and potentially by interactions in and between them
Cryptic relatedness	Genetic relatedness between individuals due to joint sharing of the gene(s) inherited from a common ancestor from the more distant past
Detectance	Probability of observing a presumed genotype given the phenotype
Effect size	The extent to which a factor influences the risk of the condition under study, rather than simply an indication of whether the factor is significantly related to the condition
Endogamy	At the level of individuals or a population, a measure describing how likely it was that the ancestors of the individuals seen presently to inhabit an area were also born and chose a spouse from the same area
Fitness	Probability of reproducing (passing the genes to the next generation)

Genetic association studies	A genetic variant is genotyped in a population for which phenotypic information is available; if a correlation is observed between genotype and phenotype, there is said to be an association between the variant and the disease or trait
Genetic drift	Changes in allele frequencies in a population from one generation to another as the result of chance events in mating, meiosis and number of offspring
Haplotype	A set of alleles that is present on a single chromosome
Heritability	The proportion of variation in a given characteristic or state that can be attributed to (additive) genetic factors
Human isolates	Also termed "founder populations"; populations that have been derived from a limited pool of individuals within the last 100 or fewer generations.
Identity by descent	See "Autozygosity"
Inbreeding coefficient	The probability that both copies of an allele are inherited from a common ancestor (identical-by-descent)
Inbreeding depression	The detrimental effects of inbreeding, typically causing a reduction in the means of fitness-related traits as a result of increased homozygosity
Linkage analysis	Mapping genes by typing genetic markers in families to identify chromosome regions that are associated with disease or trait values within pedigrees more often than expected by chance.
Linkage disequilibrium	The non-random association of alleles of different linked polymorphisms in a population
Mendelian (monogenic) diseases	Diseases inherited in simple (Mendelian) fashion, as the entire disease phenotype is due to a mutation in a single gene

Mutation	Any change in genome sequence
Penetrance	Probability of observing a presumed phenotype given the genotype
Polymorphism	A variant allele with a frequency greater than 0.01
Quantitative trait loci	Genetic loci that contribute to variations in quantitative (continuous) phenotypes
Statistical power	The probability of correctly rejecting the null hypothesis when it is truly false

1. INTRODUCTION

1.1. CURRENT APPROACHES TO STUDYING THE GENETIC BASIS OF HUMAN TRAITS AND DISEASES

1.1.1. Introduction

Human health is defined by the World Health Organisation as "...a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity." (**1**). Critics of this definition argue that health is a process rather than a state, and that well-being cannot ever be expected to be "complete" (**2**). This shows how difficult it is to define either health or disease and to clearly separate them. Human diseases represent a complex spectrum of health disturbance, ranging from those presenting early in life to those with very late onset and from mainly genetically determined to nearly entirely caused by environmental exposures (**3**).

It is certain, however, that the aetiology of some diseases is considerably simpler than of the others. Some diseases, such as Mendelian diseases, often present early in life, thus limiting the importance of environmental exposures in their aetiology in comparison to diseases of later onset (although the role of intrauterine exposures to different environmental stressors and its effects on health in later life are not yet well understood). Such diseases are usually due to the structural or functional deficiency of a single protein, resulting from random change in thegenetic code controlling the synthesis of this protein (**4**). The partial or a complete structural deficit of a single enzyme, or a building protein, eventually leads to cascade of events presenting as a disease phenotype. In this case, in all affected patients the disease phenotype is a direct consequence of a single change in genetic information. At the other end of the complexity spectrum are diseases such as cardiovascular disease and cancer. They develop for years as a result of the combination of inherited polygenic susceptibility and

lifetime exposure to the environment (**3**). Their slow development can only be monitored as a gradual breakdown of physiological mechanisms that became unable to compensate for microscopic structural damage or functional changes at various levels of the human organism. The result is a continuing deviation of monitored metabolic or biochemical parameters from the expected value. Over the course of time, this results in the first presenting symptoms and eventually leads to the development of the characteristic disease phenotype (**3**).

The investigation of genetic and environmental factors that interact in causing a disease has been a fundamental goal of the modern epidemiology of non-communicable diseases. This science recognises that each human being has a different genetic build-up (except monozygotic twins) and a personal history of lifetime environmental exposures, and therefore represents an unique evolutionary experiment in time and space. However, by designing appropriate prospective and retrospective studies, modern epidemiology attempts to identify genetic characteristics and environmental exposures that are significantly more frequent among the persons with the disease phenotype than in the general population. Such characteristics, found among those who exhibit the disease phenotype more frequently than it would be expected by chance, are then considered "disease risk factors". Once the risk factors are identified, an evaluation of the relative contribution of each of them to the development of disease and assessing whether the associations found between diseases and risk factors are causal or coincidental represent other important goals of modern epidemiology (**5-8**).

To produce results of any validity, epidemiological studies heavily rely on an investigator's ability to measure suspected risk factors (genetic characteristics or environmental exposures) precisely and to classify study subjects correctly according to the presence or absence of the disease. The presence of disease is typically "measured" by a set of diagnostic criteria. These criteria have a validity of nearly 100% for many diseases (especially if based on pathohistological examinations or various biochemical tests). Still, for some common diseases, misclassification is a serious concern even

nowadays (**9,10**). Regarding the risk factors, researchers usually investigated those that could be measured with a reasonable degree of precision. This included a multitude of variables related to the individual's environmental exposures, physiological or metabolic monitoring, or even psychological profiling. The most investigated risk factors in non-communicable disease epidemiology during the 20th century therefore included a person's age, marital status, occupational exposures, smoking, body mass index, blood pressure, cholesterol levels, food intake frequency, psychological assessment of life quality, etc. These all fostered the development of environmental, ecological, occupational, nutritional and psychological epidemiology (**5**).

Genetic epidemiology, however, was unable to expand in a similar fashion. This was because during most of the 20th century it was hardly possible to measure an individual's genetic build-up in any direct way. The investigation of genetic risk factors was limited to twin studies (**11**), heritability studies (**12**), inbreeding studies (**13**) and studies of phenotypically expressed genetic variation (e.g. classic erythrocyte antigens, HLA systems) (**14,15**). These studies had a variety of designs and could only provide indirect insights into general patterns of association between factors related to inheritance and the occurrence of diseases. This was very far from designs that researchers ideally desired to apply. However, this situation dramatically changed during the past two decades, when methods for the analysis of the variation in the DNA molecule in each human individual were first introduced (**16,17**). This development immediately attracted epidemiologists of non-communicable diseases, who realised the opportunities provided by directly measuring genetic material as a disease risk factor (**6**). The early excitement and optimism were justified, as today it is possible to measure an individual's genetic build-up more precisely than we will ever be able to measure a person's environmental exposures (such as diet). The progress in genomic research, especially in understanding the patterns of variation in the human genome, is destined to nurture a further rapid expansion of genetic (genomic) epidemiology, possibly to the levels that eventually may surpass all historic successes of other approaches to epidemiological research in non-communicable diseases (**18-21**).

In this introduction, current strategies to find genes responsible for susceptibility to human diseases will be discussed, especially those associated with the greatest burden of disability and death, such as cardiovascular diseases, cancer, diabetes or psychiatric disorders. The ongoing debates on the likely genetic architecture of most common human diseases will be addressed. The suitability of the available tools to find underlying genes to the existing knowledge gathered from both genetic and epidemiological research of human disease aetiology will be analysed. Finally, based on the results of various applied strategies to find common disease susceptibility genes, current prospects for translating genomic advances into measurable public health benefits will be discussed.

1.1.2. Advances in Genome Research Enhanced the Development of Genetic Epidemiology

The main goal of genetic epidemiology is to express the genetic build-up of an individual as a (predictor) variable, so that it can be statistically correlated to other variables of interest – disease phenotype and interacting environmental risk factors. The criterion variable in this design can be qualitative (the presence or absence of disease phenotype) or quantitative (such as levels of blood pressure, serum cholesterol or blood glucose). The advances in the study of the human genome during 1990s through the publicly funded Human Genome Project (*HGP*) and the privately funded project at *Celera Genomics* made important steps towards achieving this goal (**22,23**). The sequencing of the human genome showed that all humans are identical in about 99.9% of their genome sequence. More importantly, however, the genes (which are, in broad terms, the segments of the genome sequence that contain code for protein synthesis) formed only about 3% of that sequence, while the function of the remaining 97% of the genome remained largely unknown.

After those discoveries, it initially seemed logical to assume that the entire variation in terms of susceptibility to acquiring or avoiding diseases during a life span would have to be due entirely to the variation found within those 3% of the coding genome sequence, i.e. the genes, as only more recently comparative genomics between different mammals began to point to the presence of ultraconserved elements in the non-coding genomic regions, implying that they too may be of relevance. Efforts by *HGP* and *Celera* both initially estimated that there are about 40,000 genes scattered across the human genome, a number later corrected to about 25,000 (**22,23**). The term "gene", however, has a number of possible definitions. For the purpose of genetic epidemiological studies, a "gene" can be considered a location in a human DNA molecule, usually several thousand nucleotides (i.e. base pairs) long, in which there are shorter segments of coding sequence ("exons") and longer segments of non-coding

sequence ("introns"). During protein synthesis, the coding elements ("exons") are first transcribed into "messenger RNA" (mRNA) molecules, while the non-coding elements ("introns") are cut out. Messenger RNA then travels from the cell nucleus to the ribosomes where it is being read in "triplets" of nucleotides, each encoding the subsequent amino-acid to be built into the structure of a resulting protein. It is important to note that the location in the DNA that we refer to as a "gene" jointly assumes the coding information inherited from the person's mother (one DNA strain) and father (the complementary DNA strain). If those two sequences ("gene alleles") inherited from both of the parents are identical, then a person will be "homozygous" for that gene, and if they differ the individual is "heterozygous". If both differing sequences ("alleles") of a gene in a heterozygous individual are eventually expressed in a phenotype, then they act "codominantly". If only one of them is phenotypically fully expressed and the other is not, then they are "dominant" and "recessive" alleles, respectively (**6,22,23**).

The defined terms "homozygosity" and "heterozygosity" refer to characteristic of each gene in an *individual*. Genetic epidemiology is also interested in terms that relate to characteristics of genes in the entire *population*. A gene is "monomorphic" if there is only one ("fixed") allele for this gene present in the entire population. This means that all the individuals in the population will have to be homozygous for that gene, as they will always inherit the same allele, whoever their parents are. It is thought that about 65% of all human genes are monomorphic. Those probably include many genes with regulatory signalling function that determine that we only should have two kidneys, one liver, and five fingers on each hand. Other genes are "polymorphic", as two or more sequence variants ("alleles"), each with population frequency of at least 1%, can be found in the population at their precise genomic location. The differences between humans at their birth are thought to be mainly due to genetic variants found at polymorphic genomic loci (**22,23**).

I argued that one of the important goals of genetic epidemiology was to precisely measure the genetic build-up of each individual and to quantitatively express it as a

variable. This means that genetic epidemiologists are interested only in the varying part of human genetic material. If one accepts that 97% of the non-coding human genome sequence has no apparent function, then this part of the genome should not influence disease susceptibility. In addition to that, some two thirds of the predicted 25,000 human genes could be monomorphic ("human genetic invariance"), and their role in disease susceptibility may therefore not influence the variation in disease risk among humans. This intuitively suggests that only the differences in combinations of alleles that can be found at remaining ~10,000 polymorphic human genes determine genetic variation in susceptibility to human disease.

If this is true, it would reduce the aim of genetic epidemiology to identifying different sequence variants ("alleles") in polymorphic genes that are significantly more frequently present among the diseased than in healthy population. Subsequent goals would then include understanding how the change in sequence ("mutation"), which introduced this new allele, affects protein synthesis, and by what mechanisms does this lead to disease phenotype. Therefore, finding alleles associated with increased disease risk could, in the long term, allow genetic testing of individuals at a very early age and permit the institution of preventive measures from childhood to decrease environmental disease risks. Another apparent benefit of finding these alleles is that insights into change in function of the protein they encode could enhance our understanding of complex mechanisms that cause the disease and provide new targets for development of related drugs.

In an ideal genetic epidemiological longitudinal study, all ~10,000 polymorphic human genes would be sequenced in the entire human population. All the existing sequence variants ("alleles") would be catalogued and their population frequencies determined. Then, all humans would be followed up for disease status during their entire lifetime. After correcting for known environmental risks, gene-environment interactions and gene-gene interactions, it would theoretically be possible to assign relative risks and the population-attributable fraction of disease incidence in the population to each allelic

variant that exists for each polymorphic gene. Such an undertaking would currently be impossible, as large funds and years of work by many research groups would have to be spent at the current state of technology to perform this study even in 10 individuals (**22,23**). However, the costs of genotyping (i.e. determining existing alleles at different genome locations in DNA from individual subjects) are rapidly decreasing and the speed at which it can be performed is increasing. This led to recent calls to set the target at an affordable price for whole genome sequencing, i.e. setting a long-term goal to reduce the price of sequencing each individual's genome to about $1,000 (**24**). This is still far from realistic, but hardly anyone in the field would argue that it might not become possible in the future. Genetic screening and individually tailored drug treatment and prevention strategies could therefore become available to everyone who could afford the sequencing of his/her own genome.

Until then, we will have to design alternative approaches to locate genes underlying human diseases and find allelic variants responsible for functional changes that lead to disturbed health. The main problem is correctly associating disease phenotype with the specific allelic variant of a gene located somewhere in a genome. The advances in human genomics offered the first tools. These are based on two fundamental properties of the human genome. The first is the existence of several abundant classes of polymorphic sites scattered across the entire genome with precisely determined locations ("polymorphisms", "genetic markers"), that can be genotyped using the polymerase chain reaction (**25-28**). The second is a property of the genome called "linkage disequilibrium", which reflects the observation that markers that are physically close to each other will tend to be co-inherited, and so will the entire genomic sequence between them (**29-32**).

1.1.3. Polymorphic Markers and Linkage Disequilibrium Enabled Search for Genes Responsible for Human Diseases

The sequencing of the human genome produced evidence of the general identity of genomic sequence among humans in about 99.9% of base pairs when any two humans are compared. However, this is not to say that all 6 billion people have nearly identical sequences, and that the entire observed variation is due to the remaining 0.1% of the genetic code. Virtually any of the 3 billion nucleotides in the haploid genome can be randomly changed by mutation at any generation in any individual (**33**). However, since every person has 2 alleles for every gene (apart from those on Y-chromosome), the frequencies of such changes at the time of their first appearance will be 1 in 12 billion alleles, i.e. negligible. We speak of "polymorphic sites" in the human genome only when such newly introduced mutations gradually increase in population frequency over a number of generations, and eventually the frequency of the second allele of a previously monomorphic gene reaches frequency of at least 1% in the entire population (**33**). Loci with two or several alleles present in the population, each in frequency greater than 1%, have been found at about 0.1% of human with a genome sequence. These polymorphic sites can be exactly located by applying specifically tailored restriction enzymes. There are several classes of those "polymorphic sites" in the genome ("genetic markers", "polymorphisms"), but only two will be addressed here as they have an important potential to assist finding genes responsible for common human diseases.

The first class of polymorphisms is jointly termed "short sequence repeats" (SSR). At some locations in the genome, short nucleotide sequences tend to repeat several times (e.g. "TATATATA" represents a dinucleotide "TA" which is repeated four times; "GTCGTCGTC" is a trinucleotide "GTC" repeated three times). For some reason, when DNA is duplicated during the cell division, the number of repeating times of these nucleotide sequences can be slightly mistaken (e.g. 6 or 8 repeats are transcribed into the DNA of a separating cell instead of 7). The locations in the genome where variable

number of short tandem repeats can be found are called "STR"-s (from "short tandem repeats") or "microsatellites". The exact location of hundreds of those "markers" has been defined in the human genome, which roughly translates into one per each 1 million nucleotides (*bp*). They are very informative, as not only are they variable in a population at each individual location, but their sequence along the chromosomes (termed "haplotype") is becoming more unique for each person with introduction of each following microsatellite into a haplotype. With each new polymorphic STR added, the probability of any two persons sharing exactly the same haplotype rapidly diminishes, until it is so small that it allows the unique identification of each person based on variation found in the DNA (**25,26,34**).

Another large and promising class of polymorphisms are "single nucleotide polymorphisms" (SNP). These have been one of the major discoveries of the Human Genome Project. It is thought that there should be several million of these polymorphisms scattered throughout the human genome (**33**). As each of 3 billion nucleotides in the human genome can mutate (e.g. from "A" to "C", "T" or "G"), this gives rise to a new "allele" of extremely low frequency in the population (which equals exactly one over the doubled number of all humans). However, not many of these mutations will increase in frequency over many generations to reach an allele frequency of 1% needed for that precise location in the genome to be formally declared a "single nucleotide polymorphism" (SNP). It should be noted that each individual SNP marker is incapable of subdividing humans into as many categories as each individual STR marker. This is because each SNP locus can theoretically only have four different "alleles" ("A", "C", "T" or "G"), one of which is usually highly predominant. At the same time, each STR marker can have numerous alleles (e.g. 10 alleles with 5 to 14 repeats), many of them with quite similar allele frequencies. This makes a haplotype of any 5 consecutive STR markers far more informative, while any 5 consecutive SNP markers may be shared by a large number of humans.

However, there are also large advantages of SNPs over STRs. Firstly, they are several orders of magnitude more numerous in the human genome, and one can be found roughly in every 1 thousand nucleotides, while one STR can approximately be found per 1 million nucleotides. Thus, SNPs should allow for a very fine-scale search for genes on sequence segments that have been suspected as candidate for harbouring disease genes. This is often referred to as "fine-mapping" and it uses the property of the genome termed "linkage disequilibrium", which we will attempt to explain in more detail (**35,36**). Also, STR markers are very informative in families, where they are directly transmitted from ancestors to the offspring, but their use is very limited at the level of population because they mutate too frequently and therefore obscure the signal with the disease variants at the level of population, which is highly unlikely to happen within the several generations of extended pedigrees under study. Finally, a very practical advantage of the SNPs over STRs is that SNPs can be tested in very large numbers due to the technological development of DNA chips, while STR markers are much more difficult to work with and genotype in the laboratory.

The human genome is diploid in nature, which means that the large majority of its 3 billion nucleotide sequences exist in duplicate, with one copy inherited from each parent. These two copies coexist and functionally represent a single human genome. The only two exceptions to this rule are the sequence of the Y-chromosome (haploid in nature and always inherited from the father) and mitochondrial DNA (always inherited from the mother). The rest of the human genome is diploid, and it has the property of recombination ("crossing over"): exchanging the corresponding complementary segments of the genome sequence between parents' haploid genome sequences in each person. This happens during meiosis, when the recombined 23 chromosome haploid genome does not contain exclusively a single parent chromosomes, but rather a relatively random mixture of maternal and paternal sequences within each chromosome. This has the obvious consequence of breaking down the haplotypes of specific STR or SNP markers that were originally found along the paternal chromosomes. However, polymorphic alleles on a haplotype that are physically close to each other will have a

smaller probability of separation through recombination from those that are located far apart. Therefore, some alleles will have a probability of co-inheritance greater that the expected 50%, as if they were on different chromosomes or recombination could occur randomly between them. For such alleles that remain together on a paternal haplotype ("unrecombined"), we say that they are in "linkage disequilibrium" (**29-32**).

1.1.4. Study Designs of Genetic Epidemiology: Extended Pedigree Studies, Affected Sib-Pair Studies and Genetic Association Studies

Based on this understanding and definition of polymorphic markers and their positions, the first genetic epidemiological study designs aiming to find alleles responsible for diseases could begin to emerge. There are presently three very broad strategies to identify genetic variants associated with susceptibility to human diseases: (i) extended pedigree studies; (ii) affected sib-pair studies, and (iii) genetic association studies of population samples.

The pedigree approaches emerged from Mendelian understanding of the inheritance, and reflect primarily the interest of biologists in finding clues to the aetiology of diseases. The latter two strategies emerged from the concept of disease risk factors at the interest from the perspective of public health. Affected sib-pair linkage studies represent the extension of early genetic epidemiological studies trying to assess disease risk in relatives of the affected individuals in comparison to the general population. Genetic association studies reflect primarily the interest of epidemiologists in population-attributable fraction associated with different disease risks, which in this case are genetic variants present in different frequencies in a population.

All three approaches were made possible by the development of so-called "linkage maps", i.e. sets of hundreds of STR polymorphic markers, or hundreds of thousands of SNP polymorphic markers intended for "genome-wide scan". These "genome-wide scans" aimed to identify at least one marker, but preferably the specific haplotype, that would be found at excess frequency (with extremely high statistical significance) among affected disease cases in comparison to unaffected controls. When a marker was found in excess frequency among affected individuals in comparison to

controls, it would then be presumed that it must be physically "linked" to the disease gene. Then, the whole "candidate" region around the marker would be fine-mapped by a more dense set of markers to identify the precise position of the responsible gene (**37**).

The development of genetic maps of polymorphic markers made pedigree-based linkage studies of Mendelian disorders a routine. This was followed by the development of physical maps based on linkage disequilibrium, which further allowed fine-scale mapping. Historically, pedigree-based linkage studies have proven to be a very successful and powerful approach in finding genetic variants responsible for Mendelian (monogenic) diseases if several realistic requirements were fulfilled. Firstly, it was necessary to collect a number of nuclear (or extended) families with multiple disease cases. Secondly, it was essential that the affected persons were correctly classified as "diseased" and the unaffected as "healthy". Thirdly, all disease cases in each family needed to be caused by mutations in only one gene, resulting in an apparent pattern of disease inheritance ("Mendelian" diseases: autosomal or sex-linked, with dominant, recessive or codominant modes of inheritance). Fourthly, there should be no genetic heterogeneity in disease aetiology, meaning that all the disease cases in all collected families from the population were caused by mutations in the very same gene with a unique position in the human genome (**6, 37**). Fifthly, both "penetrance" and "detectance" needed to be reasonably high. The penetrance indicates a probability of developing the disease phenotype given the disease genotype (if less than 100% among carriers of the mutation that can cause the disease phenotype, it leads to misclassification of some study subjects and decreases the power of the study to detect underlying disease genes). The "detectance" indicates a probability of the disease genotype given the disease phenotype, as not all cases of disease phenotype in a population would have a single, or sometimes even any, genetic cause to the development of their disease, again leading to misclassification.

If the studied disease was recessive, especially if it was found in an inbred pedigree (which was usually the case), linkage analysis would be even more powerful.

The gene would have to be in the parts of the genome of the affected child that is inherited identical-by-descent from both parents (**38**). In the children of second cousins, only 1.56% of the genome is expected to be shared identical-by-descent, so if 800 STR markers were used to locate the gene, only about 10 of them would remain possible candidates after the analysis of a single case and his parents. Therefore, it was enough in some instances to find only 3 cases from an inbred isolate human population to be able to narrow down a "candidate region" to a single marker. Many genes for Mendelian diseases were indeed firstly detected in isolate and inbred human communities, as it was extremely likely that the rare genetic mutation had to be introduced on only one occasion in history by a single founder (**39**). Thus, genetic heterogeneity of the studied disease was not a concern, as was the case when the families were collected from the general population.

After the candidate marker linked to a disease gene was identified, fine-mapping of the surrounding region would be performed with a more dense set of markers, involving both STRs and SNPs. The density of selected markers for both the genome-wide scan and subsequent fine-mapping of a specific gene is always an important issue in any linkage analysis. As base pairs (nucleotides) measure physical distance between positions in the genome, Morgans measure the probability of recombination occurring between them per meiosis (generation). This makes it a more appropriate measure of distance between markers when they are used to locate the unknown gene, whereby 1 million base pairs roughly equals to 1 centimorgan (cM) (**40**). The use of a linkage set with markers that are 5 cM apart meant that there was only a 5% chance that recombination occurred between those two markers during meiosis. The initial genome-wide scans are therefore usually performed with markers spanning 5 or 10 cM, as this represents an acceptably low probability that recombination destroyed the connection between the marker and disease gene, given the large number of recruited families that would eventually point to the region of interest. Sometimes, however, the eventually determined candidate region spans over the entire haplotype of several markers, i.e. several megabases in distance, and harbours hundreds of genes. Therefore, it typically

took years until the causal genetic variant was found by fine mapping and its function understood. However, through this procedure a near-complete understanding of the genetic aetiology of those diseases was obtained, which would be impossible by any other available means.

Deciding which markers are significantly associated with disease phenotype is based on agreement among the scientists in the field on a procedure targeted to reduce the likelihood of false positive result. For linkage analysis, a sequential procedure is used for sampling and analysing extended pedigrees until the evidence in favour of linkage with a marker (usually expressed as logarithm base 10 of the likelihood ratio) reaches the level of 3. The "LOD score" value of 3 corresponds to *P*-value of 10^{-4}, when the χ^2 approximation is used for "one-sided" likelihood ratio test. This concept achieved wide acceptance because it was possible to assess the prior probability. For genetic association studies (see later), estimates of prior probability are usually very subjective, but essentially there is no prior evidence that could be used to assess and quantify prior probability. This coupled with ever larger number of polymorphic markers tested in genome-wide scans (up to 1 million SNPs or more), makes it very difficult to decide on the appropriate cut-off points of the *p*-value needed to reduce the reporting of the false positive associations. Another problem of genetic-association approaches are the phenotypes: when genetic basis of the multiple phenotypes is studied in the same set of samples, this gives rise to the statistical problem of multiple comparisons. Only the wide replication of the results in different populations may eventually help to bypass the two problems in the foreseeable future.

Extended pedigree studies are currently thought to be the most powerful approach for identifying "intermediate phenotypes", i.e., loci influencing the variation in quantitative traits that are confirmed risk factors for common complex diseases. These studies are being conducted in several isolated populations with available genealogical records, and the first results are awaited. The power of these studies to find genetic variants underlying common complex diseases of late-onset is nevertheless thought to be

limited, due to small number of disease cases in most of the isolated populations presently studied. Despite their success in finding genes underlying Mendelian diseases, extended pedigree studies of more common complex diseases yielded unrepeatable results and a rough theoretical analysis showed that they are vastly underpowered to detect the variants underlying more complex common diseases. Two alternative strategies were proposed: affected sib-pair studies (usually hospital-based) and genetic association studies (usually population-based).

Affected sib-pair studies are currently the predominant approach in genome-wide mapping of common complex diseases. Most of the conducted studies to date were underpowered to yield the promised results, but large international consortia are being formed to overcome this problem. This approach does not require genealogical reconstruction like the extended pedigree studies. It is robust to differences in genetic structure of study populations, which is a problem of genetic association studies. It does not require dense genome-wide scans and the related statistical analyses have been theoretically well developed.

Genetic association studies are the most intriguing and the least exploited of the three approaches to date. There is a hope that dense genome-wide scans using hundreds of thousands (or even millions) of SNPs in sufficiently large cohorts of diseased individuals and healthy controls will provide ultimate answers not only about genetic variants involved in disease pathogenesis, but also in public health significance of those variants at the level of the population. However, one of perhaps the main uncertainties that may determine the future success of the genetic association studies is that they too often rely on an assumption that diseases that are commonly found in population would have common allelic variants causing them. Those common variants should, ideally, all descend from a single mutational event occurring in one person in human history. In theory, this may be possible if harsh selection was introduced. For example, in a time of catastrophic historic epidemics, a person could have had a disease-causing mutation in his genome located very close to (i.e., "in linkage disequilibrium with") an important

and rare HLA variant that helped in surviving the epidemics. The selection ("epidemics") would then cause a massive reduction in the population size ("bottleneck effect"), after which the frequency of the favourable HLA variant would be dramatically increased in the surviving population, but along with it also the unfavourable disease-causing mutation. The other mechanism through which susceptibility variants for common diseases could have reached high frequencies is "antagonistic pleiotropy": a variant that was positively selected, because it controlled a fitness-improving trait early in life, may have negative health effects in the post-reproductive period (**41**). An apparent example of this "thrifty genotype" hypothesis is type II diabetes (**42**). It is very likely that humans were exposed to starvation throughout most of their evolution, thus positively selecting genetic variants for slow and highly efficient food metabolism. However, large environmental changes during the past century and the general availability of food in developed countries led to epidemics of obesity later in life, which seems to be the main determinant of type II diabetes.

In genetic association studies, all that is thought required to find the genes underlying late-onset complex diseases would be genotyping many disease cases in the population and many unaffected controls, until some differences in marker allele frequencies begin to show significant differences. Therefore, genetic association studies can be considered in some respects as linkage analyses where the entire population is considered a giant pedigree (**7**,**37**). However, it is still a matter of great debate if this approach is efficient, as it is based on a number of assumptions that in many cases may prove quite unrealistic (**8**). This will be further discussed in the chapter on the leading current approaches to find genes for complex diseases.

1.1.5. Recent Understanding of Genetic Determinants of a Spectrum of Human Diseases

1.1.5.1. MONOGENIC (MENDELIAN) DISEASES

The common feature of "Mendelian" diseases is that their entire phenotype is caused by a rare mutation in a single gene in the genome. Therefore, the segregation of affected individuals in families follows simple Mendelian predictions (**4**). The catalogue of known Mendelian diseases is regularly published and currently there have been 16,373 phenotypes with suspected or proven Mendelian basis identified to date (OMIM Statistics, Nov 13, 2005) (**43**). The last decade witnessed great successes in identifying genetic variants underlying more than 10,000 of these diseases (**11,43,44,45**). The key property of Mendelian diseases that made this success possible is that a causal genetic mutation is both necessary and sufficient for the development of disease. This ensures good correlation between disease phenotypes and underlying genotypes, given the penetrance of the genetic effect is high, which is an important requirement for successful linkage analyses or genetic association studies (**44,45**).

However, the initial successes in mapping variants underlying Mendelian diseases were soon followed by unexpected insights into the complexity of those "most simple" diseases. Two examples that were most extensively studied are retinitis pigmentosa and the thalassaemias. Following the mapping of initial variants responsible for the phenotype of retinitis pigmentosa in large pedigrees, it soon became apparent that there are many different genes, perhaps dozens, scattered throughout the genome, that may lead to the same disease phenotype when mutated. The aetiology of this condition proved even more complex when it was recognised that numerous genetic variants that underlie this condition follow very different modes of transmission: autosomal dominant, autosomal recessive, sex-linked dominant and sex-linked recessive (**37,46**).

Perhaps the most comprehensive insight into the complexity of phenotype-genotype relationships in monogenic disease was given by Weatherall (**45**). His studies on the thalassaemias across the world, arising through positive selection as a condition protective against death from malaria, but based on extremely different genetic mechanisms, showed how a remarkable diversity in phenotypes is encountered even in this relatively "simple" disease (**45**). Thalassaemias are probably the commonest human monogenic diseases, and approximately 7% of the world's population are carriers for different inherited disorders of haemoglobin. The extreme phenotypic diversity of this condition encountered throughout the world is determined by "*...layer upon layer of complexity*" (**45**). Firstly, there is a variety of primary mutations at the beta-globin genes, similar to the example of retinitis pigmentosa. Then, there is the action of two known "secondary genetic modifiers": alpha and gamma-globin genes, which affect the magnitude of the effect of excess of alpha chains. The result of the combined action of primary and secondary modifiers is then affected by an unknown number of less well defined "tertiary modifiers" (e.g. vitamin D receptor, oestrogen receptor, genes implicated in collagen synthesis, the locus for hereditary haemochromatosis, UGT glucuronyltransferase, HLA-DR locus, tumour-necrosis factor alpha, intracellular adhesion molecule 1) (**45**). Finally, it is recognized that environmental, ethnological and cultural factors also strongly affect the disease phenotype, although the underlying mechanisms are less clear (**45**). This all shows the complexity underlying even the "simplest" of genetically determined diseases and it should be taken into account when studies searching for genetic determinants of more complex diseases are designed, which is not usually the case.

1.1.5.2. "OLIGOGENIC" DISEASES

The initial successes in discovering genetic variants underlying monogenic diseases encouraged further progress with the diseases that showed high heritability and

were thought to be simpler in aetiology – "oligogenic diseases". An excellent example is Hirschprung's disease, which is the most common hereditary cause of intestinal obstruction. The pathogenesis of the disease was roughly understood, and the absence of ganglion cells in the specific plexuses of the gastrointestinal tract (myenteric and submucosal) was implicated as a cause (**47**). This understanding and the early onset of the disease led the scientists to believe that Hirschprung disease (HD) is mainly genetically determined and of relatively simple aetiology, although a clear Mendelian pattern of inheritance could not be established. A linkage analysis in pedigrees with multiple affected cases ("multiplex pedigrees") led to insights which categorised the disease by genetic aetiology and explained both the familial and population risk of the disease (**48**). There is the more common "short segment" form (S-HD), influenced by three susceptibility loci at the chromosomes 3, 10 and 19, that explain the complete population incidence of this form of HD. The gene at chromosome 3 is probably *RET*, which seems to have the crucial role in all forms of HD (but not both necessary and sufficient, as is the case with other genes that cause monogenic diseases). Other forms of the disease ("long-segment" and "syndromic" HD) are more rare and genetically more complex, with coding sequence mutations in *RET*, *GDNF*, *EDNRB*, *EDN3* and *SOX10* genes being implicated in various studies (**48**).

Another disease that seems to show an "oligogenic" determination of susceptibility is perhaps also an unexpected one – leprosy. Although this disease is infectious, development of the phenotype seems to be strongly genetically determined. A recent paper by Mira et al. (**49**) showed how the association of the disease with chromosomal region 6q25 was first implicated in a sample of 86 Vietnamese multiplex families using model-free linkage analysis. The association of the candidate region was then repeated in 208 independent simplex Vietnamese families consisting of both parents and one affected child (**50**). Fine mapping using single nucleotide polymorphisms implied that a regulatory region shared by genes *PARK2* and *PACRG* was responsible for leprosy susceptibility (**49**). The authors replicated this association in a sample of 975 unrelated cases and controls from Brazil, in whom the same variants

also showed significant association with leprosy in a candidate gene study (**49**). Another chromosomal region (10p13), earlier implicated in an Indian sample of "paucibacillary" disease cases (**51**), also showed strong association in a "paucibacillary" subset of Vietnamese cases, but not in the "multibacillary" subset. The authors concluded that variants in *PARK2* and *PACRG* are common alleles that confer susceptibility to leprosy "*per se*" globally, and that variants in 10p13 region are also common alleles that determine clinical presentation of disease as "paucibacillary" or "multibacillary" (**49,50**).

In recent years, more promising evidence is being gathered suggesting that some other diseases may have a reasonably simpler architecture of genetic susceptibility than common complex diseases of late onset. There is recently increased enthusiasm over the identification of variants underlying the susceptibility of asthma (**52-55**), systemic lupus erythematosus (**56**) and psoriasis (**57,58**).

1.1.5.3. COMPLEX POLYGENIC DISEASES OF LATE ONSET

The genetic basis of common complex diseases of late onset, responsible for most of the public health burden in wealthy countries of the world, is currently perhaps the greatest focus of interest of the entire biomedical scientific community. This is partly because the identification of common genetic risk variants in human populations would enable genetic screening and possibly provide new therapeutic targets for drugs that could be administered in the same manner to a large number of persons at increased risk. As both of those prospects would certainly be extremely lucrative and lead to an unprecedented increase in revenues for those producing genetic tests and drugs, the investments into research for genetic determinants of common late-onset diseases have been enormous during recent years. However, the output to date was hardly proportional to the investments. It appears that apparent successes in mapping genes for monogenic diseases and sequencing of the human genome prompted large number of research groups, as well as both private and public investors, into the “gold rush” (search for the

"Croesus Code") that may have been based on slightly over-optimistic assumptions (**59**). Primarily, the common diseases of greatest interest, i.e. cardiovascular disease, cancer, type II diabetes and psychiatric disorders, are frequently extremely complex phenotypes that are, contrary to most monogenic diseases, difficult to measure uniformly and define. Secondly, many of the approaches neglected the cumulative effects of the environment on disease development, being interested only in the genetic component. This is in spite of the knowledge that most of these diseases show rather low heritability and the majority of cases in general population may be due to environmental exposures. Thirdly, an appealing concept of "common disease/common variant" (CD/CV) gained popularity among the mainstream researchers in the field, based on the assumption that frequent diseases will be determined mainly by genetic variants common to all affected people from different human populations (**60**).

The outcomes of the research based on previous assumptions were not spectacular. This is not to say that there were no successes. Due to large investments, many small successes have been made and the research field is rapidly expanding. However, the massive undertaking of poorly designed genetic association studies based on possibly false assumptions resulted in a great number of reported associations of common diseases to numerous genes across the genome, but a substantial portion of published reports is likely to be false-positive. Therefore, the issues of repeatability and interpretation of such associations slowly became nearly as important as conducting the studies themselves (**61-63**). Some general conclusions can be drawn at this point. The variants that show repeatable associations with common diseases in more than one population are usually of very small effect and not always common in the populations under study. There were some isolated successes, such as the widely publicized association between APO ε4 variant in Alzheimer's disease, which indeed explained a substantial fraction of disease cases in the population, but the majority of the variants identified to date were unlikely to individually explain a substantial proportion of the disease burden in the studied population. Other variants that were implicated but not repeated may also be causal, but specific to the population under study (e.g., an unusual

gene-environment interaction in the studied population). The genetic architecture of common diseases is slowly beginning to reveal a large diversity of potential genetic causes, all of them acting through a somewhat limited number of mechanisms, with an increasingly appreciated contribution of environmental interactions (**3,44**).

Several individual efforts, however, increased our understanding of the genetic basis of complex polygenic diseases to the extent worthy of specific mention. The presented examples will be limited to cardiovascular diseases and cancer only, as those two complex diseases are jointly responsible for up to 75% of deaths in western countries and therefore represent the principal interest. The first example is the study by Ozaki et al. (**64**), in which about 1,000 cases of myocardial infarction (MI) were compared to roughly as many control individuals using 92,788 gene-based single-nucleotide polymorphism markers (SNP). The authors used this impressive number of markers in a nearly ideal high-tech genetic association study, conducted in a relatively genetically homogenous Japanese population. Although they covered about half of the entire human genome with their SNP markers, they could only find one statistically significant association (coding region of LTA gene on chromosome 6) when a recessive mode of inheritance was assumed, and no significant association ($p<10^{-6}$) under a dominant model. The increase in risk of variant carriers was modest, i.e. about 1.7. Although presented as a success, this study was actually rather discouraging for the proponents of genetic association studies that are based on extremely large numbers of SNP-markers, reasonably large samples of cases and controls from the outbred general population, and linkage disequilibrium. Even if the reported association is truly causal, it is certain that there must be far more genes underlying MI risk, but they were not identified even in this exercise that was massive in scale. The confirmation of the finding in a population other than Japanese is thus eagerly awaited.

Another interesting effort is the one performed by the deCODE Genomics company in the Iceland population. This company was founded in 1996 with the aim of identifying the genetic causes of common diseases and developing new drugs and

diagnostics based upon its research. It used a different approach from the one presented above, based on appreciating the large genetic heterogeneity and complexity underlying common diseases. The Iceland population was chosen as it offered most of the potential advantages needed to tackle this complexity – reduced genetic diversity, available disease data and reliable genealogical information. deCODE invested hundreds of millions of dollars into attempts to identify major genes involved in more than 20 of the most common diseases, and has successfully isolated genes in seven of these to date, which is possibly the greatest success rate by any group in the world. Two very recent examples related to cardiovascular disease are identification of the gene encoding phosphodiesterase 4D on chromosome 5 as a risk factor of ischaemic stroke (**65**), and the gene encoding 5-lipoxigenase activating protein on chromosome 13 as a risk factor for MI and stroke (**66**).

The investigations of genetic basis of cardiovascular diseases are still at a reasonably early stage. However, the research into the genetic changes found in human cancers has been conducted for decades (**67**). The extreme diversity and complexity of the causes, mechanisms and consequences underlying malignant transformation of the human cell is possibly a good predictor of what will be encountered in the future when studying the genetic basis of other complex diseases (**68**). It is now known that familial (monogenic) forms of cancer, such as breast cancer cases "exclusively" due to BRCA1 and BRCA2 mutations, account for only about 20% of the familial breast cancer cases, while familial cases constitute only about 5-10% of all breast cancer cases in the general population. Even among those "monogenic" breast cancers, only 25% can be explained by changes in BRCA and other known "breast cancer" genes, while the remaining 75% of familial cases are due to unknown familial predisposing genes. Non-familial cases, which constitute 90-95% of cases in general population, can therefore be explained only through the interaction of unknown polygenic predisposing genes and environmental factors (**69**).

Some of the changes in genetic material that are frequently postulated as occurring in tumour cells, although neither necessary nor sufficient in all cases in a population to lead to cancer, are mutations in coding or regulatory sequences, changes in overall ploidy, high amplification, structural rearrangements and loss of heterozygosity (**67**). The key feature of malignant cells is genomic instability, which can be due to inherited mutations in genes that monitor genome integrity, or acquired in any somatic cell during the development of cancer (**70**). However, the processes that follow are mediated through an extreme diversity of mechanisms, where it is difficult to distinguish the changes that led to cancer from the changes that arose as a consequence of cellular transformation. The amount of published results in cancer research that is becoming available on different molecular genetic aspects of the disease in recent years is so vast, that possibly the leading current problem seems to be integrating and coordinating this knowledge (**68**). It is hoped that many discovered signalling pathways act in parallel through organized networks, but the only way to find those universal principles that are somewhat more limited in number is to combine models and experiments. To achieve this, developing a system of categorization of knowledge will be essential, and one such effort is represented in the National Cancer Institute's Cancer Genome Anatomy project (**68**). It is probable that the experience with cancer genetics and genomics will soon be repeated through research into genetic causes of other complex chronic diseases of late onset that were not mentioned here (e.g. psychiatric disorders, type II diabetes, and others).

1.1.6. Current Understanding of Genetic Architecture of Common Complex Diseases

There is still a lot of uncertainty and a great deal of controversy over an understanding of the genetic architecture underlying complex chronic diseases of late onset. These diseases occur mainly in the post-reproductive period, and their genetic determinants are therefore less subject to selective impacts from the environment than is the case with more simple (monogenic and oligogenic) diseases of early onset. However, cumulative negative impacts of the environment over time are also more important in the aetiology of late-onset diseases than in early onset diseases. Late-onset diseases are therefore not only genetically more complex, but also multifactorially determined. The key questions that gave rise to recent debates are about the frequency of the responsible susceptibility variants in a population (common / rare), the number of loci in the genome that underlie these diseases (oligogenic / polygenic), and on the size of their effects (large / small).

Some argue that, because diseases of late onset are quite common in a population, their genetic determinants (variants responsible for increased susceptibility) should, intuitively, also be common and therefore evolutionarily rather old. This is known as the "common disease - common variant" hypothesis (CD/CV) (**60**). If this were true, genetic association studies would be expected to be successful and to lead to the identification of susceptibility variants. Others, however, argue that, although counter-intuitive, common diseases are more likely to be caused by a highly complex interaction of numerous genetic variants, most of them very rare, interacting among themselves and with the environment. This hypothesis is known as "common disease - rare variant" hypothesis (CD/RV), and would largely undermine currently proposed efforts to identify disease susceptibility genes using genetic association studies (**3**). It would promote the view that it is not realistic to expect one or a few mechanisms of

pathogenesis per disease, but that there are many "private" mutations coupled with a diversity of environmental exposures.

The patterns of human genetic diversity are of interest as it is thought that the variants in our genomes make individuals susceptible to specific late-onset diseases. Firstly, in more than 99% of our genome sequence there is practically no diversity, and the variants at those loci are fixed (i.e. have a population frequency of 100%). However, the figure of 100% is not entirely accurate, as it is possible that virtually any single nucleotide in this "invariant" part of the genome may be changed ("mutated") in any individual. However, this is not considered a true polymorphism, as such mutations have incredibly low population frequencies, i.e. practically one in a number equaling twice the total human population size for those occurring the first time in the autosomal part of the genome. Such a newly arisen single nucleotide polymorphism (SNP) would have to increase its frequency over the course of human population history from 1 in hundreds of millions of people to 1 in 100, e.g. 6-7 orders of magnitude, to become a true polymorphism in the population. If that occurred, this SNP can be considered a "common" variant. It is predicted that about 12 million single nucleotides in the genome, i.e. less than 1%, should be polymorphic (**33,71**). This magnitude of increase in population frequency for the newly introduced mutations should only be possible through 2 general mechanisms: long-time random genetic drift, or positive selection favouring the carriers in each new generation due to the beneficial effects of such a mutation in pre-reproductive life.

Going back to the "monomorphic" ("invariant") majority of the human genome, it is possible that even this part could confer susceptibility to the development of late-onset diseases. In that case, all humans would eventually get the disease after being exposed to their environments for sufficiently long periods. The difference in age of onset of disease cases would be determined solely by cumulative exposure to environmental risk factors during their lifetime. This is a "common disease-fixed variant" hypothesis (CD/FV), and there are good examples and plausible explanations

why this would be the case for some diseases (**42**). For example, it is very likely that starvation was a major selective pressure during most of human history, and that selection strongly favoured new variants that were protecting humans from hunger through more efficient food metabolism. If those variants became fixed, and everyone in the present human population possesses them, the large environmental change in which food became easily accessible in supermarket chains of the western world over the past 50 years would be expected to lead to a pandemic of obesity. This scenario is indeed being observed nowadays. It is likely that atherosclerosis is another example of an "universal" disease, the development of which depends only upon the sum of environmental effects during lifetime.

Two important implications of the "CD/FV" hypothesis should be noted. Firstly, in large outbred populations it would be useless to search for extremely rare variants in this nearly "invariant" part of the genome that would additionally increase risk e.g. for obesity or atherosclerosis above the "universal" risk shared by everyone. This is because the population attributable fraction of disease cases due to those specific variants would be negligible and would not lead to any feasible prevention of treatment strategies. Secondly it is apparent that changing behaviour and reducing risky environmental exposures would have much larger public health effects than any improvement in the understanding of genetic basis of diseases under the CD/FV hypothesis. It is possible that fixed variants do indeed play an important role in the genetic architecture of many common complex diseases of late onset, which could partly explain the low genetic variance and high environmental variance in many complex traits and diseases.

However, for other complex diseases of late onset, such as some types of cancer, psychiatric and neurological diseases, it is clear that significant heritability can be noted, and it is improbable that all humans would eventually develop those diseases after enough time. In such diseases (e.g., breast cancer, manic depression or multiple sclerosis), genetic factors are likely to play an important role in disease predisposition. As there is variation among humans in their predisposition to developing those diseases,

shown through demonstrated familial clustering in the population and increased disease risk in relatives or twins of affected cases, it is thought that this variation is mediated through polymorphic sites in the genome. The key question, however, remains whether the predisposition to disease is a result of the action of variants at several loci (oligogenic genetic architecture), all of which carry a reasonably large relative risk (e.g. RR>2.0) and are common in a population (CD/CV hypothesis). The alternative hypothesis is that there are many loci across the genome that interact among themselves and with the environment (polygenic genetic architecture), most of which carry a very small relative risk associated to individual variants (e.g. RR<1.5) and are very rare in a population (CD/RV hypothesis). Under the first model, the identification of several responsible variants of large effect would certainly provide clues into disease pathogenesis, and enable genetic screening, prevention and gene-based therapy. Under the second model, the identification of individual rare genetic variants that marginally increase disease risk would contribute very little to an understanding of the disease pathogenesis and would not lead to feasible diagnostic and therapeutic advances.

The two hypotheses are not necessarily mutually exclusive, and there are arguments to support both. Lohmuller et al. reviewed the replicated gene-disease associations in the world literature, the associated relative risks, and the frequencies of the implicated variants in the population, and concluded that there is support for the CD/CV hypothesis (**72**). However, the associated relative risks were usually overestimated in the first published reports and they appear rather small, so that the identified associations largely failed to improve the understanding of disease pathogenesis. It is still thought, however, that the general lack of success in mapping complex disease genes is due to most of the current studies being under-powered (using too few cases and genomic markers to detect associations), and that improved designs and meta-analyses should detect more common variants (**8,73,74**). Others doubt that even the increase in number of subjects or number of markers should necessarily help (**3**). They argue that if variants became common in a population, they are likely to either be neutral during the pre-reproductive period (and thus increased in frequency by

genetic drift), or to have beneficial effects on fitness in the pre-reproductive period (and therefore be positively selected). This would imply that common variants with detrimental effects in the post-reproductive period would have had to be very old in evolutionary terms and neutral or even beneficial ("antagonistic pleiotropy" hypothesis) in early life. The authors consider this to be unlikely, based on a summary of the evidence from experimental organisms, or at least to not be a leading mechanism of common disease pathogenesis. Finally, there is a third scenario in favour of the CD/CV hypothesis which cannot be easily dismissed. It hypothesises that some very rare variants became extremely useful in times of large pandemics of infectious diseases, and rapidly increased in frequency over shorter periods of human history. A variant physically close to (in tight linkage with) the protective variant could then also increase in frequency via a "hitch-hiking" effect, as its detrimental effects on fitness were considerably smaller than the beneficial effects of the linked protective variant under the selective pressure of epidemics. This could explain at least some of the numerous reported associations between specific HLA groups and some relatively common human diseases (**75**).

The proponents of the CD/RV hypothesis mainly use arguments that rely on decades of fundamental research in population genetics and human evolution. As the human population underwent a massive expansion over the past several generations, modelling the predicted number of newly arisen mutations during recent human history implies that the majority of genetic variants contributing to current human genetic and phenotypic variation is predicted to be rare (**3**). This argument has recently been strengthened by the discovery that the estimate of the number of mutations per generation per gamete could have been historically underestimated by an order of magnitude (**76**). The recent work by Cohen et al. is the first prominent paper that empirically shows that the high-risk tail of the distribution of a complex quantitative trait - HDL cholesterol - is determined mainly by rare variants at the population level (**77**). Although it may seem counter-intuitive to some scientists (and certainly less attractive for industrial investment) that common diseases of late-onset are mainly caused by a

large number of rare variants with small effects, this long neglected view appears to finally be gaining some support.

To summarise the current state in this debate, it is generally accepted that the allelic frequencies in the population and their effect size have an "L"-shaped distribution. The alleles with very large effects, that could provide new insights into disease pathways and mechanisms, are predicted to be very rare in the population. At the same time, the alleles with tiny effects in pre-reproductive period are allowed by selection to become more common (**78**). The effects of many rare and possibly some common variants interacting among themselves and with the environment would cumulatively lead to the breakdown of intrinsic compensation mechanisms of the human organism and eventually manifest as the disease phenotype.

1.1.7. Leading Current Approaches to Identify Common Complex Disease Genes

Based on everything discussed so far, it can be concluded that two main approaches to identify common complex disease genes are emerging in the post-genome sequence era. The first approach, which currently represents mainstream research as shown by the large number of publications of its proponents in high-profile journals in recent years, is advocating expensive efforts in the general populations of western countries, such as U.K. and U.S.A. ("BioBanks") (**79,80**). Such studies are attempting to generate large population cohorts (of up to 1 million people) with great quantities of information on individual genetic background and environmental exposures. Then, massive genetic association studies are planned with tens of thousands of affected cases and hundreds of thousands of unaffected controls. Their genomes would be scanned using a large number of single-nucleotide polymorphisms (SNP), which may run into hundreds of thousands of markers per person. The key assumption that would eventually determine the success of this general approach is that the genes underlying most common complex diseases and their underlying haplotypes are common in the population (common disease/common variant hypothesis, "CD/CV"). A recent meta-analysis of genetic association studies published to date was supportive of a significant contribution of common variants to common disease susceptibility (**72**). The leading current effort following this direction is the "International Hap-Map Project" (**81**). This project assumes that most human genomic diversity is common, and is organised into distinct "haplotype blocks". Those blocks are presumably also common in the population, and so are the disease susceptibility mutations arising on those blocks. Ultimately, the catalogue of all variants of haplotype blocks in the human genome would enable associating them with common diseases, which would be a more feasible strategy than a genome-wide scan if the initial assumptions are correct. Moreover, each haplotype would be defined by a minimum informative number of SNPs needed to

distinguish it from other haplotype variants ("haplotype tagging"), which would greatly reduce the costs and effort of genotyping (**82**).

An alternative approach is based on the assumption that the key to success in mapping complex disease genes will be through decreasing their aetiologic heterogeneity, or subdivision by more detailed phenotyping into distinct disease subgroups, and improving correlation between genotypes and phenotypes in populations under study. This approach advocates the use of isolate human populations with a defined number and origin of founders, known ethnic history, possibility to define disease phenotypes, and reconstruct individual genealogical records (**39,83**). Some of the obvious advantages of this approach are that it is orders of magnitude less costly, and that linkage analyses and genetic association studies can be performed at the same time to support each other and increase the power of the study. However, the main advantage may be that this approach should work even if the variants underlying common complex diseases are rare in general population. This is because such rare variants with large effects may still be common enough in an isolate population to be detected by a genetic association study. If they are also rare in an isolate population, they still may be detected by linkage analysis, through an approach that is similar to the mapping of monogenic (Mendelian) diseases in isolate populations, which already proved successful in the past. Therefore, even if most of the genetic diversity underlying common diseases proves to be rare, which is somewhat counter-intuitive but predicted by population genetic theories (common disease/rare variant hypothesis, "CD/RV"), the variants could still be identified (**3**). The problem with studying isolate populations are small sample sizes and that the results may not be relevant for and applicable to wider, general populations. However, it seems that this approach is being taken more seriously in the research community due to recent successes in Iceland (**65,66**). This is especially true because Iceland's most recent success, finding a gene that increases susceptibility to myocardial infarction, was mapped on a rather common haplotype in Iceland. The association with the suspected genomic location was later declared confirmed in the population of United Kingdom by a candidate gene approach and genetic association study, but with a

different and less common haplotype involved in the latter population. This was presented as an example how the initial finding would not be possible in the general U.K. population, but only in an isolated population of Iceland, although there is an open remaining question whether finding an association with different alleles in the same gene could indeed be considered "a replication" of the original finding. The results from other isolate populations that are currently under study are eagerly expected, which include the populations of Newfoundland (**84**), the Saami (**85**), Sardinia (**86**), Israel (**87**), Netherlands (**88**), Croatia (**89**) and Dagestan (**90**), to mention a few.

In addition to those two leading over-arching approaches that are attempting to discover genetic variants underlying complex common diseases in a comprehensive way, and to find genes for many different diseases and traits within the same study, there are also other, more specific approaches that led to important advances. Many disease-oriented groups throughout the world formed multi-centre initiatives to gather a large number of patients with a specific diagnosed disease of interest, mainly using affected sib-pair design. In such cases, unlike in the two approaches mentioned above, the recruitment of the disease cases is not population-based, but rather hospital-based. When an adequate number of cases is recruited for the study, the "candidate regions" harbouring genetic variants of interest are initially sought for by relatively small number of markers (usually several hundreds). These approaches are presumed especially powerful if the diseases in collected cases were of early onset. This general approach, although only single disease specific, and subject to a number of confounding effects (such as population stratification) (**91**) and often disregarding environmental effects, has still been quite successful. Some repeatable associations in the current literature were reported after this initial approach (**92,93**). The comparative strengths and weaknesses of pedigree-based approaches, affected sib-pair studies and genetic association studies were already extensively discussed in Chapter 1.1.4.

1.1.8. Translating Genomic Knowledge into Public Health Benefits

There are two main expectations from genetic epidemiology in terms of delivering results that would have major impact on public health. The first one is the association of different genetic variants with specific health risks and the translation of that knowledge into the development of commercially available genetic tests that could predict diseases. The second one is the understanding of disease mechanisms and obtaining new insights in disease pathogenesis, which would reveal new targets for the development of drugs that could prevent or reverse the course of human diseases (**94**). Although several years ago those targets appeared far from reality, today there is a growing economic sector of biotechnology, in which a large number of private companies are attempting to deliver one or both of those goals, and some tests and genome-based drugs are already being offered on the market. We will briefly address the current status of advancement towards those two main goals of genetic epidemiology.

In terms of genetic testing, recently also called "genomic profiling", it is based on an expectation that knowing most of the genetic variants that could increase the risk of disease would enable the development of "DNA chips" containing this information. Those chips could then scan for the presence of an extremely large number of such variants in any individual's genome at birth. After the scan of the genome, the chip would compute the lifetime risk of various diseases, thus being a powerful tool of a "personalised medicine". Although a number of private companies already offer genomic profiling for "oxidative stress", "susceptibility to obesity or osteopenia", "nicotine or alcohol dependance", etc., these could hardly have any scientific basis, as the genetic architecture of those traits and responsible variants are simply not known with any accuracy at present. Therefore, the problem of regulating the marketing of such tests is growing recently, as it is entirely unlikely that we could have useful and reliable genetic tests that could predict individual risk of common complex diseases in the foreseeable future (**94**). However, in the meantime it is certain that the ease of marketing

of those tests, coupled with the desire of consumers in some western countries to actively control their health at any cost, may result in the creation of a smaller market for these tests of unproven value. Similar or even more dramatic examples have already been seen with the popularity of various diets and food supplements (**20,21,95,96**).

However, although predicting complex diseases in individuals based on their unique genome sequence may still be far from reality, there have been some positive developments in achieving the second target – associating genes with diseases to understand aetiology which eventually led to new drug discoveries. The two frequently cited examples are imatinib and trastuzumab (**97,98**). Imatinib (Gleevec) was developed following the discovery that a chromosome translocation created a new gene structure in some patients with chronic myeloid leukaemia, and the drug binds to the protein product of this gene and fights disease progression where other treatments fail (**97**). Trastuzumab (Herceptin) did not appear to significantly improve the survival of breast cancer patients, until it was realised that it is very efficient, but only in a subset of Her-2 positive breast cancer cases. Those examples based on a molecular understanding of disease pathogenesis may seem spectacular, but the more general view is that the successes in finding the genes underlying common complex diseases have been very modest in relation to the unprecedented investments into this research from both industry and the academic community. Although numerous associations of various genes with a spectrum of diseases have been reported, only a few of those associations have been replicable. Moreover, the risks associated with implicated variants were usually very modest. The findings to date promised little hope for contributing to an improvement in the understanding of disease aetiology. As the current level of investment in unsustainable, it is becoming apparent that the successes in mapping genes for Mendelian diseases will not be easily repeated with complex diseases, and that more rational strategies for associating genes and disease phenotypes will need to be developed.

Based on these premises, the recent review by Merikangas and Risch (**99**) provided a more sober assessment of the current status of research for common complex

disease genes and strategies for future investments. The authors argue that investments are justified for the diseases that are: (i) common in the population (associated with substantial public health burden); (ii) can be diagnosed precisely (to avoid misclassification of cases, which dramatically reduces the power of genetic association and linkage studies); (iii) show substantially increased risk in relatives of diseased cases (to demonstrate the role of genetic effects as opposed to environment); and (iv) for which no preventable environmental risks are known. It is worrying that enormous funds are being invested in searching for the genetic basis of diseases or conditions that hardly show any heritability, cannot be diagnosed with any precision, or for which the funds would be far better placed in fighting environmental risks rather than searching for genetic clues. For example, investing funds in finding genes for increased individual "nicotine dependence" or "alcohol dependence" is entirely misplaced, as those traits have been shown to cluster more strongly in social groups of different genetic background rather than in families. In addition, the benefit of public health intervention on reducing nicotine and alcohol consumption in a population, which are cheaper than gene searches, far outweigh any possible benefit that could come out of knowing genes that predispose an individual to alcohol consumption or smoking. The majority of cases of cardiovascular diseases or diabetes type II in a population can probably be explained by environmental risks such as unhealthy diet, lack of physical activity and smoking. It is unlikely that finding genetic variants that mildly increase the risk of those diseases, and even developing therapies based on that knowledge, would lead to an appreciable decrease of their public health burden in a population.

Examples of diseases where the genomic revolution could prove helpful, however, are Alzheimer's disease, multiple sclerosis, autism or schizophrenia, where the relatives of diseased individuals are clearly at greater risk, there are no known preventable environmental risks, and which are sufficiently common in the population to justify large investments. For other diseases, the funds would be better placed into research of determinants of human behaviour and motivation for leading more healthy lifestyles. This is all particularly relevant for the population of western countries, where

an estimated 150 million people already have type II diabetes and are overweight. However, only a minority of world's population lives in developed countries. In recent years, calls have been made upon the international scientific community not to forget the majority of world's population that does not represent a lucrative market for the pharmaceutical industry, but could also perhaps benefit from new genomic and molecular technologies, even more than the western world (**100**). The fact that 11 million children under five years still die annually of mainly preventable or easily treatable causes, such as pneumonia, diarrhoea, malaria or malnutrition, and that more than two thirds of people with AIDS live in countries with virtually non-existent health systems, are more than worrying. Those people could greatly benefit from recombinant vaccines that use genomic technology, or from molecular tests that could precisely diagnose the aetiology of their infections and thus enable more efficient use of sparse medicines available to those populations. It remains to be seen whether the genomic revolution of the 21st century will truly revolutionise medicine and result in major public health benefits for all of humanity. The alternative scenario is that it may only deliver partial successes which will become available to the rich minority and thus further increase the gap between the world's rich and the poor, as was the case with recent revolutions in informatics and telecommunication technologies in 1980s and 1990s.

1.2. STUDYING THE EFFECTS OF INBREEDING ON HUMAN TRAITS AND DISEASES

1.2.1. Introduction

In many parts of the developing World and in many communities within the developed World large proportions of all marriages are still among close relatives. The reasons include geographic, tribal, cultural or religious isolation or socio-economic motivation such as preservation of property, particularly the land (**101,102**). The degree of inbreeding in offspring of such marriages can be measured by the genetic term "inbreeding coefficient" (*F*), which indicates the proportion of the autosomal genome which is expected to be homozygous through inheritance of identical genes from common ancestors (i.e. proportion of alleles identical by descent (IBD) or "autozygosity"). The *F* value is calculated from genealogical information and it amounts to about 6% in the offspring of first cousin parents and 25% in the offspring of incestuous unions of first-degree relatives (**103**). The apparent risk in the individuals with a considerable proportion of their genes homozygous for identical allelic variants is the occurrence of "Mendelian" (monogenic) diseases caused by rare and recessive deleterious autosomal mutations (**104**).

The effects of inbreeding have historically focused on early-onset diseases, mainly recessively inherited monogenic (Mendelian) diseases, birth defects, decreased fertility and early mortality. This was due to the long-term experience that consanguineous unions were more likely to result in genetic diseases of children, most of which had an apparent phenotype that was very notable in the communities. Therefore, the vast majority of research on inbreeding effects had been focused on pre-reproductive health problems, and the risks have been thoroughly evaluated by numerous groups and individual authors (**102,105-111**).

During recent years, the rapid development of molecular genetic techniques enabled precise measurement of genetic characteristics of humans. This has led to the discovery of many genes responsible for Mendelian (monogenic) diseases which usually have early onset and are very rare. However, success in finding the genes responsible for the susceptibility to complex chronic diseases that currently represent most of the public health burden in both developed and developing world has been limited (**8,10**). The genetic architecture underlying late-onset diseases such as cardiovascular diseases, malignancies, adult-type diabetes and psychiatric disorders is still the matter of open debate. The genetic model that has recently begun to emerge from the experiments in plants, animals and human studies favours the hypotheses that the genetic variants associated with the increased risk of complex chronic diseases are more likely to be: (a) rare than common in the population (**73,74**); (b) numerous than rare across the genome (**3**); (c) of a small than of a large effect (**72**); (d) recessive than dominant (**18**).

The concern is that those hypotheses, if valid, imply that inbreeding in humans could have a considerable effect on the occurrence of complex late-onset diseases. The proposed mechanisms will be discussed in detail in the discussion part of this thesis, but the main reasons why this may be the case are presented here. First, modest levels of inbreeding observed in human populations are expected to have much larger effects on the population distributions of polygenically determined traits than on oligogenic traits and diseases. Second, inbreeding is predicted to have larger effects on the population-attributable fraction of disease cases if the underlying variants are rare rather than common. This is because common recessive variants will occasionally become homozygous in the population by chance, without a need for inbreeding to bring them together. If the variants are very rare in the population, and inbreeding is almost the only realistic scenario under which they can become homozygous in an individual, then the fraction of disease cases in the population who are the offspring of related parents will be much larger. Third, the risk of genetically determined disease in pre-reproductive age may be rather minor because the responsible mutations are dramatically reducing fitness and are therefore constantly being cleared from the gene pool by selection. In the post-reproductive

period, however, it is possible that numerous rare and recessive genetic variants underlying complex traits and susceptibility to common late-onset diseases have not been selected against during the human history, and that the genomic target for the effects of inbreeding is an order of magnitude larger. Fourth, the inbreeding depression (a change, usually a decrease, of the value of the trait due to the effect of increased homozygosity of the genes underlying the trait) should, at the same time, affect many phenotypes that are known as late-onset disease risk factors, such as blood pressure, body mass index, and possibly cholesterol and glucose levels. Even if the effect of inbreeding depression on each of those phenotypes individually is rather small, their risk is mediated in multiplicative manner, so the cumulative net effect of inbreeding depression on all the risk phenotypes during a lifetime could be more substantial than appreciated (**112,113**).

The concern over the proposed risk mechanisms of inbreeding on late-onset diseases is exaggerated by the fact that remarkably few publications addressed it, as the research into inbreeding was focused on assessing the risk of early mortality due to rare recessive deletorious mutations. In this chapter, a review of the literature is presented from 1965 onwards to assess the human population exposure to the "risk" of inbreeding, i.e. to assess the prevalence of inbreeding in contemporary human populations. Also, the empirical evidence on the effects of inbreeding on human biology and health is reviewed in detail, especially concerning biomedically relevant quantitative traits and late-onset complex diseases. After empirical evidence is reviewed, theoretical arguments on how inbreeding studies can improve an understanding of the genetic architecture of late-onset complex traits and diseases will be presented, which was the main aim of this thesis.

1.2.2. Global Prevalence and Distribution of Inbreeding

To estimate the global prevalence and distribution of inbreeding, a thorough search of the PubMed database was performed, looking for the references published from 1965 to date mentioning the terms "inbreeding", "consanguinity", "identity by descent" and/or "autozygosity", with a limit set to "human". This yielded more than 10,000 references, but only about 800 were relevant to the scope of this estimation process. Another 200 important references that were either book chapters or published before 1965 and found in bibliography of the 800 primarily identified references. However, the most informative resource on which I heavily relied to double-check my own abstraction process was a comprehensive web-site edited by Professor Alan H. Bittles, which can be found at **www.consang.net/global_prevalence/tables/**.

The populations studied in the retrieved references range from entire continents (Europe, Iberoamerican populations), countries and subcontinents (India, Pakistan, Oman, Jordan, Kuwait), to specific geographic, religious or ethnic isolates (**114,115**). As the aim of this review was to produce estimates at the level of specific countries, and the large majority of identified studies focused on specific isolates and minorities, only a limited number of studies was eventually informative for this work (**116-165**). **Table 1-1** reviews the studies which aimed to assess the proportion of consanguineous marriages and/or coefficient of inbreeding in different countries or their sub-populations using community-based study designs, so their results can be considered representative for the populations under study. For countries without published information on the prevalence of inbreeding, conservative estimates were based on information available for neighbouring countries with a similar ethnic structure.

Table 1-1 presents the country-level prevalence of consanguineous marriages defined (according to WHO) as a union of spouses that are related at the level of second cousins or closer (**114**). The prevalence of consanguineous marriages

presented in this table is always conservative, in a way that it was consistently rounded to the closest multiple of 5% that was lower than the actual estimate from the study. This was done as the aim was to assess the lower limit of the persons globally who are inbred at the level corresponding to offspring of second-cousin marriage or greater, i.e. have an inbreeding coefficient of 1.56% or greater. Under the assumption that consanguineous unions produce a similar number of offspring to those from non-consanguineous unions, this could then be translated into an estimate of the proportion of the population of those countries who are inbred. This assumption was recently supported with empirical data, because although there were studies implying decreased fertility in consanguineous marriages and increased childhood mortality of their children, these marriages tend to compensate these effects by having more children (**166**). It should be noted that in countries where inbreeding is most prevalent, the large majority (over 70%) of marriages tend to be between first cousins (expected inbreeding coefficient in offspring of 6.25%) (**106,125,128,132,137**). Therefore, it is reasonable to assume that the total share of the human population that show inbreeding as defined by WHO criteria are expected to have an average inbreeding coefficient greater than 3.0%. This estimate is based on the expected ≈200 million people in Arab countries being inbred according to WHO criteria, with more than 70% of them being the offspring of first-cousin marriages (F=0.0625). Another ≈200 million people are from India, with a substantial share of them also being the offspring of first-cousin marriages. With some 250 million people having F value of at least 0.0625 only in Arab nations and India, and the rest of those fulfilling WHO criteria having F greater than or equal to 0.0156 (by definition), it becomes apparent that those who we refer to as "inbred" have an average F value of greater than 0.03 when this is considered globally.

Table 1-1: Conservative estimates (multiples of 5%) of the proportion of consanguineous marriages for different countries and regions (using WHO criteria of consanguineous marriages - second cousins or closer) and WHO population estimates for 2000 classified by WHO regions*.

WHO Region / Country	Population 2001 (millions)	Percentage of consanguineous marriages	Reference	Inbred population (millions)
AFRO D				
Algeria	31.5	20%	**116**	6.3
Nigeria	111.5	50%	**117**	55.8
Angola, Benin, Burkina Faso, Cameroon, Cape Verde, Chad, Comoros, Equatorial Guinea, Gabon, Gambia, Ghana, Guinea, Guinea-Bissau, Liberia, Madagascar, Mali, Mauritania, Mauritius, Niger, Sao Tome and Principe, Senegal, Seychelles, Sierra Leone, Togo	150.7	**20%	---	30.1
AFRO E				
Botswana, Burundi, Central African Republic, Congo, Cote d'Ivoire, DR Congo, Erithrea, Ethiopia, Kenya, Lesotho, Malawi, Mozambique, Namibia, Rwanda, South Africa,	337.5	**15%	---	50.6

Swaziland, Uganda, Tanzania, Zambia, Zimbabwe				
AMRO A				
Canada, Cuba, USA	320.7	1%	**118,119**	3.2
AMRO B				
Antigua and Barbuda, Argentina, Bahamas, Barbados, Belize, Brazil, Chile, Colombia, Costa Rica, Dominica, Dominican Republic, El Salvador, Grenada, Guyana, Honduras, Jamaica, Mexico, Panama, Paraguay, St Kitts and Nevis, St Lucia, St Vincent and the Grenadines, Suriname, Trinidad and Tobago, Uruguay, Venezuela	431.2	2%	**118,119**	8.6
AMRO D				
Bolivia, Ecuador, Guatemala, Haiti, Nicaragua, Peru	71.3	2%	**118,119**	1.4
EMRO B				
Iran	67.7	30%	**120**	20.3
Jordan	6.7	40%	**121**	2.7
Kuwait	2.0	35%	**122**	0.7
Lebanon	3.3	25%	**123**	0.8
Oman	2.5	30%	**124**	0.8
Saudi Arabia	21.6	40%	**125**	8.6
Tunisia	9.6	25%	**126**	2.4

United Arab Emirates	2.4	30%	**127**	0.7
Bahrain, Cyprus, Libyan Arab Jamahiriya, Qatar, Syrian Arab Republic	23.7	**25%	---	5.9
EMRO D				
Egypt	68.5	25%	**128,129**	17.1
Iraq	23.1	45%	**130,131**	10.4
Pakistan	156.5	50%	**132-136**	78.3
Sudan	29.5	50%	**137**	14.8
Yemen	18.1	40%	**138**	7.2
Afghanistan, Djibouti, Morocco, Somalia	61.8	**25%	---	15.5
EURO A				
European Union (EU-15 in 2000), Andorra, Croatia, Czech Republic, Iceland, Malta, Monaco, Norway, San Marino, Slovenia, Switzerland	404.7	1%	**139**	4.0
Israel	6.2	10%	**140-142**	0.6
EURO B				
Turkey	66.6	20%	**143,144**	13.3
Azerbaijan, Kyrgyzstan, Tajikistan, Turkmenistan, Uzbekistan	47.4	**20%	---	9.5
Albania, Armenia, Bosnia and Herzegovina, Bulgaria, Georgia, Poland, Romania, Serbia and Montenegro, Slovakia, TFYR Macedonia	102.9	**1%	---	1.0
EURO C				

Kazakhstan	16.2	**20%	---	3.2
Belarus, Estonia, Hungary, Latvia, Lithuania, Republic of Moldova, Russian Federation, Ukraine	229.5	**1%	---	2.3
SEARO B				
Indonesia	212.1	15%	**145**	31.8
Sri Lanka	18.8	20%	**146**	3.8
Thailand	61.4	**5%	---	3.1
SEARO D				
Bangladesh	129.2	10%	**147**	12.9
India	1,013.7	20%	**106,148-57**	202.7
Bhutan, DPR Korea, Maldives, Myanmar, Nepal	95.9	**5%	---	4.8
WPRO A				
Australia	18.9	1%	**158,159**	0.2
Japan	126.7	5%	**160,161**	6.3
Singapore	3.6	5%	**118**	0.2
Brunei Darussalam, New Zealand	4.2	**2%	---	0.1
WPRO B				
China	1,284.5	5%	**162-165**	64.2
Malaysia	22.2	5%	**118**	1.1
Cambodia, Cook Islands, Fiji, Kiribati, Laos, Malaysia, Marshall Islands, Micronesia, Mongolia, Nauru, Niue, Palau, Papua New Guinea, Philippines, Republic of Korea, Samoa, Solomon Islands, Tonga,	228.8	**5%	---	11.4

Tuvalu, Vanuatu, Viet Nam				
TOTAL - ALL REGIONS				
World	6,044.9	11.8%	**116-165**	718.7

* The regions are defined by WHO region and child and adult mortality stratum (see statistical annex of World Health Report 2000 at http:/www.who.int/whr2001/2001/archives/2000/en/pdf/Statistical Annex.pdf). Thus, Afr, Amr, Emr, Sear and Wpr refer to the United Nations African, American, Eastern Mediterranean, South East Asian and Western Pacific regions; and the B-E suffixes refer to the economic grouping of the country according to UN classifications, with E being the group with the lowest level of economic development

** Conservative estimates based on information available for neighbouring countries with similar ethnic structure

Table 1-1 summarises conservative estimates of the prevalence and distribution of consanguineous unions and inbred individuals globally, based on the reviewed material (**116-165**). The countries are grouped according to the statistical annex of World Health Report published by the WHO in 2000 (**167**). Although the relative share of consanguineous unions is estimated to be largest in Pakistan and Sudan, the greatest number of inbred individuals is certainly found in India (about 200 million people). Pakistan (78 millions), China (64 millons), Nigeria (56 millions) and the rest of sub-Saharan Africa (81 millions) are thought to harbour at least 275 million additional inbred persons. Large numbers of inbred individuals can also be found in Indonesia (nearly 32 millions), Iran (20 millions), Egypt (17 millions), Sudan (15 millions), Turkey (13 millions) and Bangladesh (13 millions). They are followed by Afghanistan, Morocco and Somalia jointly (16 millions), Iraq (10 millions), Saudi Arabia (9 millions), Yemen (7 millions), Japan (6 millions), Algeria (6 millions) and former Soviet Union republics with Muslim population majorities (10 millions). Along with sporadic cases in Latin America, Asia and Eastern Europe and geographic, ethnic and religious isolate communities of the Western world, the conservative estimate of the number of inbred persons in the World amounts to 718.7 million (11.8%).

These are just the people with apparent recent consanguinity in their genealogies, at the level of second cousin marriage or closer. They are thought to have, on average, about 3% of their genome additionally identical by descent (see page 45 for calculation of this estimate; it should be very robust and apply to the current human population). With a recently corrected estimate of the number of human genes to 25,000 (**168**), and a mutation rate estimate per generation increased by an order of magnitude (**169**), this translates into these individuals having about 750 more autosomal genes autozygous in comparison to general population of their respective countries, which is the only relevant variable needed when assessing the population attributable fraction of the inbreeding effects. This is to say that the absolute levels of inbreeding are not as important as defining the relative increase in autozygosity in the subgroup of inbred individuals in comparison to the general population of similar genetic background that is exposed to the similar effects of the environment. As the estimated figure of 3% brings to homozygosity potentially numerous recessive mutations in those 750 genes that are mildly deleterious, the effects of this late in life are uncertain, but should not be neglected. For example, if each of those 750 genes would have a highly deleterious mutation in the population in frequency of 1:1000, there would be about 75% chance for each consanguineous individual to develop a disease, assuming the full penetrance.

To complicate things further, the demographic history of the entire human species allows the possibility that contemporary humans are more inbred than it is generally appreciated, but the amount of cryptic relatedness ("hidden" or "background" autozygosity) is difficult to estimate. Broman and Weber demonstrated that the amount of IBD homozygosity at the genome level in families sampled from isolate populations worldwide is up to several times greater than the genealogies in the last 4-5 generations would suggest (**170**). Still, it is not clear if this could also be true for the large outbred populations. Recent modelling of the period since the most recent common ancestor (MRCA) of all humans also showed that this time was surprisingly much shorter than widely anticipated (**171**). The most illustrative proof of this hidden autozygosity are probably the autosomal recessive Mendelian diseases in Western societies that are caused by extremely rare variants

showing no allelic heterogeneity, as their sheer occurrence in population is in many cases a reminder of cryptic autozygosity present in contemporary populations.

These hypotheses still must be taken with caution. Although I tried to estimate the "minimum" prevalence of inbred individuals globally, there is a chance that a significant portion of the studies from **Table 1-1**, upon which the estimates are based, are likely to be biased as they usually studied the population in which inbreeding was prevalent. Therefore, their results may not be applicable to countries as a whole. Also, the countries in which there were no studies of inbreeding prevalence may have a smaller prevalence of inbreeding than the neighbouring countries, from which their estimates were eventually derived. Still, it is not questionable that the number of people globally who are the offspring of related parents is substantial and that it runs into hundreds of millions. It is also likely that incestuous relationships still occur in many parts of the world, both developing and developed, especially in rural areas. Such cases are characterized by the highest values of individual inbreeding coefficients, and thus form the upper tail of the distribution of the values of inbreeding coefficients in human populations.

1.2.3. Effect of Inbreeding on Pre-Reproductive Traits and Diseases

1.2.3.1. EARLY MORTALITY

Most inbreeding studies have addressed the problem of early mortality, as the main concern over the effects of inbreeding focused on the possibility that every person carries at least several rare recessive genes which if homozygous would result in death before reproductive age in any and all environments ("lethal-equivalents"). Historically, reviews implied that the offspring of first-cousin marriages were 1.41 times more likely to die before they reach adulthood, but when the miscarriages and stillbirths were included, the relative risk of dying before the age of 16 increased to as much as 2.0 (**104**). However, it is possible that combinations of non-genetic factors such as lower socio-economic status of inbred families and generally high incidence of childhood mortality contributed considerably to those estimates. The most serious attempt to resolve this problem was carried out by Bittles and Neel, who reanalysed the reports on mortality increase in children of first-cousin marriages from 38 populations representing all major ethnic groups. The increase in mortality was assessed starting from about six months gestation (which included late miscarriages) to a median age of 10 years. The authors concluded that the increase in mortality in offspring of first-cousin marriages as compared to non-inbred controls is fairly constant across all populations and mortality rates. In their study it only amounted to about 4.5%, a significant deflation from previous estimates. This equals to an average human being heterozygous for 1.4 lethal equivalents (**13**). Later, Cavalli-Sforza et al. derived an even lower estimate through an extensive research conducted among the first cousin offspring in Italy (**172**).

1.2.3.2. EARLY MORBIDITY

The most comprehensive reviews of the effects of inbreeding on early morbidity were presented by Freire-Maia and Elisbao (**173**) and Khlat and Khoury (**111**). The studies of non-lethal childhood morbidity related to inbreeding are usually

confined to attempts to investigate the increase in the proportion of consanguineous unions among parents of children with rare recessive monogenic (Mendelian) diseases in comparison to the proportion of consanguineous unions in general population. This can then be used to compute a population-attributable risk for such diseases that is due to recent inbreeding. The complete list of all reported rare monogenic (Mendelian) diseases and syndromes is being regularly updated by the Online Mendelian Inheritance in Man (OMIM) service (**4**), and inbreeding usually represents a significant risk factor for those disorders that are recessively inherited. The diseases and syndromes most frequently studied in this context were cystic fibrosis, cystinosis, spinal muscular dystrophy, albinism, eye defects and hearing anomalies. However, although there are at least several thousands of such diseases, it should be understood that their impact on the total health burden is still relatively small, regardless of the prevalence of inbreeding in a population, although it may be more significant in terms of child health burden in certain areas where inbreeding is prevalent (**102**).

It is worth mentioning that in human isolates several population genetic mechanisms (that usually include inbreeding) act together to expose the phenotypic effects of rare recessive mutations, which proved to be very useful during the recent years in finding the genes responsible for monogenic (Mendelian) diseases. Due to the small number of founders of such populations, the allelic frequency of rare recessive mutations carried by founders increases by several orders of magnitude in comparison to general populations. Then, in these (usually very small) populations genetic drift can lead to further increase in the allelic frequency of a mutation of interest. If the populations are not small, genetic drift can only have an effect over a longer period and is therefore more important in older populations, and not in the new isolates. Finally, consanguinity (due to a reduced number of potential mates in a small population) brings those recessive alleles to homozygosity, exposing a disease phenotype in an affected individual. The theory behind this process helped molecular geneticists and genetic epidemiologists developing a "homozygosity mapping" approach which proved highly successful in mapping genes responsible for a large number of monogenic (Mendelian) diseases (**38**), although success with complex

chronic diseases of late onset has been limited to date.

1.2.3.3. COGNITIVE PERFORMANCE

Although intelligence can be considered a quantitative trait, while mental retardation is a qualitative trait that falls under the category of early morbidity, they are presented here separately as in many societies and cultures there is a strong belief that inbreeding affects these traits. There is a high degree of consistency between numerous reports in suggesting that inbreeding leads to a depression in IQ value with a mean regression coefficient of about $b = -45$ (**174**). Studies of the effects of inbreeding on mental retardation show that the absolute risk in children of apparently unrelated parents amounts to 0.012 and it increases to 0.062 in the offspring of first-cousin mating, a relative risk of about 5 (**175**).

Using this information, Morton (**174**) suggested that both the decline in IQ and the increase in mental retardation are consistent with a highly polygenic model of inheritance of these traits, and that there is little evidence that natural selection throughout human history favoured the individuals with greater IQ values in any way. The responsible genetic variants under Morton's model were rare, recessive and numerous. As a considerable reported effect of inbreeding was the consistent result, this further underscores the problem of the potential effect of inbreeding on highly polygenic late-onset complex diseases that were not subjected to natural selection throughout human history (**175**).

1.2.3.4. OTHER EFFECTS

The studies conducted on the effects of inbreeding in human populations showed a large variety of aims, and many of them do not fall within any of the categories mentioned in this review. As the increase in prenatal mortality was expected as a result of autozygosity, a number of studies investigated the effects on fertility and sterility. It is generally accepted that the risk of abortions and miscarriages is greater among consanguineous couples (**176-177**), but it has also been shown that those couples tend to have more children, possibly as a result of

reproductive compensation (**166,178**). Other interesting papers included studies of inbreeding effects on maternal-zygotic gene conflict over sex determination (**179**), twinning rate (**180-182**), developmental stability (**183**), erythrocyte antigen incompatibility (**184**) and X-heterosis (**185**). Some studies even investigated the frequency of consanguineous matings due to multiple use of donors in artificial insemination (**186**). The results of most of these studies were usually consistent with the hypotheses of some effects that could possibly be attributed to inbreeding, but those were predominantly of small size.

Another (indirect) way to point to the importance of inbreeding effects is to demonstrate that outbreeding and admixture between isolated populations and individuals of reduced genetic diversity increases fitness, improves health, and affects biological quantitative traits in the direction opposite to inbreeding depression. This phenomenon is called "heterosis" ("hybrid vigour"), and is well-established in plants and animals, although the studies in human populations are very rare (**187**).

1.2.4. Effects of Inbreeding on Post-Reproductive Traits and Diseases

1.2.4.1. QUANTITATIVE TRAITS

The predicted effects of inbreeding on quantitative traits (QT) were summarised by Falconer (**188**) and further discussed by Charlesworth and Charlesworth (**189**). As yet, only a few studies addressed this problem empirically in human populations. The most frequently investigated traits were simple anthropometrical traits (e.g. height) and some physiological traits (such as blood pressure). Here, the interest in reviewing the effects of inbreeding on human quantitative traits will be restricted only to those traits that are bio-medically relevant as they represent known risk factors for late-onset diseases in human populations. In 2002, the World Health Organization identified 5 leading risk factors for human late-onset mortality and disease burden globally. Three of them were human biological quantitative traits that represent leading risk factors for cardiovascular diseases, the most frequent cause of death in the world (**190**): high blood pressure, cholesterol and obesity (which can be thought of as high body mass index). In this chapter, the known effects of inbreeding on those 3 traits will be reviewed.

In a review of more than 8,000 references from 1961 to date only a few that studied the relationship between inbreeding and blood pressure (BP) were found. Most of them had an ecologic epidemiological design, i.e. they reported an increased prevalence of hypertension in populations that were inbred, without correcting for further potential confounding effects, such as specific controls being included taking into account e.g. socio-economical variables (**191-194**). In **Table 1-2**, the results of the three studies that included large number of examinees and reported a regression of BP values on inbreeding coefficient (F) estimated from genealogies are presented (**195-197**). This design, which is based on measures of both variables (F and BP) in many individuals, is superior to the ecologic design. All three studies were based on observations on several thousand examinees. All three studies reported a large

increase in blood pressure values with inbreeding, with the effect apparent on both systolic and diastolic blood pressure.

Several studies reported the effects of inbreeding on body mass index, or on components of its computation (height, weight). It should be stated that the effect on height is more important, as its value is factored into the calculation of BMI to the second power. Similar to the blood pressure example, there were studies investigating the effects on height using an ecologic design (**198-200**), but the same three studies from the previous blood pressure example were also used here to present this relationship (**195-197**). Although the effects reported on height and weight weren't consistent, in all three studies the net effect on BMI was showing an increase in value in persons with higher inbreeding coefficients.

Finally, only a single study investigating the effects of inbreeding on cholesterol levels could be identified (**196**). This study showed an effect known in human evolution studies as "antagonistic pleiotropy", i.e. opposing effects in early and later life, with decreased values in young age (<20 years) and significantly increased values later in life (>= 40 years).

Studying the effects of inbreeding on human quantitative traits can be potentially informative of the genetic architecture underlying those traits. Given the likely multifactorial determination of QT values in humans and the large environmental and genetic differences between populations, it will be difficult to disentangle the contributing effects of genetic and environmental factors on the observed distributions. A good discussion on this problem was provided by Rose (**201**). The most valuable information will probably come from populations that represent extremes either in their genetic structure, environmental exposures or unusual phenotypic mean and/or variance in QT. The methods and strategies to address the problem of variance component analyses of human quantitative traits in isolate population using available genealogical records are rapidly developing (**202-204**).

Table 1-2: Review of the studies investigating the effect of inbreeding on quantitative traits that represent known risk factors for late-onset diseases in human populations.

Quantitative trait	Author& Year	Ref*	Reported effect
Blood pressure	Krieger 1968	**195**	Significant increase in diastolic blood pressure (BP) values, systolic not measured.
	Martin et al. 1973	**196**	Significant increase in systolic BP, no consistent significant effect on diastolic BP.
	Badaruddoza 2004.	**197**	Significant increase in systolic and diastolic blood pressure.
Body mass index (height, weight)	Krieger 1968	**195**	In 3,465 children, no significant effects on height or weight, increase in body mass index
	Martin et al. 1973	**196**	In 12 age-gender defined samples ranging from 166 to 657 subjects, increase in weight, decrease in height, increase in body mass index
	Badaruddoza 2004.	**197**	In 3,253 children, decrease in height and weight, increase in body mass index
Cholesterol levels	Martin et al. 1973	**196**	Decrease in population <20 years, large increase in population >=40 years

1.2.4.2. COMPLEX LATE-ONSET DISEASES

Table 1-3 summarizes the effects of inbreeding on late-onset complex chronic diseases. It seems that only 15 studies provided direct information on the relationship between inbreeding and adult-age diseases to date (**205-219**). Study designs included an ecological comparison between more and less inbred populations in 4 studies, a population-based case-control design in 7 studies, a hospital-based case-control design in 3 cases, and self-reporting for disease and inbreeding status by questionnaire in 1 study. Only three studies did not detect possible effects of inbreeding on the diseases under study, but those were also the ones using a study design prone to multiple confounding effects. In the studies by Jaber et al. (**207**) and Denic and Bener (**212**), self-reporting of inbreeding status could have been a likely cause of misclassification. Also, in the study by Saugstad and Odegard in Norway (**214**) the authors reported that a change of diagnostic practice for psychiatric illnesses (especially schizophrenia) during the period of study could have acted as a major confounding effect. All other studies report inbreeding effects on many late-onset diseases, and the estimated or projected risks were considerable. It appeared that in the inbred communities, individuals or affected cases the relative risk ranged from 1.5 to 4.5 when compared to non-inbred controls (**205,206,208-211,213,215-219**).

Table 1-3: Review of the studies investigating the effect of inbreeding on complex late-onset diseases. (*These studies are based on self reported consanguinity, which in our experience tends to be of low reliability, as most "inbreeding loops" are found in the third and fourth parental generations and thus unknown to most people).

Disease	Author/Reference	Study Design	Reported effect of inbreeding
Coronary heart disease	Shami et al. (**205**)	Case-control study Hospital-based	Hospital cases had significantly greater inbreeding coefficients than controls (individuals in population from which they were recruited)
	Puzyrev et al. (**206**)	Ecological study	Increased risk of myocardial ischaemia

			among endogamous males
	Jaber et al.* (**207**)	Case-control study (self-reported exposure and outcome status)	No difference in self reported prevalence among students with or without consanguineous parents
	Ismail et al. (**208**)	Case-control study Hospital-based	Increased risk (OR 3.8) of myocardial infarction South Asians under 45 years related to parental consanguinity
Cancer	Simpson et al. (**209**)	Case-control study Population-based	Increased risk of breast, endometrial and ovarian cancer associated with greater inbreeding coefficient in women under 45 years
	Lebel & Gallagher (**210**)	Case-control study Population-based	Increased risk associated with greater inbreeding coefficient, (especially for multiple and early-onset cancers)
	Shami et al. (**205**)	Case-control study Hospital-based	Hospital cases had significantly greater inbreeding coefficients than controls (individuals in population from which they were recruited)
	Rudan (**211**)	Ecological study	Stepwise increase of 20-year cancer incidence associated with greater inbreeding in 5 island communities
	Denic & Bener * (**212**)	Case-control study Population-based (self-reported exposure status)	Decreased risk of breast cancer among women who self-reported being offspring from consanguineous unions, no effect on cervical cancer
Psychi-atric Disorders	Abaskuliev & Skoblo (**213**)	Case-control study Population-based	Increased frequency of consanguinity among parents of schizophrenia cases
	Saugstad & Odegard (**214**)	Case-control study Population-based	No increase in first-cousin matings among parents of psychiatric patients (major changes in diagnostic criteria

			over time reported by authors as important confounder)
	Gindilis et al. (**215**)	Ecological study	Severe schizophrenia 2-3 times more prevalent in inbred communities
Alzheimers Disease	Vezina et al. (**216**)	Case-control study Hospital-based	205 autopsy-confirmed cases of late-onset Alzheimer's disease significantly more inbred than controls
Multiple sclerosis	Roberts et al. (**217**)	Case-control population-based (with parallel ecological study)	Average inbreeding coefficient greater among cases than controls; increased prevalence in genetic isolate population)
	Roberts et al. (**218**)	Case-control population-based (with parallel ecological study)	Average inbreeding coefficient greater among cases than controls; increased prevalence in genetic isolate population)
Type 2 diabetes	Jaber et al.* (**207**)	Case-control study (self-reported exposure and outcome status)	No difference in self reported prevalence among students with or without consanguineous parents
Gout	Ombra et al. (**219**)	Ecological study	Increased prevalence of hyperuricemia and uric stones in a highly inbred community
Asthma	Jaber et al.* (**207**)	Case-control study (self-reported exposure and outcome status)	No difference in self reported prevalence among students with or without consanguineous parents
Peptic ulcer	Jaber et al.* (**207**)	Case-control study (self-reported exposure and outcome status)	No difference in self reported prevalence among students with or without consanguineous parents

1.2.5. Inbreeding Studies can Improve Understanding of Genetic Architecture of Late-Onset Complex Traits: Theoretical Considerations

It is widely accepted that one of the main reasons why genetic association studies of late onset complex diseases did not yield more success is a relatively poor understanding of the genetic architecture underlying susceptibility to those diseases. Many studies have been based on assumptions of the number of susceptibility alleles underlying those diseases, their population frequency, effect size, penetrance, dominance/recessivity and the population frequency of underlying haplotypes. However, it is fair to say that none of those parameters is currently known with any degree of certainty, which is why most of the studies showing adequate power to detect disease susceptibility alleles under a set of assumptions eventually failed. Obtaining any insight into the parameters of the genetic architecture of complex human traits and diseases, such as the number of susceptibility alleles underlying those diseases and the distribution of their effect size, would therefore be very valuable. This information would help in improving the design of future studies into the genetic basis of complex late-onset diseases.

Late-onset complex diseases can be considered to result when a threshold of quantitatively varying risk or liability is exceeded (**3,188**). Liability is thought to result from the net effect of many quantitative traits that are influenced by genes and the environment. These genetic effects can be described collectively by analysing the components of genetic variance. Those components can be estimated based on a measurement of resemblance between relatives for disease or a quantitative trait. The total genetic variance (V_G) of a complex trait can be partitioned into its components (**3,188**): (a) additive genetic variance (V_A), the component of variance due to genetic effects that are directly transmissible from parent to offspring; (b) dominance variance (V_D), the component of variance due to interactions (departures from additivity of effects) between alleles at the same locus; and (c) epistatic variance (V_I), the component of variance due to interactions between alleles at different loci (**3,188**). The total phenotypic variance (V_P) in the trait is the sum of V_G and any non-

genetic effects (V_E), together with the effects of interactions between genotype and environment (V_{GE}). The heritability is the ratio of V_A to V_P. These components can be estimated from correlations between relatives, such as parents and offspring, and full–sibs. In practice, it is difficult to separate V_D and V_I, and they are often treated as a single component of non-additive variance. For example, fitness-related traits generally show substantial additive variance; whereas relatively few show non-additive variance. Late-onset traits are predicted to show higher values of V_A and V_D. (**113, 220**). It is, however, very difficult in practice to disentangle environmental from genetic variance using nuclear family information to compute heritabilities, as nuclear families share common environments. This usually results in an over-estimation of genetic effects.

Therefore, one of the rare feasible approaches to obtain information on genetic variation influencing complex late-onset traits and diseases is to measure the effects of inbreeding. Inbreeding can contribute to disease prevalence in a population and to the inbreeding depression of complex quantitative traits (**13,221**). This results from increased homozygosity of trait alleles, which either show recessive effects on the trait acting in the same direction (directional dominance), or show heterozygous advantage. The value B, which is the negative of the regression coefficient of trait mean on inbreeding coefficient, F, provides a useful summary statistic for the genetic damage that would occur if all deleterious recessives were made homozygous (equals to $F = 100\%$). When fitness components, such as survival to adulthood, are measured on a scale of natural logarithms, B provides a measure of the deleterious effect of complete homozygosity, often called the "inbreeding load".

Theory shows that the value of B due to deleterious mutations depends only on the genomic mutation rate and the dominance of individual mutations (**221,222**). If all mutations have the same effects, B for a given trait that is positively correlated with fitness is:

$$B = U\{(1/h) - 2\}\alpha \qquad \text{(Eq. 1)}$$

where U is the mutation rate per diploid individual to deleterious alleles affecting the trait; h is the extent to which fitness is reduced in heterozygous mutation, relative to its effect in homozygotes; and α is a constant of proportionality relating the effect of a mutant allele on the trait in question to its effect on fitness. Relating the measurement of variance components and inbreeding effects to the predictions of models of the maintenance of genetic variation provides an important means of testing the models, which will be one of the main aims of this thesis (**12,221,222**).

Apart from providing information on genetic variation influencing disease through testing the above model based on genetic load computations, inbreeding studies can be useful for the estimation of the number of genetic loci influencing a complex trait or disease. In theory, this can also be achieved through two other aproaches: (a) to cross inbred or natural populations that show a large difference in trait value and estimate the number of "equivalent effect" genes accounting for the difference on the basis of the trait variances in the hybrid offspring (F1, F2 and backcross generations); (b) to undertake systematic genetic mapping of trait loci distinguishing a pair of lines, yielding a direct estimation of gene numbers. In practice, however, those estimates often have large sampling variances and require assumptions that are not always met. Neither of those two methods provides a direct estimate of the number of loci segregating within a population. If the lines concerned have been derived from a population, they at least provide a lower bound to the number of such loci (**12,223**).

However, another method based on studying inbreeding effects can be applied directly to population data and is useful for estimating the number of recessive or partially recessive loci contributing to a trait that shows inbreeding depression (**222**). The effect of inbreeding, B, and the dominance variance, V_D, are measured as suggested earlier in the text (Eq. 1). Then, a lower bound to the number of genes, n, affecting the trait is provided by:

$$n \geq B^2 / V_D \qquad \text{(Eq. 2)}$$

To conclude, properly designed inbreeding studies can be used to assess both genetic variation influencing complex traits and late-onset disease susceptibility, and to estimate the minimum number of loci influencing those traits.

2. EXAMINEES AND METHODS

2.1. STUDY POPULATION: THE CROATIAN ISLAND ISOLATE RESOURCE

In this part of the thesis, the study population (Croatian island isolate resource) will be described and its apparent advantages for the analyses of inbreeding effects on human traits and diseases listed. In further sections, the methods of recruitment of the examinees, determination of their individual inbreeding coefficients (risk exposure) and measurements of quantitative traits of interest and diagnosis of diseases will be presented, along with study designs and statistical methods of data analyses. As the study population, study designs and methodological approaches were chosen in order to maximise the potential of analysis of the effect on each particular trait or disease, the methods will be presented accordingly, by studied traits / diseases.

Croatia has 15 Adriatic Sea islands with a population greater than 1,000. The villages on the islands have unique population histories and have preserved their isolation from other villages and the outside world through many centuries. The history, demography and genetic structure of those villages have been investigated for more than 50 years. The research, mainly carried out by the Institute for Anthropological Research in Zagreb, Croatia, resulted in over 100 publications in international journals (**89,224,225**). **Figure 2-1** shows the geographic location of the main inhabited islands that will be mentioned in this research. **Table 2-1** provides a comprehensive review of the reference sources with detailed information on population genetic structure and variation, monogenic (Mendelian) diseases and rare genetic variants identified in particular islands and reported in the literature (**226-252**).

It can be noted that there is continuity in the research of these unique populations. The initial investigations of population structure were performed using classic genetic polymorphisms such as blood group antigens (**238-241**), HLA markers

and immunoglobulin allotypes (**234-237**). These findings have been later supported by more informative analyses, using STR and VNTR polymorphisms (**226-230**), and more recently mtDNA and Y-chromosome based analyses (**231-233**). At least seven rare Mendelian diseases characteristic for particular islands have been reported (**242-249**), which are believed to be due to a combination of founder effect (increasing population frequency of rare mutations in an isolate) and subsequent inbreeding (which increases the probability of the recessive rare variants becoming homozygous). Reports of high population frequencies of several extremely rare genetic variants found in these islands support this explanation (**231, 250-252**).

Figure 2-2 presents a brief overview of ethnic history of Croatian island populations. The earliest inhabitants known with any degree of certainty are thought to be proto-Illyrians, inhabiting the islands around 2,000 B.C. During the B.C. era, they were succeeded by Illyrians, who admixed with Greeks and succeeding Romans to form the population "sub-stratum". In 7th century AD the first Croats immigrated into the area and admixed with the other "*sub-stratum*" populations. The populations then remained relatively isolated for centuries, adopting the "*ča*" dialect of Croatian language, until the next immigration wave ("*super-stratum*") occurred during medieval period, with arriving immigrants who fled the mainland in fear of Ottoman (Turkish) expansions and brought the "*što*" dialect (**89,224,225,253**). The present-day population structure, "*ad-stratum*" was then formed by very limited population movements from 1800 onwards, which ensured gene flow between the settlements, and further immigration from the mainland on a small scale.

Table 2-1: Brief review of population genetic research undertaken in Croatian island isolates in the past 25 years.

TYPE OF RESEARCH	ISLAND	REF.
A. Studies of population genetic variation		
STR polymorphysms	Krk, Brac, Hvar, Korcula	**226-229**
VNTR polymorphysms	Hvar	**230**
Y-chromosome haplogroups	Krk, Brac, Hvar, Korcula	**231**
MtDNA haplogroups	Krk, Brac, Hvar, Korcula	**232,233**
HLA markers or Immunoglobuline allotypes	Krk, Hvar, Silba, Olib, Pag	**234-237**
Serogenetic polymorphisms	Brac, Hvar, Korcula, Silba, Olib,	**238-241**
B. Reports on autochtonous Mendelian diseases		
Dwarfism	Krk	**242-244**
Albinism	Krk	**244**
Progressive spastic quadriplegia	Krk	**244**
Familial cognitive dysfunction	Susak	**245**
Familial congenital hip dislocation	Lastovo	**246**
Familial ovarian cancer	Lastovo	**247,248**
Keratoderma palmoplantaris transgrediens	Mljet	**249**
C. Reports of high population frequencies of extremely rare genetic variants		
Deleted/triplicated alpha-globin gene	Silba	**250**
PGM1*W3 phosphoglucomutase-1 variant	Olib	**251**
MtDNA haplogroup F	Hvar	**252**
Y-chromosome haplogroup P*	Hvar	**231**

Figure 2-1: Geographic location of the island populations that were studied to investigate inbreeding effects on various quantitative traits and late-onset diseases in this research.

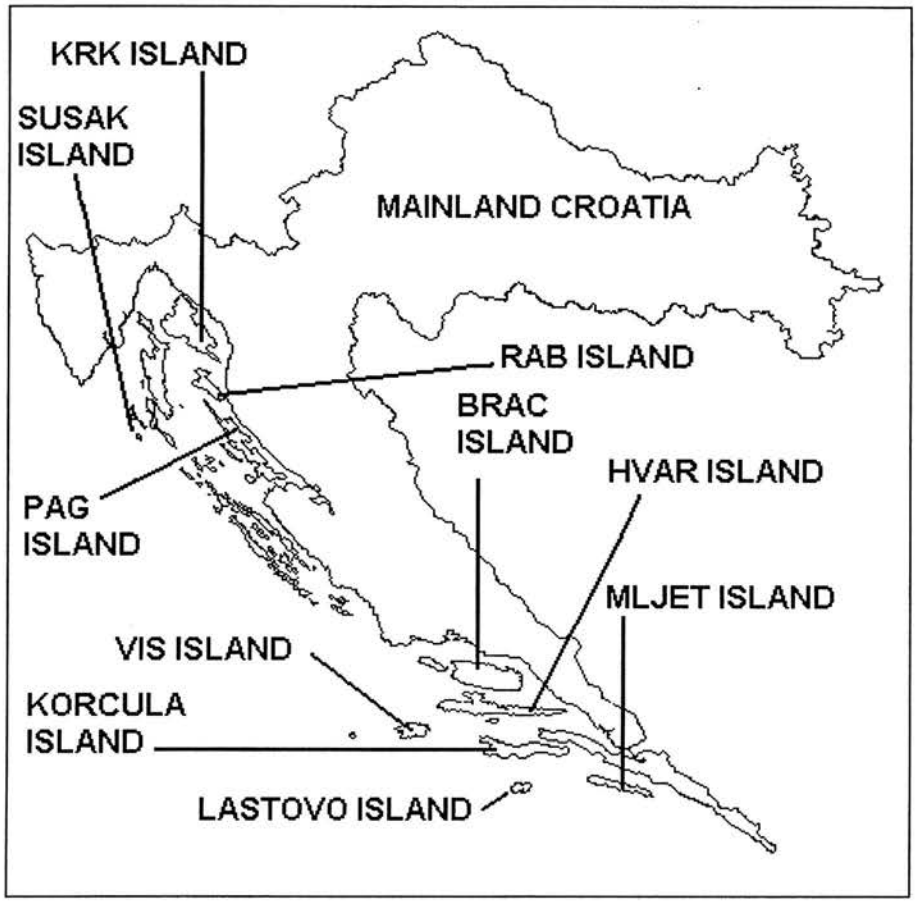
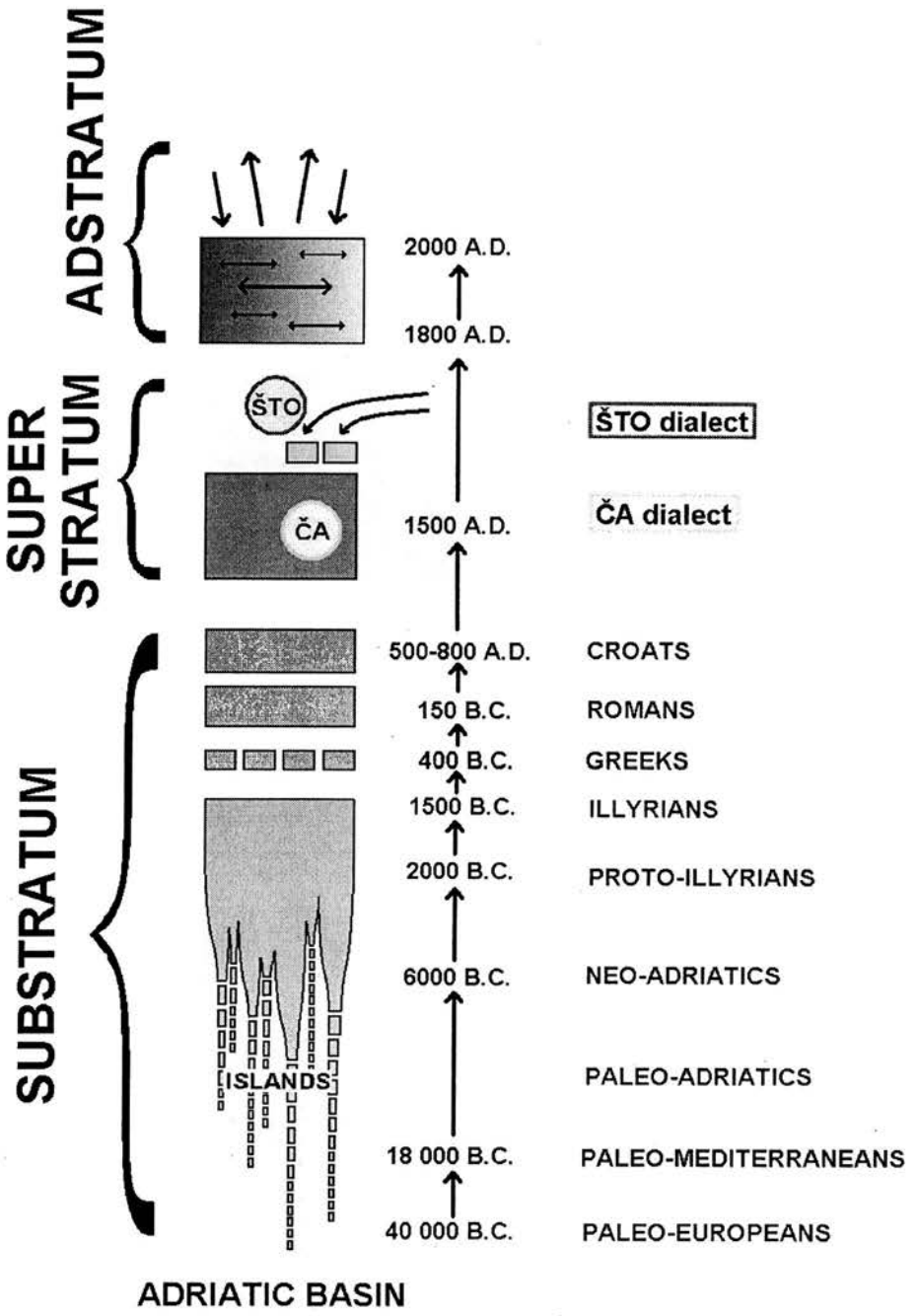


Figure 2-2: Brief schematic presentation of long-term population history of island populations in middle Dalmatia, Croatia.



The tendency towards inbreeding in each village has been influenced by several factors. The first one was geographic isolation, which limited mate choice in the villages leading to frequent marriages between relatives. Even nowadays, in most of these settlements the levels of parental and grand-parental endogamy are among the highest reported for human populations **(89,224,225)**. Other important factor were political (*"Pastrovic")* privileges given to residents of certain communities for serving the Venetian Republic. These privileges introduced large social differences and contributed to subdivision of the populations on particular islands. Finally, inbreeding was also favoured due to purely socio-cultural reasons (**89,224,222**), as in times of great poverty and very limited resources the individuals on the islands were forced into consanguineous unions in order to preserve the land and resources within the families. Therefore, the Croatian island populations present a range of inbreeding patterns at both individual and subpopulation levels, as documented in previous studies reporting endogamy, isonymy, mating choice, genealogical information, and genetic marker distributions (**254-257**).

The relevance of the study population for the research undertaken in this thesis will be discussed in detail in the following sections, as study designs, samples and methods differed and were tailored to appropriately address each specific research question. The author's specific role in data collection processes in these populations was explained in detail in the Foreword of the thesis, and partly also in the acknowledgement.

2.2. INVESTIGATION OF THE EFFECTS OF INBREEDING ON BLOOD PRESSURE

2.2.1. Sample Selection

Blood pressure, height and weight were measured in the late 1970s and early 1980s in 2,760 adult individuals selected at random from voting lists from 25 isolate villages on the 3 islands (Brac, Hvar and Korcula) in middle Dalmatia, Croatia. The detailed geographic location of all 25 villages is presented in **Figure 2-2**. The recruited sample represented about 20% of the total village populations. In addition, data were collected on body mass index, diet, education level, occupation and smoking status. This was carried out with the informed consent of participants by the Institute for Anthropological Research in Zagreb, Croatia in collaboration with the Smithsonian Institute in Washington, USA. None of the examinees had ever received antihypertensive treatment.

2.2.2. Computation of Individual Inbreeding Coefficients in 25 Villages

A single researcher (author of the PhD thesis) computed individual inbreeding coefficients independently and blind to BP status for each study participant. The computation of the inbreeding coefficient was initially based on pedigree information on 2-3 ancestral generations collected during the initial field work in 1970s and 1980s. These genealogical data were expanded by a study of parish registries stored in local churches during 1997–2000, to 4 ancestral generations for each individual (five generations where these occurred over a similar timeframe). The coefficients (F) were computed according to Wright's path method (**258**) and gave values ranging from zero to 0.125. In addition, average inbreeding measures for each of the 25 villages based on

isonymy were calculated, as it is widely accepted that those estimates provide an upper limit to the true inbreeding estimate in a population (**259,254,255**).

2.2.3. Blood Pressure Measurements

Blood pressure (BP) was measured by a single observer in local health centres and dispensaries between 6 am and 12 noon. Blood pressure was measured on the right forearm in a sitting position. Two measurements of both systolic and diastolic blood pressure were taken 5 minutes apart in each individual following standard procedures (**260**). Hypertension was defined as systolic BP $\geq$ 160 or diastolic BP $\geq$ 95 mmHg. Individual BP values were adjusted for the major determinants of BP: age, height, weight and smoking status in the analyses and reported separately in males and females, before the relationship with inbreeding status was investigated. The basic distribution of all 2,760 examinees by village (and island) of residence, gender, mean value of inbreeding coefficient in each village (estimated from genealogies and isonymy) and body mass index and environmental covariates is presented in **Table 2-2**.

Figure 2-2 (next page): Exact location of 25 villages from the islands of Brac, Hvar and Korcula in which the sample of 2,760 adult individuals previously untreated for hypertension was recruited during the field work in 1970s and 1980s.

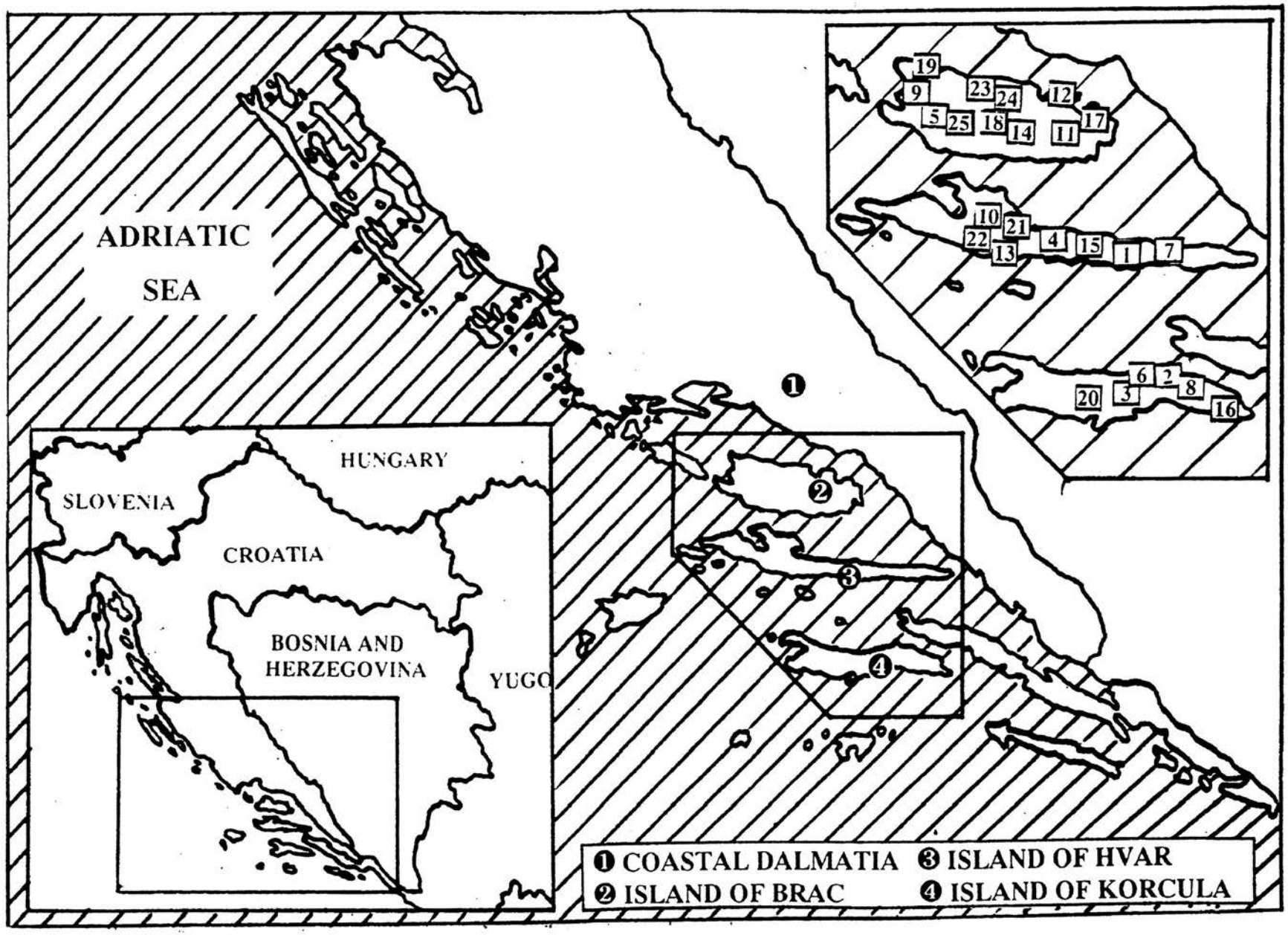

ADRIATIC
SEA
HUNGARY
SLOVENIA
CROATIA
BOSNIA AND
HERZEGOVINA
YUGO
19
9
23
24
12
5
25
18
14
11
17
10
21
22
13
4
15
1
7
6
2
8
20
3
16
❶ COASTAL DALMATIA
❷ ISLAND OF BRAC
❸ ISLAND OF HVAR
❹ ISLAND OF KORCULA

Table 2-2: The basic distribution of all 2,760 examinees by village (and island) of residence, gender, mean value of inbreeding coefficient in each village (estimated from genealogies and isonymy), body mass index. Environmental covariates are presented as proportion with university degree (EDU), employed in agriculture / fishery (OCU), consuming Mediterranean diet (DIE), and smokers (SMO).

VILLAGE	Sex	N	AGE (x (SD))	*F* (gen)	*F* (iso)	BMI (x (SD))	EDU (%)	OCU (%)	DIE (%)	SMO (%)
1. Gdinj (H)	M	54	53.6 (10.9)	0.048	0.107	24.4 (3.3)	3.7	92.6	94.4	9.3
	F	76	53.1 (11.7)	0.049		25.6 (4.7)	1.3	72.4	93.4	1.3
2. Pupnat (K)	M	46	42.8 (12.9)	0.046	0.034	27.3 (3.6)	2.2	89.1	95.7	13.0
	F	51	44.4 (11.8)	0.043		26.1 (3.7)	2.0	90.2	88.2	5.9
3. Cara (K)	M	63	44.3 (11.7)	0.035	0.040	26.1 (3.7)	3.2	90.5	95.2	36.5
	F	74	45.6 (12.5)	0.030		26.6 (3.8)	0.0	87.8	94.6	10.8
4. Poljica (H)	M	28	42.3 (9.1)	0.032	0.037	25.5 (2.7)	0.0	89.3	92.9	10.7
	F	20	43.0 (7.6)	0.032		25.4 (4.2)	0.0	85.0	85.0	0.0
5. Dracevica (B)	M	20	46.8 (16.4)	0.026	0.031	27.0 (4.2)	5.0	90.0	90.0	35.0
	F	22	50.9 (14.2)	0.036		30.0 (4.3)	0.0	81.8	90.9	13.6
6. Racisce (K)	M	40	43.6 (11.2)	0.029	0.034	27.8 (4.0)	2.5	82.5	92.5	25.0
	F	64	43.2 (11.8)	0.026		27.0 (4.1)	1.6	90.7	78.1	21.9
7. Bogomolje (H)	M	48	57.6 (15.9)	0.025	0.030	22.4 (2.6)	2.1	87.5	95.8	8.3
	F	33	57.7 (13.4)	0.025		23.9 (4.1)	3.0	78.8	90.9	0.0
8. Zrnovo (K)	M	98	45.0 (13.0)	0.022	0.029	26.3 (2.9)	5.1	79.6	92.3	42.9
	F	104	45.0 (13.9)	0.019		26.8 (4.3)	2.9	78.9	85.6	19.2
9. Lozisca (B)	M	17	56.9 (12.2)	0.022	0.029	26.3 (4.3)	5.9	82.4	88.2	35.3
	F	55	49.9 (14.1)	0.018		28.7 (4.4)	1.8	72.7	90.9	9.1
10. Vrbanj (H)	M	65	38.2 (10.4)	0.020	0.010	26.5 (2.7)	4.6	90.8	95.4	10.8
	F	53	42.3 (9.6)	0.018		26.8 (3.9)	1.9	86.8	88.7	1.9
11. Selca (B)	M	117	54.4 (14.4)	0.016	0.026	26.8 (3.3)	4.3	78.6	89.7	41.9
	F	185	52.0 (14.9)	0.016		27.7 (4.4)	2.2	76.8	90.8	17.8
12. Povlja (B)	M	25	56.3 (14.7)	0.014	0.024	27.6 (3.2)	0.0	84.0	92.6	40.0
	F	47	54.1 (14.3)	0.017		27.5 (4.8)	2.1	85.1	87.2	10.6
13. Vrisnik (H)	M	52	38.0 (10.4)	0.012	0.023	24.6 (2.6)	3.8	90.4	94.2	19.2
	F	44	42.4 (9.6)	0.019		26.5 (3.6)	0.0	72.7	88.6	6.8

14. Gornji Humac (B)	M	32	51.0 (12.5)	0.014	0.016	25.7 (3.5)	3.1	87.5	90.6	59.4
	F	48	47.4 (13.6)	0.012		28.3 (4.5)	2.1	77.1	85.4	25.0
15. Zastrazisce (H)	M	70	49.6 (14.8)	0.013	0.013	24.7 (2.8)	4.3	85.7	94.3	12.9
	F	54	50.7 (12.1)	0.013		25.6 (3.9)	3.7	87.0	92.6	0.0
16. Lumbarda (K)	M	54	43.4 (11.3)	0.014	0.025	26.8 (2.7)	7.4	85.1	90.7	37.0
	F	54	43.0 (12.4)	0.010		25.6 (3.2)	1.9	85.2	88.9	5.6
17. Novo Selo (B)	M	27	46.6 (16.6)	0.011	0.014	26.6 (3.9)	0.0	85.2	88.8	33.3
	F	30	54.6 (13.3)	0.009		28.5 (4.3)	0.0	80.0	86.7	16.7
18. Praznice (B)	M	53	48.7 (13.2)	0.012	0.016	27.2 (3.8)	3.8	88.7	94.3	45.3
	F	67	53.0 (14.6)	0.007		28.9 (4.5)	1.5	82.1	89.5	10.4
19. Sutivan (B)	M	61	50.2 (15.3)	0.009	0.012	27.9 (3.7)	6.6	72.1	90.2	27.9
	F	88	50.5 (15.1)	0.009		28.0 (4.6)	3.4	68.1	89.8	23.9
20. Smokvica (K)	M	52	43.2 (10.5)	0.009	0.019	25.8 (3.0)	3.8	84.6	90.4	42.3
	F	45	45.7 (12.5)	0.008		26.2 (3.6)	2.2	77.8	71.1	11.1
21. Svirce (H)	M	74	40.2 (11.0)	0.008	0.008	25.6 (2.5)	2.7	91.2	93.2	18.9
	F	75	38.5 (10.9)	0.007		25.8 (3.5)	1.3	69.3	73.3	6.7
22. Dol Hvarski (H)	M	53	41.1 (10.7)	0.005	0.004	26.5 (2.5)	5.7	79.2	94.3	13.2
	F	42	42.4 (9.9)	0.005		25.6 (3.9)	2.4	71.4	76.2	2.4
23. Skrip (B)	M	47	51.4 (10.8)	0.003	0.008	27.7 (3.2)	2.1	85.1	93.6	31.9
	F	56	49.8 (14.5)	0.005		29.6 (4.7)	1.8	83.9	87.5	16.1
24. Dol Bracki (B)	M	41	44.7 (15.3)	0.003	0.005	28.0 (3.2)	2.4	85.4	92.7	29.3
	F	24	41.8 (15.5)	0.001		27.8 (4.3)	0.0	79.2	75.0	25.0
25. Nerezisca (B)	M	50	44.7 (12.9)	0.002	0.004	27.3 (3.4)	4.0	66.0	90.0	54.0
	F	62	44.9 (11.9)	0.002		28.5 (5.0)	0.0	56.5	71.0	19.4

2.2.4. Statistical Analyses

Comparisons of BP among *villages* were based on systolic and diastolic BP measurements adjusted for age, body mass index (weight / height^2) and smoking status. A step-down multiple regression analysis was performed using MINITAB 12.21 software to investigate the correlation between *individual* BP measurements and inbreeding coefficients. The model explored the relationship between systolic and diastolic blood pressure (as dependent variables) and a number of explanatory variables:

individual inbreeding coefficient (F), island and village of residence, smoking status and the major known risk factors for hypertension - age, sex, (log transformed) height and (log transformed) weight. Variables which made the least contribution to the explained variation were dropped one at a time until all the remaining variables were statistically significant (defined as $P<0.05$ for main effects and $P<0.01$ for higher order effects).

A lower bound, n_L, for the number of genetic loci of equivalent effect contributing to the variability in BP, was calculated as

$$n_L = D_T^2 / V_G = D_T^2 / (H^2 \times V_p)$$

where D_T is the overall slope of the regression on F, V_G is the total genetic variance, V_P is the residual phenotypic variance and H^2 is the broad-sense heritability. To correct for possible unobserved background inbreeding preceding the earliest generation of which we had knowledge, we inflated F-values by a factor equal to the ratio of the median of village inbreeding levels based on isonymy methods to the median based on pedigree methods. The above statistical analyses were planned, designed and performed under the leadership and in collaboration with Dr Andrew Carothers from the Human Genetics Unit, Medical Research Council, Edinburgh, UK, and then repeated under his guidance by the author of this thesis.

The population attributable fraction (PAF) was calculated by multiple logistic regression allowing for individual differences in the variables: village, sex, age, height, weight and smoking. The appropriate regression was determined as a function of all associated variables (including F). After that, each individual's probability of being hypertensive if their F was set equal to 0 was noted. The sum of all such probabilities, P_{sum}, is an estimate of the number affected in the absence of inbreeding, but with other variables remaining unaltered. Then, population attributable fraction (PAF) is computed as: $PAF = 1 - P_{sum}/N_{pop}$, where N_{pop} is the total population size.

2.3. INVESTIGATION OF THE EFFECTS OF INBREEDING ON CORTICAL INDEX AND OSTEOPOROSIS

2.3.1. Sample Selection

During the field research between 1979 and 1981, undertaken by the Institute for Anthropological Research in Zagreb, Croatia, the measurements of cortical index were performed in 14 villages: Dracevica, G. Humac and Nerezisca (Brac island), Gdinj, Bogomolje, Vrisnik, Zastrazisce, Svirce and Dol (Hvar island), and Pupnat, Cara, Racisce, Lumbarda and Smokvica (Korcula island). The sample included 1,389 adult individuals (682 males and 707 females). They were selected randomly from voting lists to form approximately 20-30% of the total population of these 14 villages. The sample represents the subset of 14 (of 25) villages presented in **Table 2-2**. The collected data characterizing environmental variation included the proportion of inhabitants with some college education (EDU), occupation in agriculture and fishery (OCU), regular consumption of traditional Mediterranean diet (DIE), smoking habits (SMO) and measurement of body mass index (BMI). As this population cohort is a sub-sample of that presented in detail in **Table 2-2**, all the relevant epidemiological parameters presented there are also reflected in this sub-sample.

Table 2-3 supports the hypothesis of decreased variation in most of the studied characteristics related to environment. This is important, as socio-economic status, occupation, diet, obesity and climate are usually highlighted as potential environmental risk factors for many common late onset diseases.

Table 2-3: Genetic and environmental characteristics of 14 village populations (*N(sam), N(pop)*) ranked according to average inbreeding coefficient computed from genealogical data (*F(gen)*). Also presented: average inbreeding coefficients computed from isonymy (*F(iso)*) and serogenetic polymorphisms (*F(sgp)*), proportion of examinees with some college education (*EDU*), working in agriculture and fishery (*OCU*), regularly consuming traditional mediterranean diet (*DIE*) and smoking (*SMO*), and the average body mass index (*BMI*).

VILLAGE*	N¶ (sam)	N# (pop)	*F* (gen)	*F* (iso)	BMI (avg)	EDU (%)	OCU (%)	DIE (%)	SMO (%)
Gdinj (H)	135	153	0.049	0.107	25.1	2.3	80.8	93.8	4.6
Pupnat (K)	96	326	0.044	0.034	26.7	2.1	89.7	91.8	9.3
Cara (K)	139	464	0.032	0.040	26.4	1.5	89.0	94.9	22.6
Dracevica (B)	20	66	0.031	0.031	28.6	2.4	85.7	90.5	23.8
Racisce (K)	103	290	0.027	0.034	27.3	1.9	87.5	83.6	23.1
Bogomolje (H)	90	102	0.025	0.030	23.0	2.5	84.0	93.8	4.9
Vrisnik (H)	105	216	0.015	0.023	25.5	2.1	82.3	91.6	13.5
G. Humac (B)	43	214	0.013	0.016	27.3	2.5	81.3	87.5	38.8
Zastrazisce (H)	133	210	0.013	0.013	25.1	4.0	86.3	93.6	5.9
Lumbarda (K)	59	354	0.012	0.025	26.2	4.7	85.2	89.8	21.3
Smokvica (K)	97	866	0.008	0.019	26.0	3.1	81.4	81.4	27.8
Svirce (H)	152	375	0.008	0.008	25.7	2.0	80.1	83.2	12.8
Dol (H)	102	154	0.005	0.004	26.1	4.2	75.8	86.3	8.4
Nerezisca (B)	65	106	0.002	0.004	28.0	1.8	60.7	79.5	34.8

* distributions by sex and age not shown as later results were standardized by those two variables;

¶ sample obtained between 1979-1981 as random 20-30% of the total village populations;

total village populations in the year 2000, in which disease prevalence was determined;

§ blood sample for the analysis was obtained from 88.8% of the examinees.

2.3.2. Computation of Individual Inbreeding Coefficients in 14 Villages

This has been done using pedigree information on 4 ancestral generations (five generations where these occurred over a similar timeframe), recorded during the initial field work and supplemented by a study of parish registries stored in local churches during 1997-2000, as described in Section 2.2.2. The individual inbreeding coefficients (F) were then computed according to Wright's path method (**258**):

$$F = \Sigma_{(1\rightarrow c)}(1/2)^{(n_i+m_i+1)}$$

where m and n refer to the number of paths from a common ancestor, and c refers to the number of common ancestors. The genealogical inbreeding coefficient for each village was then computed as the average of all individual F values. To further support these estimates, F was calculated from isonymy as suggested by Tay and Yip (**259**), and mean values were derived for each of the 14 villages. Estimates based on isonymy are generally thought to be positively biased, and so to provide an upper bound for true F values.

2.3.3. Cortical Index Measurements and the Definition of Osteoporosis

The osteometric dimensions of metacarpal bones are an efficient and practical method for the investigation and monitoring of bone mass. At the time when data used in the present analysis was gathered, the radiological measurements of the metacarpal bones was widely used as the best available screening method of bone status in population studies. The procedure of osteometry of metacarpal bones, as thoroughly described by Barnett and Nordin (**261**), was followed in field studies performed on all

three investigated islands. Hand-wrist radiographs were taken using a single portable X-ray. Total diaphysis width (D) and medullary canal width (d) of the second left metacarpal bone was determined by a single, experienced observer (**261, 262**). Measurements were performed by one investigator using a millimeter ruler and a magnifying glass (x10) with a scale permitting 0.05-mm accuracy. Measurements were rounded to 0.1 mm. For each individual and for each bone, the cortical index (CI) was computed as:

$$CI = (D-d) * 100 / D.$$

Table 2-4. Descriptive statistics of age and cortical index by village in male examinees (see Table 2-3 for comparative inbreeding levels in villages under study).

		Age				Cortical index			
Village	N	X	SD	Min.	Max.	X	SD	Min.	Max.
Gdinj (H)	51	54.53	11.72	24	81	55.97	7.89	37.78	72.22
Pupnat (K)	46	42.76	12.91	23	71	55.17	6.84	44.85	68.35
Čara (K)	63	42.59	11.87	20	71	53.11	8.42	32.37	72.07
Dračevica (B)	20	46.80	16.37	23	74	66.28	7.15	47.62	80.00
Račišće (K)	40	43.30	11.22	20	62	55.14	7.30	36.23	73.11
Bogomolje (H)	46	57.85	15.65	24	81	48.90	6.17	32.14	64.44
Vrisnik (H)	44	37.68	10.52	23	54	57.22	6.77	42.11	69.89
G. Humac (B)	27	51.04	16.81	22	85	62.87	6.92	49.49	76.14
Zastražišće (H)	69	50.09	14.78	20	77	53.41	7.45	32.43	76.09
Lumbarda (K)	53	43.21	11.36	22	77	59.70	7.60	41.09	77.83
Smokvica (K)	52	40.87	10.53	24	59	53.46	7.36	37.75	69.41
Svirče (H)	70	39.94	11.10	21	55	59.36	9.78	39.81	82.76
Dol (H)	50	41.44	10.37	20	56	56.73	6.15	40.00	70.51
Nerežišća (B)	51	45.59	13.86	22	81	62.31	9.07	44.12	82.02
Total	682	45.25	13.73	20	85	56.50	8.59	32.14	82.76

Table 2-5. Descriptive statistics of age and cortical index by village in female examinees (see Table 2-3 for comparative inbreeding levels in villages under study).

		Age				Cortical index			
Village	N	X	SD	Min.	Max.	X	SD	Min.	Max.
Gdinj (H)	71	53.48	12.28	27	82	55.64	9.99	37.21	84.15
Pupnat (K)	50	44.76	11.74	21	61	56.84	8.67	34.53	75.00
Čara (K)	75	43.63	12.38	20	63	56.29	8.74	37.63	83.09
Dračevica (B)	23	52.48	15.82	21	87	65.11	12.46	42.22	92.41
Račišće (K)	63	43.29	11.83	21	71	58.22	8.58	40.18	77.14
Bogomolje (H)	32	58.50	15.62	19	78	50.44	9.89	29.35	75.00
Vrisnik (H)	28	41.43	8.69	23	54	61.20	9.28	43.37	76.40
G. Humac (B)	44	49.09	13.81	25	82	64.93	11.24	39.76	91.21
Zastražišće (H)	54	52.37	12.93	24	83	54.46	9.41	37.76	83.33
Lumbarda (K)	55	43.09	12.27	22	74	59.58	9.94	37.70	80.00
Smokvica (K)	42	44.48	12.41	23	61	58.28	11.17	32.08	82.45
Svirče (H)	65	39.25	10.88	20	55	63.90	9.81	40.26	87.18
Dol (H)	43	41.86	10.08	22	56	63.30	9.29	46.07	84.00
Nerežišća (B)	62	44.76	12.15	19	77	65.02	10.82	37.21	91.55
Total	707	46.15	13.21	19	87	59.31	10.60	29.35	92.41

The cortical index is a useful and precise measure that could be used as a predictor of osteoporosis development. It measures the thickness of the bone mass by analyzing the relationship between total diaphysis width and the width of the medullary canal in which there is no bone mass. **Tables 2-4 and 2-5** show the sample sizes and descriptive statistics of age and cortical index by village in male and female examinees. As expected, a larger effect of age on cortical index and on osteoporosis prevalence is observed in females than in males (**Figures 2-3 and 2-4).** It is apparent that the effects of age are minimal in younger subjects (plateau) and around menopausal ages a decline

in cortical index could be observed. This trend is more pronounced in females, which is in accordance with the findings in other populations (**262**).

The prevalence of osteoporosis was established according to statistical criteria. It was based on the distribution of cortical index values in people under the age of 45 in each gender separately. A "cut-off" value of cortical index was defined as the mean minus two standard deviations of the distribution in each sex, which is a suggested criterion for osteoporosis when predicted by cortical index (**260-262**). As this criterion has been widely used, and the measurements were performed by a single device and technician and analyzed by a single experienced assessor, I suggest that the likelihood of substantial procedural errors is minimal.

Figure 2-3: Effect of age on cortical index in males.

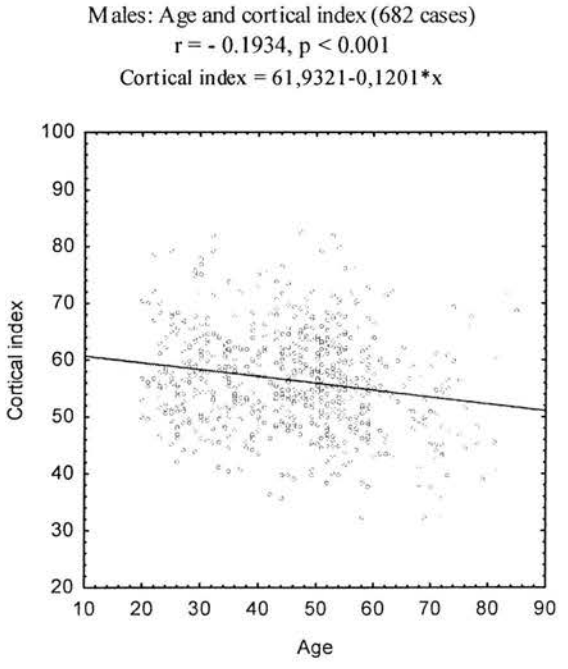
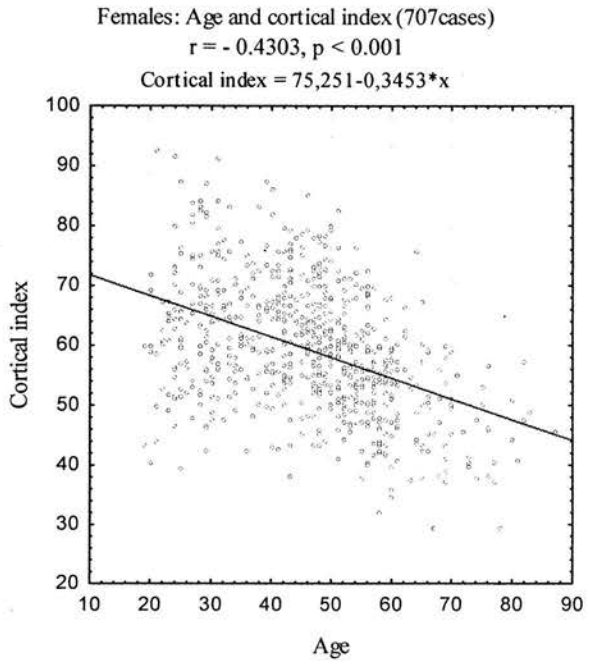


Figure 2-4: Effect of age on cortical index in females.



2.3.4. Statistical Analyses

To investigate the relationship between inbreeding and the cortical index and prevalence of osteoporosis, the ecologic epidemiological design was used. The prevalence of osteoporosis was compared between the villages with various levels of mean inbreeding. As the expected prevalence of the disease could be considerably influenced by variations in sex and age distribution of the different village populations under study, the prevalence rates were presented separately for each gender, and standardized by age and age^2 according to the regressions presented in **Figures 2-3 and 2-4**.

The mean values of cortical index and prevalence rates were adjusted accordingly in all 14 villages included in this part of the study, using the regression of prevalence in the entire sample on combined effect of sex and age, and weighted

according to sample size (**5**). The statistical significance of correlations between the standardized means of cortical index values, standardized prevalence of osteoporosis and mean inbreeding coefficients were determined in a linear regression model using "Statistica" software.

2.4. INVESTIGATION OF THE EFFECTS OF INBREEDING ON LEARNING DISABILITY

2.4.1. Sample Selection

The prevalence of learning disability (LD) was determined in 10 isolate villages on 5 different Croatian islands: G. Humac (Brac island), Gdinj and Svirce (Hvar island), Pupnat, Cara, Racisce, Lumbarda and Smokvica (Korcula island), Susak (Susak island), and Lastovo (Lastovo island) (see **Figures 2-1 and 2-2**). Those 10 villages are characterised by reduced environmental variation and their inhabitants share very similar environmental factors (climate, nutrition, socio-economic status, occupation, education, housing), as has been demonstrated in previous studies (see **Table 2-2**). In theory this should create a favourable setting for the study since it should help limit socio-economic and cultural bias in the interpretation of the results. Another favourable characteristic of these populations for our study is the diversity of the attitudes towards inbreeding, as was discussed in section 2.1.

2.4.2. Estimation of Inbreeding Coefficients in 10 Villages

Genetic characterisation of these villages included the computation of average inbreeding coefficient of each village based on reconstruction of genealogies of a sample of examinees which formed 20-30% of adult village population, as detailed in section 3.2.2. The individual inbreeding coefficients (F) were then computed according to Wright's path method (**258**):

$$F = \Sigma_{(1\rightarrow c)}(1/2)^{(n_i+m_i+1)}$$

where m and n refer to the number of paths from a common ancestor, and c refers to the number of common ancestors. The genealogical inbreeding coefficient for each village was then computed as the average of all individual F values.

2.4.3. Estimation of Learning Disability Prevalence

For the purpose of this study, learning disability was defined as the inability to attend the public school system. As the elementary schools (grade 1-8) in the place of the study are both public and compulsory, the assessment of child's ability to attend the school is performed at the age of six, based on a combination of IQ score measurement and behavioural assessment. The assessment is based on standard set of tests, as required by Croatian Ministry of education and culture (**263**). These tests include: (a) perception test, test of point linkage, test of knowing facts, drawing test and numerical test; (b) intelligence test based on drawing a human image; (c) "Bender Gestalt" test; (d) Raven's progressive coloured matrices (**263**). Data on the individuals unable to attend school were retrieved from local general practitioners and were considered to be complete. The prevalence of LD was calculated as the proportion of individuals unable to attend school in the total population of each village (as of January 2001). Ethical approval for this study was obtained from the Ethics Committee of the University Medical School in Zagreb, Croatia.

2.4.4. Statistical Analyses

Linear regression analysis of LD prevalence on F was performed using the data from all ten villages. The corresponding Pearson's coefficient of correlation (r) and the regression coefficient (b) were determined using the SYSTAT 7.0 software. The observed prevalence of LD in each of the studied populations was considered to approximate reasonably well the absolute risk of LD at pre-school age in each population. The relative risk for each unit increase of 5% inbreeding was inferred from

the slope of the linear regression curve as the ratio of the expected LD prevalence at the points of $F = 5\%$ and $F = 0\%$.

As pointed out by Freire-Maia (**264**), in certain instances the absolute and relative risk measures can be artificially low. An index - called "the genetic effects of inbreeding" (GEI) - was suggested as an alternative (**264**), and is calculated as:

$$GEI = (Pi-Po) / (1-Po),$$

where Pi is a probability of the event (in this case LD) in an inbred person (in this case a village with an average F greater than 3%), and Po is the probability of the event in a non-inbred person (in this case a village with an average F less than 1%).

The population-attributable fraction (PAF) was calculated by defining each village's LD prevalence value if their F was set equal to zero (P) according to the regression line. The intercept of the regression line on the Y-axis when F equals zero (X-axis) is an estimate of the LD prevalence in the absence of inbreeding. Then:

$$PAF = 1 - Psum / Npop,$$

where Npop is the total population size.

2.5. INVESTIGATION OF THE EFFECTS OF INBREEDING ON 10 COMMON COMPLEX DISEASES OF LATE ONSET (CORONARY HEART DISEASE, STROKE, CANCER, SCHIZOPHRENIA, UNI/BIPOLAR DEPRESSION, ASTHMA, TYPE II DIABETES, GOUT, PEPTIC ULCER, EPILEPSY)

2.5.1. Sample Selection

This study was undertaken in a subset of the original 25 villages described in Section 2.2.1. and in presented in **Table 2-2** and **Figure 2-2**. In 14 of the 25 villages, a field study was undertaken between 1997-2000 to compute the inbreeding coefficients of 1,339 adult individuals recruited during the initial field study in the late 1970s and early 1980s. At the same time, the disease status of the 10 most frequent complex diseases of late onset was determined in the entire current population of those 10 villages. The comparison of prevalence between the group of villages with "high", "moderate" and "low" inbreeding was then performed, as an ecologic study. Special attention was devoted to finding the examinees from the original 1,339 who were still alive and living in the same village today, i.e. nearly 30 years after the initial study. A total of 480 such individuals could be found, and they were used for a cohort study at the individual level.

Table 2-3 lists the villages chosen for the study, presents their categorization according to inbreeding prevalence into "high", "moderate" or "low" group, and shows the average inbreeding coefficients computed from genealogies (*F(gen)*) and isonymy (*F(iso)*). The six villages categorized into "high inbreeding" group were located in all three studied islands, and the mean individual inbreeding coefficient in the sample from each village was between 0.025 and 0.049. The four villages categorized into the "moderate inbreeding" group had mean individual inbreeding coefficients between

0.012 and 0.015, and were also situated on all 3 islands. Similarly, the "low inbreeding" group also included four villages from all three studied islands, and the mean inbreeding coefficient of their populations was not greater than 0.008. All these estimates weew based on *F(gen),* while *F(iso)* are overestimates of *F(gen)* by an average factor of 1.53.

The variables defined in each village to control for non-genetic confounding effects, especially socio-economic status, were the proportion of examinees with some college education (13 or more years in school) (*EDU*), the proportion of the population working in agriculture or fishery (*OCU*), the proportion regularly consuming a traditional mediterranean diet (*DIE*) and the proportion regularly smoking (*SMO*), and the average body mass index (*BMI*) (**Table 2-3**). The data on these variables were collected by a single researcher through a questionnaire (yes/no answers to the first four questions), and the height and weight of all examinees were measured by a "Hospitalija" anthropometer according to standard procedures defined in a guide to field methods (**260**).

The figures in **Table 2-3** are based on the initial sample of 1,339 adult individuals recruited during the initial field study in late 1970s and early 1980s. Population census data in 1981, 1991 and 2001 from the study villages show significant depopulation with minimal immigration over the last two decades. Thus only 480 individuals still resident in the study villages were available for follow up from the initial sample of 1,339 adult individuals recruited during the field study in the late 1970s and early 1980s.

2.5.2. Computation of Individual Inbreeding Coefficients in 14 Villages

This has been done using pedigree information on 4 ancestral generations (five generations where these occurred over a similar timeframe), recorded during the initial field work and supplemented by a study of parish registries stored in local churches

during 1997-2000. The individual inbreeding coefficients (F) were then computed according to Wright's path method (**258**):

$$F = \Sigma_{(1 \rightarrow c)}(1/2)^{(n_i+m_i+1)}$$

where m and n refer to the number of paths from a common ancestor, and c refers to the number of common ancestors. The genealogical inbreeding coefficient for each village was then computed as the average of all individual F values. To further support these estimates, F was calculated from isonymy as suggested by Tay and Yip (**259**), and mean values were derived for each of the 14 villages. Estimates based on isonymy are generally thought to be positively biased, and so provide an upper bound for true F values.

Table 2-6: Criteria adopted for establishing diagnoses of 10 selected late-onset diseases.

1. CORONARY HEART DISEASE: Includes angina pectoris and/or myocardial infarction. *Angina pectoris* may be diagnosed by a GP and represents a clinical syndrome due to myocardial ischaemia characterised by repeating episodes of precordial discomfort or pressure, typically precipitated by exertion or relieved by rest or sublingual nitroglycerine with or without reversible ischaemic ECG changes. *Myocardial infarction* must be diagnosed by a consultant in a local general hospital based on presenting symptoms (deep substernal radiating pain) and supported by a combination of ECG findings (e.g. deep Q-waves, elevated or depressed ST segments, deeply inverted T-waves), elevated white blood cell count and elevated myocardial component of creatine kinase and lactic dehydrogenase with or without myocardial imaging.
2. CEREBRAL STROKE: It must be diagnosed by a consultant in a local general hospital based on presenting symptoms. These include variable neurological defects that increase over 24-48 hours due to an enlarging infarction of the brain caused by stenosis, embolism or thrombosis of the intra- or extracranial arteries. Clinical diagnosis may be supported by CT/MRI scan or arteriography.
3. CANCER: It must be diagnosed by a consultant in a local general hospital. It is defined as any abnormal cellular growth of any site which was histologically confirmed as malignant.
4. DIABETES TYPE II: It may be diagnosed by a GP. Diagnosis can be established in persons older than 30 years if any of the following conditions are met: symptomatic hyperglycaemia (polyuria, polydipsia, polyphagia, weight loss) or diabetic ketoacidosis or non-ketotic hyperglycaemic-hyperosmolar coma or plasma (or serum) glucose level greater than 140 mg/dL after an overnight

fast on 2 occasions or the development of any of the late complications (e.g. retinopathy, nephropathy, atherosclerotic changes on peripheral or coronary arteries, neuropathy).

5. SCHIZOPHRENIA: It must be diagnosed by a consultant in a local general hospital. It is a mental disorder with a tendency towards chronicity which impairs functioning and is characterised by psychotic symptoms involving disturbances of thought, perception, feeling and behaviour. Six specific criteria include psychotic symptoms such as delusions, hallucinations and formal thought disorder; deterioration from previous level of functioning; continuous signs of illness for at least 6 months; a tendency toward onset before the age of 45 years; symptoms not due to mood (affective) disorder; symptoms not due to organic mental disorder or mental retardation.
6. UNI/BIPOLAR DEPRESSION: It must be diagnosed by a consultant in a local general hospital. Diagnosis is based on a combination of symptomatic picture of depression, chronic course, family history, response to treatment, TRH stimulation test, dexamethasone suppression test and sleep EEG.
7. EPILEPSY: It must be diagnosed by a consultant in a local general hospital. It is a recurrent paroxysmal disorder of cerebral function characterised by sudden brief attacks of altered consciousness, motor activity, sensory phenomena or inappropriate behaviour caused by abnormal excessive discharge of cerebral neurones. Diagnosis may be supported by abnormalities in EEG, CT, or MRI.
8. ASTHMA: It may be diagnosed by a GP. It is a lung disease characterised by three features: airways obstruction that is usually reversible, presenting with attacks of tachypnoea, tachycardia and audible wheezes; airway inflammation; and increased airways responsiveness to a variety of stimuli. Diagnosis may be supported by a family history of allergy or asthma.
9. GOUT: It may be diagnosed by a GP. Diagnosis is based on a characteristic set of clinical features of acute gouty arthritis (recurrent acute mono/polyarticular pain in peripheral joints, often nocturnal, progressively more severe, with swelling, warmth, redness and tenderness). It may be supported by any of the following: elevated serum urate (greater than 7 mg/dl), demonstration of urate crystals in tissue or synovial fluid, or dramatic response (within 24 hours) to colchicine.
10. PEPTIC ULCER: It must be diagnosed by a consultant in a local general hospital. It is a circumscribed ulceration of the gastric or duodenal mucous membrane causing a chronic and recurrent burning, gnawing, aching, soreness or empty feeling in the epigastrium. Diagnosis must be supported by endoscopic findings and/or X-ray studies with barium.

2.5.3. Defining Disease Status and Estimating Prevalence

Inspection of local medical records revealed that the 10 most commonly reported medical conditions (other than minor self-limiting disorders) with adult onset were coronary heart disease, stroke, cancer, schizophrenia, epilepsy, uni/bipolar depression, asthma, adult-type diabetes, gout and peptic ulcer. Specific diagnostic criteria were established for each of these 10 conditions following those presented in the 16th edition of Merck's Manual (see **Table 2-6** for details). Two medical doctors, who were blind to the inbreeding status of each individual, inspected the medical records of the 480 individuals and recorded whenever appropriate diagnostic criteria were met. The doctors visited each village between March and October 2000 and reviewed the medical records of all inhabitants in collaboration with local general practitioners, who typically had lived in the community for a number of years and were familiar with each patient's history. Diagnoses were supported wherever possible by medical records from consultant specialists at the local general hospital in Split.

2.5.4. Statistical Analyses

The effects of inbreeding on the prevalence of 10 complex diseases were first studied through an ecologic study design. In this approach, the prevalence of diseases were simply compared between the groups of villages of "high", "moderate" and "low" inbreeding. Disease prevalence rates were standardized by sex and age to the total population of all 14 villages included in the study, using 10-year age intervals and a direct standardisation method. The statistical significance of the differences between the groups were assessed using a Chi-square test for independent samples.

In an attempt to overcome the possible confounding effects of unmeasured environmental exposures or population stratification, the relationship between individual

inbreeding coefficients and disease outcomes was investigated among the 480 individuals from the original cohort who were still alive and resident in the study villages in 1997-2000. Data on age, sex, smoking status, village of residence, height, weight, and individual inbreeding coefficient (F) were analysed in a logistic regression with disease status as the outcome. Age and sex were forced into the prediction model irrespective of whether they were formally significant. Other main effects (inbreeding, smoking, height, weight and village) were entered with p=0.05 and all higher order effects and interactions with p=0.01. The statistical analyses in this cohort study design were planned, designed and performed under the leadership and in collaboration with Dr Andrew Carothers from the Human Genetics Unit, Medical Research Council, Edinburgh, UK, and then repeated under his guidance by the author of this thesis.

Population attributable fraction (PAF) estimates for inbreeding were calculated by logistic regression allowing for individual differences in the variables: village, sex, age, height, weight and smoking. The appropriate regression was determined as a function of all associated variables (including *F*), then the probability for each individual of having the disease outcome was calculated assuming an *F* value set at 0. The sum of all such probabilities, P_{sum}, is an estimate of the number affected in the absence of inbreeding, but with other variables remaining unaltered. Then

$$PAF = 1 - P_{sum}/N_{aff},$$

where N_{aff} is the actual number of affected individuals. In deriving the PAF values, the effects estimated from the subset of 14 villages were applied to the full data set from all villages.

2.6. INVESTIGATION OF THE EFFECTS OF HETEROSIS ON MATING SUCCESS, FITNESS, HEIGHT, BLOOD PRESSURE, CHOLESTEROL AND TRIGLYCERIDE MEASUREMENTS

2.6.1. Sample Selection

The studies of inbreeding effects up to this point were based on a historic sample from the islands of Brac, Hvar and Korcula, from about one generation ago. In attempt to repeat the main effects of inbreeding observed in that setting and to reaffirm the arguments by investigating the effects of heterosis (also known as "hybrid vigour", theoretically expected to result from admixture and outbreeding due to an increase in individual heterozygosity), a new study was designed and conducted. During 2002 and 2003, the field work was carried out in 9 villages on the four neighbouring Adriatic islands (*Rab*, *Vis*, *Lastovo* and *Mljet*) (**Figure 2-5**).

Figure 2-5 presents the location of 9 villages on the islands, while the 10th population was deliberately recruited from immigrants from the Croatian mainland into all the studied villages, to represent a genetically diverse and outbred control population inhabiting a similar environment. The 9 settlements were carefully chosen in 2002 to present a wide range of differing ethnic histories, fluctuations in population size, admixture and bottleneck events, known founding times, accessibility of genealogical records and population willingness to participate in the research program.

Figure 2-5: Location of 9 populations from the islands of Rab, Vis, Lastovo and Mljet in which an additional sample of 1,001 adult individuals was recruited during the field work in 2002 and 2003 (V1-V9). Immigrants from the mainland formed the 10th population (V10).

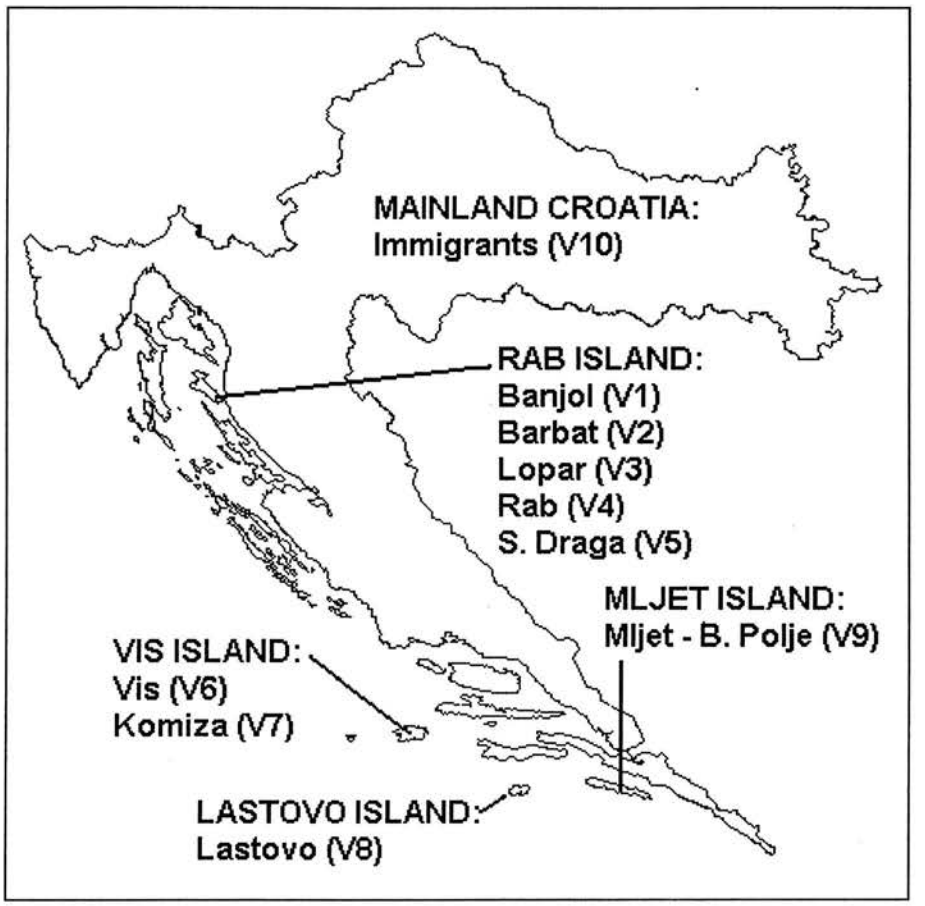


Table 2-7 shows some of the basic demographic variables for each studied village. The sources of the data for **Table 2-7** were derived from more than 30 data sources in Croatian, which included multiple books on the ethnic history of these populations, demographic records and statistical yearbooks of the Institutes and Ministries of the Croatian government. I believe that the data presented in the table are robust, as they were consistent throughout the data sources from the authorities of the Venetian Republic, the Austrio-Hungarian Empire, the Dubrovnik Republic and

Croatian authorities, who all governed those populations at various periods in history. Two variables (proportion of inbred population and proportion of grandparental endogamy) were derived from the field study undertaken under my supervision in 2002 and 2003.

Table 2-7: Some of the basic demographic variables for each studied village (Pop. size = total population in 2001; Admix. episodes, time since last = number of major admixture episodes in history and time elapsed since the last such episode; Last bottleneck and % Reduc. pop. size = Percentage of population that vanished during last bottleneck event and the corresponding percent reduction in population size; Trend (year): demographic index describing population increase or decrease, defined as the ratio of current population size and that in the given year; % inbred pop. = proportion of consanguineous examinees in the sample; % GP endog. = proportion of examinees' grandparents born in the village under study).

Village	Barbat (1)	Banjol (2)	Lopar (3)	Rab (4)	S. Draga (5)	Vis (6)	Komiza (7)	Lastovo (8)	Mljet (9)
Sample size	100	100	100	100	100	100	100	100	100
Pop. size (2001)	1,205	1,971	1,191	554	1,164	1,776	1,523	835	1,111
% inbred populat.	13 %	7 %	10 %	4 %	5 %	12 %	27 %	3 %	11 %
% GP endog.	90.3	41.9	98.4	39.4	46.9	87.7	91.3	71.7	93.5
Founded (ybp)	1,450	1,600	1,600	3,000	950	3,000	640	1,200	1,200
Admix. episodes	2	3	3	4	1	4	0	0	0
Time since last	350 yrs	350 yrs	350 yrs	350 yrs	350 yrs	350 yrs	640 yrs	1200 yrs	1200 yrs
Last bot-tleneck	550 ybp	550 ybp	550 ybp	550 ybp	550 ybp	50 ybp	50 ybp	1200 ybp	1200 ybp
% Reduc. pop. size	60 %	60 %	60 %	95 %	60 %	53 %	44 %	0 %	0 %
Maxim. populat.	1,300	1,971	1,500	5,000	1,164	4,300	3,572	1,602	2,106
Maxim. (year)	1950 AD	2001 AD	1400 AD	1400 AD	2001 AD	1910 AD	1910 AD	1931 AD	1948 AD
Trend (1750)	402 %	340 %	657 %	55 %	333 %	127 %	585 %	76 %	101 %
Trend (1875)	280 %	229 %	505 %	62 %	162 %	58 %	68 %	83 %	77 %
Trend (1925)	110 %	167 %	208 %	64 %	116 %	55 %	46 %	58 %	57 %

Two villages (*Rab* and *Vis*) date back to Illyrian period (e.g. some 1,000 years B.C.) and were later fortified by arriving Greeks and the succeeding Romans. By that time, both villages already experienced two major admixture events – first between Illyrians and Greeks, and later between Greeks and Romans. *Banjol* and *Lopar* were both founded by Greeks in 4th century as military camps. Two centuries later, Romans founded *Barbat* as their place of worship. The first Croats arrived to the region in the 7th century AD, admixing with populations of all existing settlements and establishing four more of the settlements under study – *Lastovo* and *Mljet* (9th century), *S. Draga* (11th century) and *Komiza* (14th century).

Finally, wars with the Turkish Empire in the Balkans forced immigrant refugees from the Croatian mainland to the islands. This occurred between 1570-1650 AD and resulted in the last major admixture event. It affected mainly *Rab* and *Vis*, i.e. the fortified settlements, but also the other villages, with the exception of *Komiza*, *Lastovo* and *Mljet* (**Table 2-7**). For the purpose of this study, a population bottleneck was considered a reduction in population size that occurred within the maximum time of 2 generations (50 years) and resulted in a decrease in population size greater than 40%. In 1449 and 1456, two plague epidemics killed or made refugees of 95% of the inhabitants of *Rab* village and 60% of the inhabitants from the other Rab island villages. Due to their geographic isolation from the mainland, the villages *Vis*, *Komiza*, *Lastovo* and *Mljet* avoided the epidemics. However, the isolation that saved the latter four villages during the 15th century attributed to economic hardship during the 20th century and a 44-88% reduction in population size. The degree of recent isolation of the villages studied was assessed as the percentage of the subject's grandparents who were born in the same village, ranging from 39.4% to 98.4%, and it loosely correlated with the prevalence of inbred individuals in the sample, showing that limited mate choice was not necessarily the main determinant of inbreeding in the villages. The current population size in the villages studied ranged from about 500 to about 2,000. The fluctuations in size through history were large, even during the 20th century (**265,266**). The percentage of persons

who are apparently inbred based on 2-generation parental genealogy ranged from 3% to 27%.

The field work to collect the data of interest - on marital status, number of children, height, blood pressure, serum cholesterol and triglycerides - was performed during 2002 and 2003. It was carried out by a team that included employees of the School of Public Health "Andrija Stampar" of the University of Zagreb Medical School and the Institute for Anthropological Research in Zagreb, Croatia. In each of the nine villages, a random sample of 100 adult inhabitants has been collected. Sampling was based on computerised randomisation of the most complete and accessible population registries in each village, which included medical records (Mljet and Lastovo islands), voting lists (Vis island) and household numbers (Rab island). An additional 101 examinees were recruited from second-generation immigrants into all 9 villages who volunteered to take the part, to form a genetically diverse control population that shares the same environment. The details on the ethical issues are presented in Section 2.7. The Ethics approvals, information sheets and informed consent forms are attached to the thesis as **Appendix 1**.

2.6.2. Defining Inbred, Endogamous, Admixed and Outbred Individuals in the Study Sample

Several days before the field examinations were performed in each settlement, all of the subjects selected for the study were sent a form to complete their individual two-generation pedigree, including the dates and places of birth and full names and marital surnames of both parents and all 4 grandparents. With respect to their genetic background, examinees were divided into 5 categories:

(a) inbred: when the same original (non-marital) surname, specific to the settlement, was found in at least one of father's and one of mother's parents, and further genealogical information revealed inbreeding;

(b) endogamous: when all four grandparents were born in the subject's village of residence, but there was no signal of inbreeding from surnames;

(c) admixed: when both father's parents were born in one village, and both mother's parents in another, different village;

(d) outbred: when less than four of the grandparents were born in the subject's village of residence, with some born in other village(s) or mainland Croatia.

(e) immigrants: when all 4 grandparents were born outside the four investigated islands.

The predicted individual heterozygosity is expected to be greatest in the admixed group, followed by outbred examinees. Endogamous individuals are expected to have smaller average individual heterozygosity, while the inbred group should have the lowest levels of heterozygosity.

2.6.3. Data Collection and Trait Measurements

All of the examinees were first interviewed by one of the trained surveyors, based on a questionnaire that was developed for this research program. The questionnaire collected extensive information on personal data (e.g. name, date and place of birth, gender, occupation and different lifestyle variables), with a special interest in marital status (single / married / divorced / widowed) and number of children. Information was retrieved on health complaints, drug intake and hospitalisation records. The survey also included the WHO (Rose) angina questionnaire (**267**), the WHO claudication questionnaire (**268**), the WHO non-communicable diseases questionnaire (**269**), the EU respiratory health questionnaire (**270**), two simple questionnaires on socio-economic status and nutrition habits developed for this specific population, and

the SF-36 questionnaire on quality of life (**271**). It usually took 25-40 minutes per examinee to complete the entire questionnaire. The copy of the large questionnaire comprising all these aforementioned questionnaires is attached to the thesis as **Appendix 2.**

Blood pressure (mmHg), height (mm) and weight (kg) were each measured by a single observer in local health centres and dispensaries between 8 and 11 AM following standard procedures (**260**). Blood pressure was measured on the right forearm in a sitting position. Two measurements of both systolic and diastolic blood pressure were taken 5 minutes apart in each individual. The resting time before each measurement was 5 minutes, systolic pressure was defined as the level (mmHg) at the 1st Korotkoff sound, and diastolic as the level at the 5th sound. Stature (height) and body mass (weight) were measured using a single anthropometer ("Hospitalija", Zagreb; height scale in millimeters, weight in decagrams). Biochemical analyses of total cholesterol and tryglicerides were measured from a blood sample that was first drawn into the 10 ml BD Vacutainer tube. This was performed in fasting individuals between 8:30 and 9:30 AM. Those tubes were then stored in the freezer at –20°C in all cases, and transported frozen to a single biochemical laboratory based in Zagreb. The laboratory was chosen as it had been internationally accredited for performing the analysis of interest and included in quality assessment and monitoring program RIQAS (www.riqas.com, accessed on June 1st, 2005).

2.6.4. Statistical Analyses

Analyses of the effects of inbreeding, admixture and outbreeding were performed on mating success (measured as entering at least one marriage and having at least 1 child in that marriage), average number of children (in examinees over 40 years, as this was justified as the age of completed family size in the overwhelming majority of cases from genealogical records; the average age difference between males and females at marriage was 2.1 years), mean height, blood pressure, cholesterol and triglyceride values. The

hypothesis was that heterosis may enhance "positive" effects, such as success in finding a spouse, number of children and personal height, while inbreeding should lead to increased blood pressure, cholesterol and triglycerides. All the measured individual values for studied QTs were first standardised for the effects of age using a standard regression model (**5**). The mean values between the groups were compared using an ANOVA test.

2.7. ETHICAL ISSUES

Ethical approval for this research was sought from appropriate research ethics committees in Croatia and Scotland. Both approvals were granted, and they are attached in **Appendix 1**. The approval in Croatia was obtained from Ethics Committee of the Faculty of Medicine of the University Medical School in Zagreb, Croatia, on May 7th, 2001. The approval in Scotland was obtained from Multi-Centre Research Ethics Committee (MREC) for Scotland on October 24th, 2001.

All the examinees enrolled in the study have read the information booklet (attached in **Appendix 1**). In each village, this booklet was sent to 120 persons randomly chosen to form the sample representative of the village. The information booklet explained why has their village been chosen for the project, what were the aims of the project, who lead the project, what would they be asked to do, what examinations were involved, what results would they be given, how would they benefit from the study, were there plans for further work, and whether there had been an ethical approval granted.

Informed written consent was obtained from all participants in the study, where they agreed to take part based on information provided in information booklet. The consent form is attached in **Appendix 1**. Participation rate was over 85% in all villages, thus enabling us to obtain a desired sample size of 100 in all of the villages under study by including the first 100 invited examinees that agreed to participate in the study.

3. RESULTS

3.1. EFFECTS OF INBREEDING ON BLOOD PRESSURE

3.1.1. Ecologic Study at the Village (Population) Level

Measurements recorded during a survey in 1979-81 in an *untreated* population permitted the analysis of BP as a quantitative trait. The prevalence of hypertension among non-inbred individuals in the study population was ~ 20%. Average inbreeding measures for each of the 25 villages were based on Wright's path method and isonymy. A highly significant linear correlation was found between mean inbreeding coefficient of study individuals in each village and the prevalence of hypertension (**Figure 3-1a,b**).

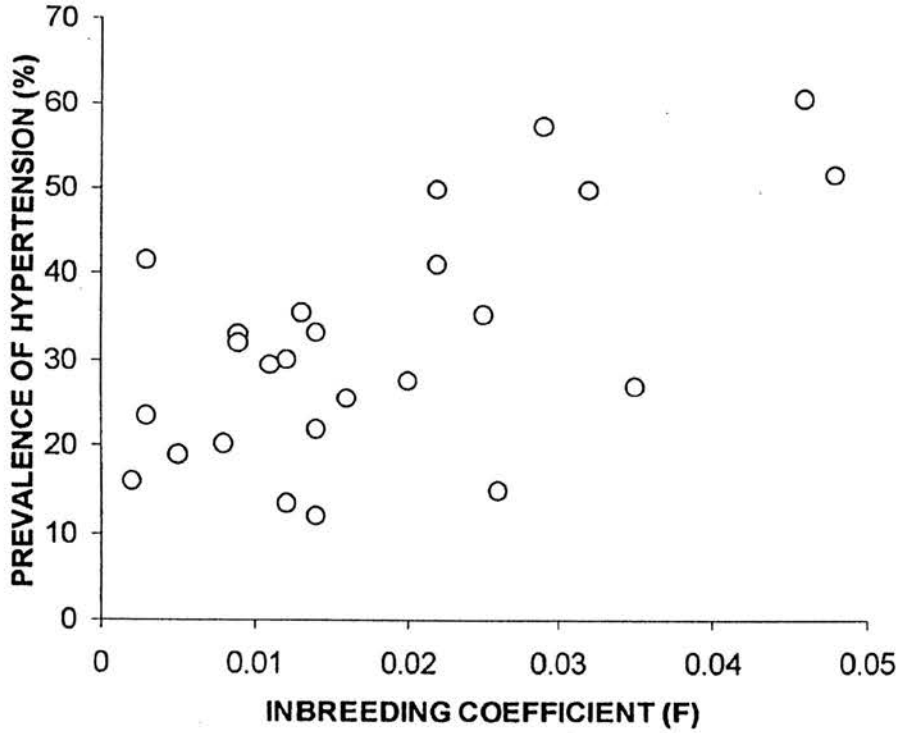
3.1.2. Regression Analysis at the Individual Level

To explore this observation further, multiple regression analysis of systolic and diastolic BP and individual inbreeding coefficients (F) was performed on the data from 2760 individuals in the 25 Dalmatian villages, controlling for the main recognised determinants of BP (age, sex, height and weight), village of residence and smoking status. A strong linear correlation between F and adjusted systolic and diastolic BP was found in both males and females (**Figure 3-2a,b**).

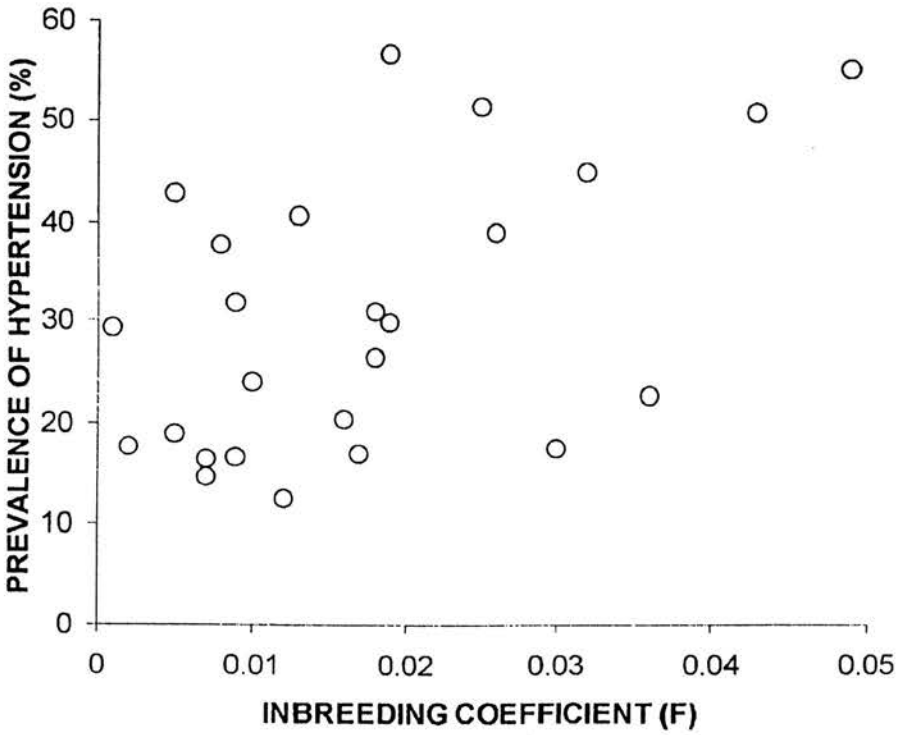
Figure 3-1 (next page): Relationship between prevalence of hypertension and average genealogical inbreeding coefficient in 25 villages from the islands of Brac, Hvar and Korcula.

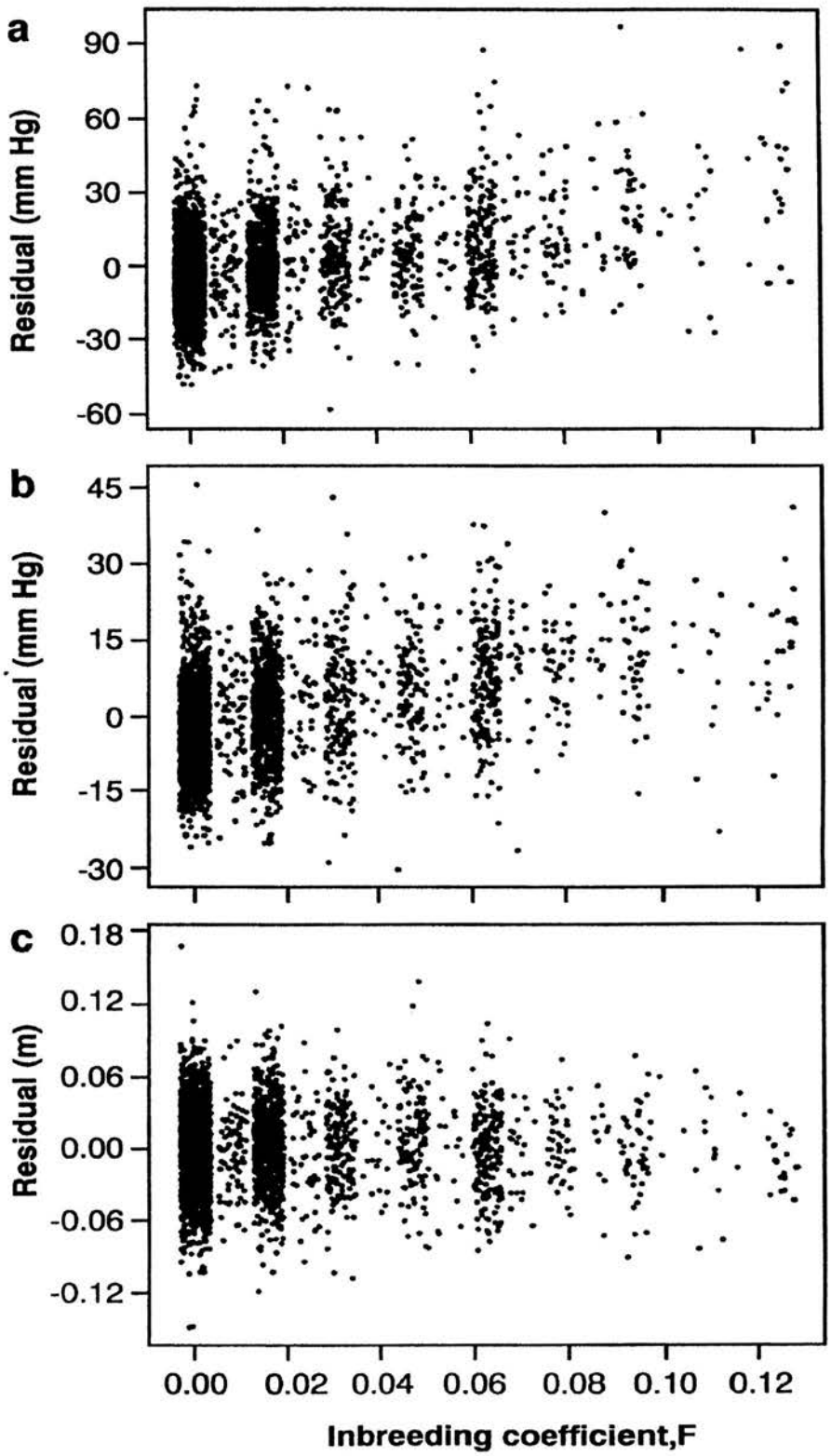
Figure 3-2 (page after next): Relationship between measured blood pressure (BP) and individual genealogical inbreeding coefficients (F) in 25 villages from the islands of Brac, Hvar and Korcula (individual residual values on BP on F plotted for 2,760 study subjects for systolic BP, diastolic BP and height).

MALES



FEMALES




a
Residual (mm Hg)
90
60
30
0
-30
-60
b
Residual (mm Hg)
45
30
15
0
-15
-30
c
Residual (m)
0.18
0.12
0.06
0.00
-0.06
-0.12
0.00
0.02
0.04
0.06
0.08
0.10
0.12
Inbreeding coefficient,F

The multiple regression model which incorporated all the predictors from **Table 2-2** showed that both systolic and diastolic BP levels correlated positively with age, weight and individual inbreeding coefficients and negatively with height and smoking status in both males and females. The regression model explained 40-50% of the phenotypic variance in BP. The strongest effect was clearly individual inbreeding coefficients, which alone explained approximately 15% of the variance in males and 10% in females in both systolic and diastolic levels. An increase in F of 0.01 corresponded to an increase of approximately 3 mm Hg in systolic and 2 mm Hg in diastolic BP in both sexes (**Carothers A, personal communication and Appendix 3**).

Assuming the effect of inbreeding (F) on BP depends on the number and dominance properties of QTL alleles, the equation (2) from section 1.2.5 uses the slope of the regression of F on BP to predict the minimum number of QTL (n) affecting the trait, assuming equal effects. The value B, which is the negative of the regression coefficient of trait mean value in a population under study on inbreeding coefficient, F, provides a useful summary statistic for the genetic damage that would occur if all deleterious recessives were made homozygous ("inbreeding load", equals to $F = 100\%$). The value of B due to deleterious mutations depends only on the genomic mutation rate and the dominance of individual mutations (**221**,**222**). If all mutations have the same effects, B for a given trait that is positively correlated with fitness is:

$$B = U\{(1/h) - 2\}\alpha \qquad \text{(Eq. 1)}$$

where U is the mutation rate per diploid individual to deleterious alleles affecting the trait; h is the extent to which fitness is reduced in heterozygous mutation, relative to its effect in homozygotes; and α is a constant of proportionality relating the effect of a mutant allele on the trait in question to its effect on fitness. If the dominance variance,

V_D, of the trait is known, a lower bound to the number of genes, n, affecting the trait is provided by:

$$n \geq B^2 / V_D \qquad \text{(Eq. 2)}$$

Taking the total phenotypic variance as an upper limit for the value of V_D, this analysis shows that the genetic component of blood pressure variability in this population is influenced by not less than several hundred QTL. The predicted number of equal-effect loci from these equations is 306 and 405 for systolic BP and 375 and 615 for diastolic BP in males and females, respectively (**Carothers A, personal communication and Appendix 3**). There were more QTL contributing to effects in females than in males, which may be partially explained by the substantial gender differences in physiological regulation of traits such as blood pressure or bone mass index (see later in the example of the effects on osteoporosis), which also leads to significant differences in life expectancy between the sexes.

Height was analysed in a similar fashion (**Figure 3-2c**) since it shows additive variance but no major dominance component (**195,272**). Therefore, it seemed as an ideal "control" quantitative trait, which is known to be considerably affected by socio-economic factors, but should not be expected to be affected by inbreeding (due to its low dominance variance and high additive variance). The results showed that the slope of the regression of F on height did not differ significantly from zero, as predicted. This is a very important finding, which shows that the observed effects of inbreeding are not easily explained by the positive correlation between inbreeding and socio-economic status of the examinees.

3.2. EFFECTS OF INBREEDING ON CORTICAL INDEX AND OSTEOPOROSIS

Another quantitative trait was studied, which is an important intermediate phenotype for a common late-onset disease (osteoporosis) (**260-262**). **Figures 3-3 to 3-6** show the results of an ecologic-epidemiological study of the association between the mean inbreeding coefficient in 14 villages (*F*) and mean standardized cortical index (CI) and osteoporosis (OP) prevalence. The coefficient of correlation (r) between *F* values and CI was -0.28 in males (p=0.08) and -0.42 in females (p=0.005). The "outlier" village in males (**Figure 3-3**), Dracevica, has the smallest sample size (N<20), so it is possible that chance deviation due to small sample size is responsible for high average values of CI and without that village the association would also formally reach statistical significance in males.

This relationship is much clearer in females (**Figure 3-4**), where the same small sample from Dracevica is again an "outlier", but the other examined villages followed the predicted relationship of CI and average inbreeding more closely than was observed in males. This may point to greater heritability or a more polygenic nature of bone mass loss control in females, than in males. However, the fact that the measured average inbreeding correlates significantly with the two biomedically important quantitative traits (blood pressure and cortical index), that are not significantly correlated mutually in this analysis at either population or individual level, makes it more likely that the effect is real. This points to conclusion that the villages characterized by higher rates of consanguinity indeed have greater prevalence of hypertension and osteoporosis.

Figure 3-3: Effect of average inbreeding coefficient (*F*) on cortical index (standardized age-adjusted residuals weighted according to sample size) in male examinees in 14 villages (r=-0.28, p=0.077, y = 0.168 – 8.823*x).

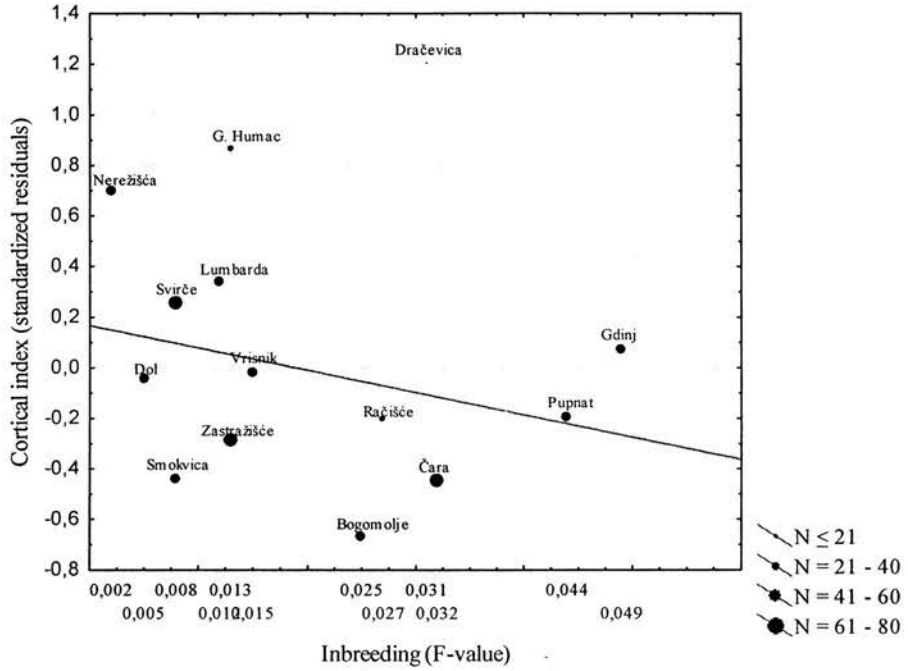


Figure 3-4: Effect of average inbreeding coefficient (F) on cortical index (standardized age-adjusted residuals weighted according to sample size) in female examinees in 14 villages (r=-0.42, p=0.005, y = 0.255 – 11.057*x).

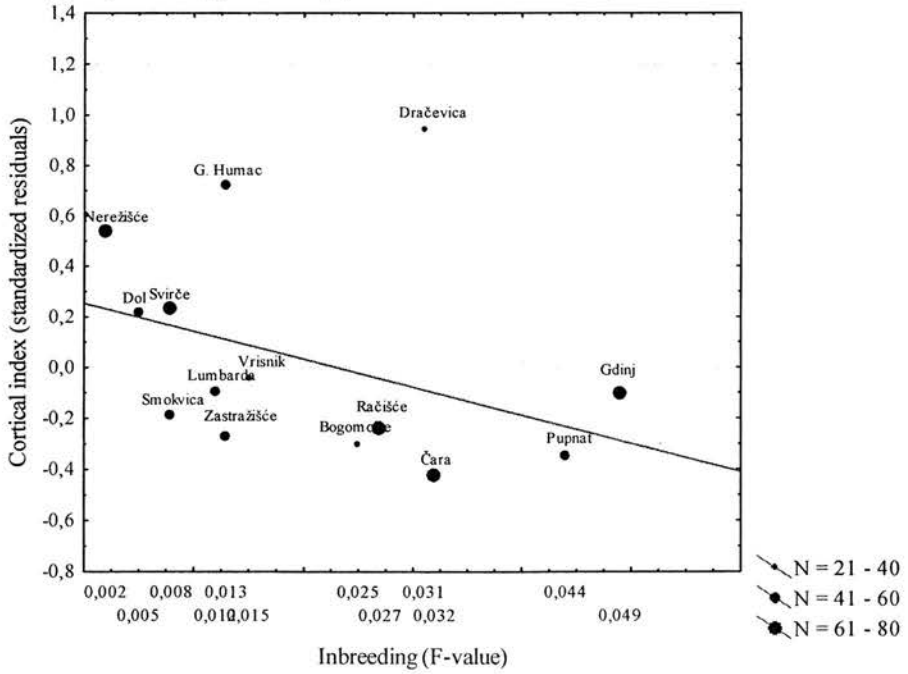


Figure 3-5: Effect of average inbreeding coefficient (*F*) on prevalence of osteoporosis (standardized by age, weighted according to sample size) in male examinees in 14 villages (r=0.32, p<0.001, y = 1.269 + 74.598*x).

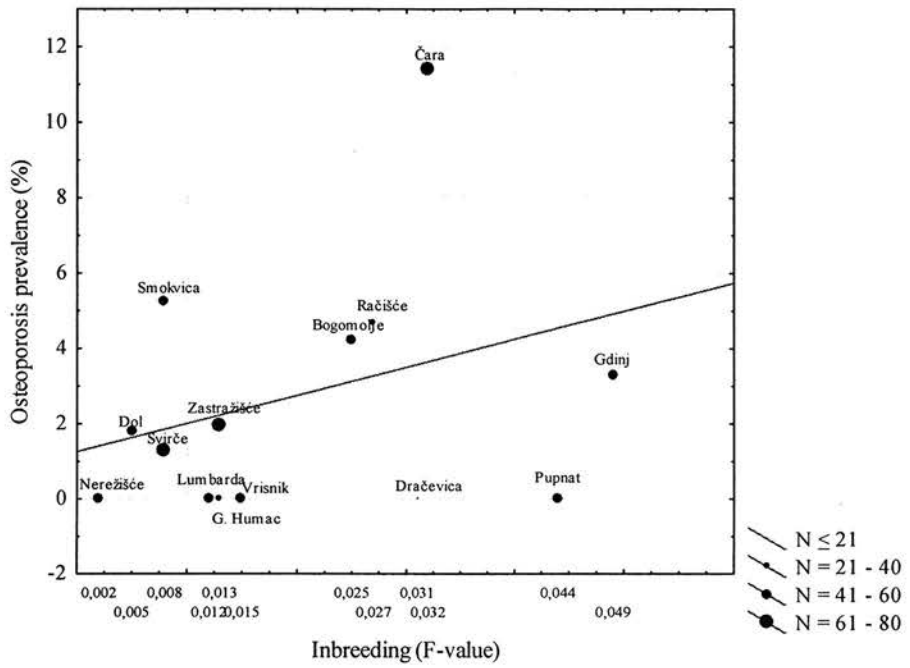
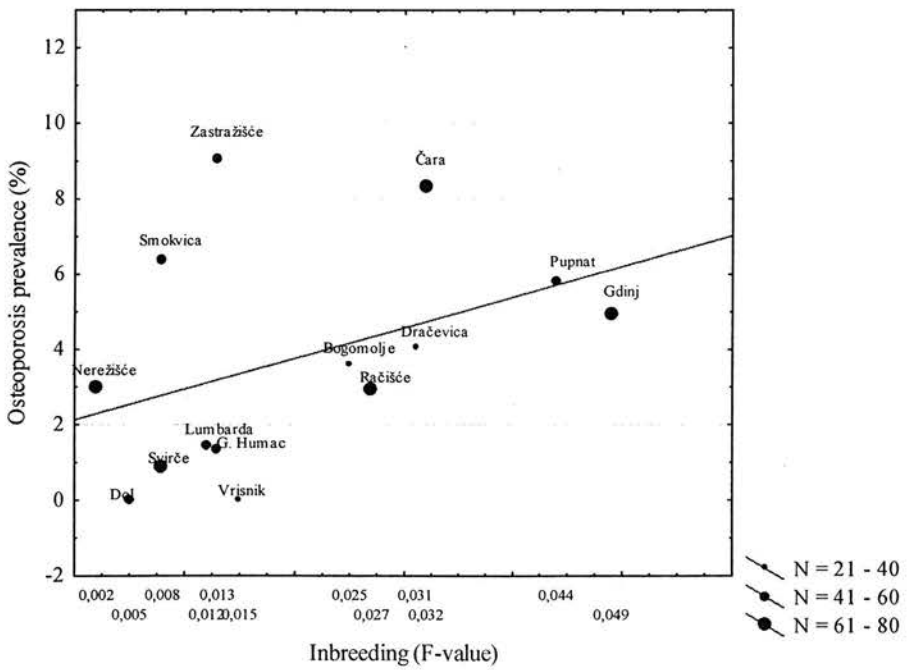


Figure 3-6: Effect of average inbreeding coefficient (F) on prevalence of osteoporosis (standardized by age, weighted according to sample size) in female examinees in 14 villages (r=0.43, p<0.001, y = 2.136 + 81.568*x).



The coefficients of correlation between F and OP prevalence were 0.32 in males ($p<0.001$) and 0.43 in females ($p<0.001$) **(Figure 3-5,3-6)**. These results may indicate a significant inbreeding depression on a biomedically important quantitative trait (CI), which leads to an increase in susceptibility to osteoporosis and its prevalence in inbred communities of Croatian islands. All the correlations were more significant in females. Furthermore, there is a general concordance between values for both males and females from the same village, which supports the hypothesis that the findings are indeed due to village-specific effects rather than chance findings.

There is a possibility that the high prevalence of large-effect variants have substantial effect on cortical index in particular villages, making them the apparent outliers. An example is the village of Cara in **Figure 3-5**. There is a concern that such outliers may considerably affect the regression and thus drive the conclusions. Also, there are many other possible confounding effects due to different levels of exposure to risk factors in different villages. It is not possible to obtain information on prevalence of main risk factors for osteoporosis in the population under study in this historic sample. Still, the consistency of the effects of inbreeding on the two non-correlated quantitative traits and the plausibility of the effect based on a large body of evidence on inbreeding depression in animals and plants argues that the observed effects may indeed be real.

3.3. EFFECTS OF INBREEDING ON LEARNING DISABILITY

Previously conducted studies of inbreeding effects on cognitive performance in the literature compared the prevalence of LD in an inbred cohort with non-inbred controls (**174**). This raised issues about social and cultural comparability of controls and the possible clustering of a Mendelian disease (a single large effect gene) in the inbred cases. In contrast, this study investigated 10 populations with similar environments and culture but with a spectrum of inbreeding coefficients and quite different founding populations. If this study found a comparable prevalence of LD in all 10 populations, this would not support inbreeding effect and the LD prevalence would be assumed to be determined mainly by factors related to the environment. However, if a positive correlation between inbreeding levels and LD prevalence would be found across 10 villages, this would imply effects of inbreeding. A further advantage of having 10 distinct populations under investigation is to rule out the possibility of a rare Mendelian disease clustering, as it is unlikely that the same rare variant would be present in all 10 populations having very different founding populations and ethnic history.

Another hypothesis tested through this analysis was that if a modest increase in sharing the variants identical by descent leads to a significant change in prevalence of LD across several isolate populations that share very similar environmental effects, this would be consistent with a very large number of genomic loci influencing the condition. An in-depth theoretical consideration of the problem was presented by Morton (**174**).

Table 3-1 presents the villages studied, their respective total populations (in the year 2000), the average coefficients of inbreeding (computed as above), the number of cases of learning disability (LD) and the prevalence of LD in each village. The inbreeding coefficients in these villages ranged from 0.8% to 4.9%, and the prevalence

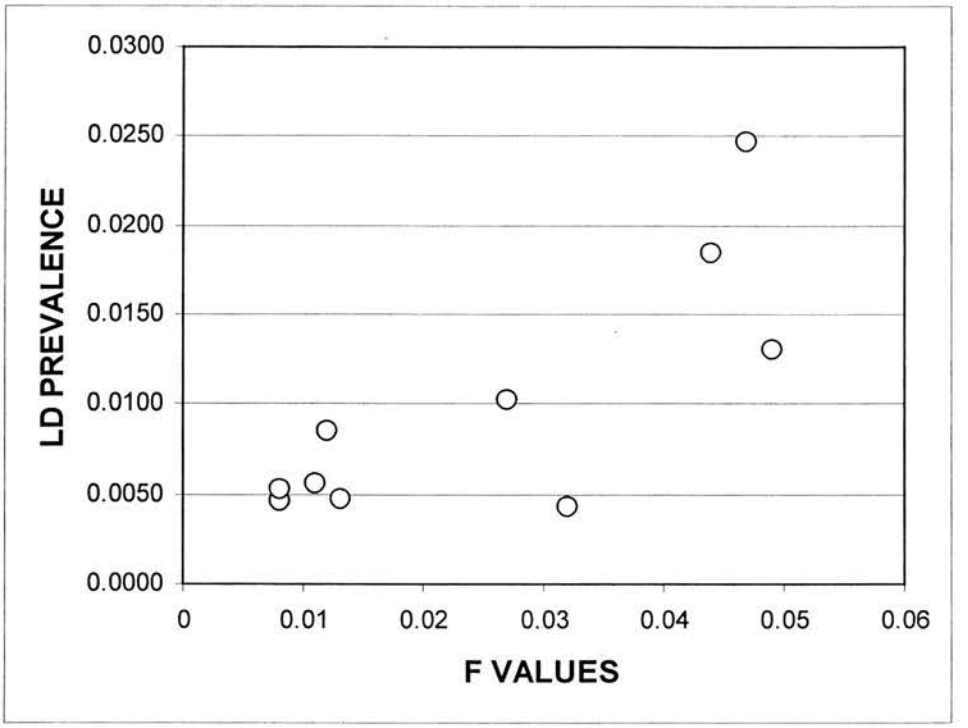
of LD from 0.43% to 2.47%. The mean prevalence of LD of 1.0% is somewhat below the world average figures for this type of population (**174**).

Table 3-1: Total population of 10 studied villages (as of January 2001), average population inbreeding coefficient (*F*) determined from genealogies, number of learning disability (LD) cases and the prevalence of learning disability.

VILLAGE, ISLAND	N (population)	*F* (genealogical)	LD cases	LD prevalence
Gdinj (Hvar)	153	0.049	2	1.31%
Susak (Susak)	81	0.047	2	2.47%
Pupnat (Korčula)	326	0.044	6	1.84%
Čara (Korčula)	464	0.032	2	0.43%
Račišće (Korčula)	290	0.027	3	1.03%
G. Humac (Brač)	214	0.013	1	0.47%
Lumbarda (Korčula)	354	0.012	3	0.85%
Lastovo (Lastovo)	899	0.011	5	0.56%
Smokvica (Korčula)	866	0.008	4	0.46%
Svirče (Hvar)	375	0.008	2	0.53%

Figure 3-7 presents the relationship between *F* and LD and the corresponding Pearson's correlation coefficient, which amounted to 0.80 ($p<0.01$). Although the relative risk per increase of inbreeding coefficient from 1% to 5% appeared to be quite high (about 4-5), the absolute risk only increased from 0.5% to 2.0%. This means that in the absence of inbreeding only 1 in more than 200 births should be expected to result in some form of learning disability in this population. When inbreeding reaches the level close to that of first-cousin matings (6.25%), some form of learning disability may be expected only in about 1 in every 50 births.

Figure 3-7: Plot of the values of prevalence of learning disability against average coefficients of inbreeding in ten studied villages.



The genetic effect of inbreeding (GEI) was 0.69%, which is derived from the difference in expected prevalence of LD when the villages with F greater than 3% and less than 1% are compared. The population-attributable fraction of cases due to inbreeding in all ten populations was very high, and it amounted to 76.6% (**Table 3-2**).

Table 3-2: Pearson's coefficient of correlation between inbreeding coefficient and prevalence of learning disability, regression coefficient, estimates of relative risk, genetic effect of inbreeding (GEI) and population attributable fraction (PAF).

MEASURE	VALUE
Pearson's coefficient of correlation (r)	0.80
Regression coefficient (b)	0.33
Relative risk (F=5% vs. F=0%)	10.23
GEI (F>3% vs. F<1%)	0.69%
Population attributable fraction	76.6%

The estimates in **Table 3-2** should still be taken with caution, as when only a single risk factor is studied, especially in a study of an ecological design such as this one, which is especially prone to confounding effects, it is more likely that the effects are overestimated than underestimated. However, a large number of studies in the literature using different designs in different time periods, populations and in different settings already confirmed a strong negative association between inbreeding and cognitive performance and positive association with learning disability, which makes it likely that the effects observed in this study are also real, and that inbreeding is prevalent enough in the populations under study to enable thorough investigations of inbreeding effects.

3.4. EFFECTS OF INBREEDING ON 10 COMMON COMPLEX DISEASES OF LATE ONSET (CORONARY HEART DISEASE, STROKE, CANCER, SCHIZOPHRENIA, UNI/BIPOLAR DEPRESSION, ASTHMA, TYPE II DIABETES, GOUT, PEPTIC ULCER, EPILEPSY)

3.4.1. Ecologic Study at the Village (Population) Level

Table 3-3 presents village data in three groups ("high", "moderate" and "low") defined by the average level of inbreeding which was presented in detail in **Table 2-2**. The allocation of villages to these three groups according to the estimates of F based on genealogy (F_{gen}) is broadly consistent with those based on isonymy (F_{iso}). On a log scale, the correlation between the two measures of inbreeding across villages was 0.92, with F_{iso} exceeding F_{gen} on average by a factor of 1.53. To explore the confounding effects of other variables, such as socio-economic factors, differences in nutritional factors and smoking, the prevalence of these risk factors was presented for each village in **Table 2-2**. Overall, this presentation confirmed the low levels of higher education and the high proportion of the population eating a traditional diet and employed in agriculture and fishing, with very little variation across the villages. Levels of reported smoking, however, varied widely from village to village.

In **Table 3-3**, the prevalence of 10 complex diseases standardized by age and sex to the total population of all 14 villages is presented separately for the groups of villages with relatively "high", "moderate" and "low" average inbreeding coefficient values. The levels of disease prevalence were consistently higher in the groups of villages with

"moderate" in comparison to those with "low" average inbreeding. The differences reached statistical significance for gout, uni/bipolar depression, peptic ulcer, schizophrenia and cancer.

Table 3-3:
Age and sex-standardised prevalences of 10 complex diseases in groups of villages with relatively "high", "moderate" and "low" average inbreeding coefficient values. *Fgen* = weighed average inbreeding coefficient computed from genealogical data; *Fiso* = weighed average inbreeding coefficient computed from isonymy data; Statistical significance (*p values*) in highly and moderately inbred groups is calculated against the low inbreeding group (*p<0.05; **p<0.01; ***p<0.001).

	"High inbreeding" (N = 1,401)	"Moderate inbreeding" (N = 998)	"Low inbreeding" (N = 1,501)
	Average (*Fgen*)=0.036	Average (*Fgen*)=0.013	Average (*Fgen*)=0.006
	Average (*Fiso*)= 0.052	Average (*Fiso*)= 0.019	Average (*Fiso*)= 0.009
Coronary heart disease	13.28%	11.95%	11.23%
(95% confid. intervals)	(11.50% - 15.06%)	(9.94% - 13.96%)	
Stroke	2.43%	2.79%	1.73%
(95% confid. intervals)	(1.62% - 3.24%)	(1.77% - 3.81%)	(1.07% – 2.39%)
Cancer	*** 4.54%	* 3.44%	1.93%
(95% confid. intervals)	(3.45% - 5.63%)	(2.31% - 4.57%)	(1.23% - 2.63%)
Schizophrenia	*** 1.23%	* 0.96%	0.14%
(95% confid. intervals)	(0.65% - 1.81%)	(0.35% - 1.57%)	(0.00% - 0.33%)
Uni/bipolar depression	*** 10.26%	** 7.63%	4.51%
(95% confid. intervals)	(9.45% - 11.07%)	(5.98% - 9.28%)	(3.46% - 5.56%)
Asthma	3.63%	2.64%	2.60%
(95% confid. intervals)	(2.65% - 4.61%)	(1.64% - 3.64%)	(1.79% - 3.41%)
Type II diabetes	6.02%	7.35%	6.77%
(95% confid. intervals)	(4.77% - 7.27%)	(6.01% – 8.69%)	(5.50% - 8.04%)
Gout	*** 9.25%	*** 7.19%	3.96%
(95% confid. intervals)	(7.75% - 10.77%)	(5.58% - 8.80%)	(2.97% - 4.95%)
Peptic ulcer	*** 6.92%	** 4.29%	2.18%
(95% confid. intervals)	(5.59% - 8.25%)	(3.03% - 5.55%)	(1.44% - 2.92%)
Epilepsy	*** 1.47%	0.78%	0.31%
(95% confid. intervals)	(0.84% - 2.10%)	(0.23% - 1.33%)	(0.03% - 0.59%)

Villages in the "high" inbreeding group showed further increases in disease prevalence compared to the "moderate" group for all disorders except type 2 diabetes and stroke. Thus, the following disorders showed a stepwise increase in disease prevalence across villages stratified by average inbreeding coefficient: gout, depression, peptic ulcer, schizophrenia, cancer, epilepsy, coronary heart disease and asthma (the last two not statistically significant).

3.4.2. Cohort Study at the Individual Level

Table 3-4 shows the results when the analysis was repeated at the individual level (rather than village level), similar to the analysis of effects on blood pressure (Section 3.4.1.). The analysis included data from 480 individuals from 14 villages for whom individual coefficients of inbreeding were available, as they were included in the original (historic) sample. The aim was to explore if inbreeding coefficients represent a variable with a significant effect on the "disease presence" variable. Age and sex were forced into the multiple logistic regression model on the basis that these would be expected to be relevant in all conditions. Other main effects (inbreeding, smoking and log_weight) were entered with p=0.05, and all higher-order effects and interactions with p=0.01. Schizophrenia and epilepsy were excluded from this analysis because of the small numbers of outcome events and therefore low study power to investigate predictors for these conditions.

Table 3-4:
Summary of results of multiple logistic regression (* excludes village effects, even though they were significant in the data set from 14 villages, since they could not be estimated for the remaining 11 villages in the full set; [1] Population attributable fraction (PAF) estimated from individual inbreeding levels (see Text)).

Disease	Predictor	Coefficient	SE	p value	PAF (%)[1]
C.H.D.	Constant	-6.20	1.27	0.000	34.3 *
	SEX	0.47	0.26	0.070	

	AGE	0.69	0.14	0.000	
	VILLAGE	---	---	0.022	
	F1	26.63	4.88	0.000	
Stroke	Constant	-3.28	0.71	0.000	26.3
	SEX	-0.30	0.40	0.464	
	AGE	0.34	0.20	0.095	
	F	13.37	5.85	0.022	
Cancer	Constant	10.17	6.27	0.105	23.2
	SEX	0.83	0.54	0.128	
	AGE	0.02	0.20	0.909	
	LN_WT	-3.49	1.40	0.013	
	F	13.0	6.17	0.035	
Depression	Constant	-3.86	0.58	0.000	42.2
	SEX	0.69	0.31	0.028	
	AGE	0.02	0.14	0.896	
	F	24.54	4.43	0.000	
Asthma	Constant	-3.74	0.80	0.000	47.7
	SEX	-0.40	0.45	0.374	
	AGE	0.36	0.23	0.110	
	F	23.10	6.04	0.000	
Diabetes	Constant	-28.16	4.58	0.000	0.0
(type II)	SEX	0.97	0.31	0.002	
	AGE	0.31	0.14	0.027	
	LN_WT	5.60	0.99	0.000	
Gout	Constant	-14.90	4.08	0.000	31.0
	SEX	-0.17	0.29	0.552	
	AGE	0.38	0.14	0.006	
	LN_WT	2.85	0.89	0.001	
	F	17.69	4.16	0.000	

Peptic ulcer	Constant	-3.23	0.63	0.000	30.6
	SEX	-0.18	0.35	0.615	
	AGE	0.36	0.18	0.043	
	F	15.68	5.06	0.002	

Table 3-4 shows that inbreeding coefficient (*F*) remained a significant (positive) predictor for every condition except type 2 diabetes. The forced inclusion of age and sex in the model acted to reduce slightly the significance of the effect of inbreeding, because of a small positive correlation between inbreeding and age. Village of residence was found to be a significant predictor only for coronary heart disease (CHD). Body weight (LN_WT) was a significant negative predictor for cancer, and positive for type 2 diabetes and gout.

In terms of health impact, the results on population attributable fraction (PAF) show that as much as 23-48% of the prevalence of these disorders (other than type 2 diabetes) in this population can be attributed to recent inbreeding. However, the highest PAF estimates are associated with the least common diseases, so there may be considerable uncertainty around them and some of them are likely to be overestimates.

3.5. EFFECTS OF INBREEDING, ENDOGAMY, ADMIXTURE AND OUTBREEDING ON FITNESS, MATING SUCCESS AND QUANTITATIVE TRAITS (HEIGHT, BLOOD PRESSURE, CHOLESTEROL AND TRIGLYCERIDES)

To support the observed effects of inbreeding on quantitative traits and complex diseases of late onset in a historic sample from the islands of Brac, Hvar and Korcula islands, a new large-scale field study was undertaken between 2002-2003. During this investigation, samples of 100 individuals of the present-day population were recruited from 9 different villages from the neighbouring islands of Rab, Vis, Lastovo and Mljet (see **Figure 2-1**) - a total of 900 persons. An additional 101 persons were recruited as genetic controls from the immigrants into the villages, although for 10 of them it was later realised through genealogical investigation that one of the four of their grandparents was born in the village of study.

The purpose of this analysis was to repeat the analysis of the relationship between inbreeding and blood pressure. In this study, values of triglycerides and cholesterol were also measured, which are additional established risk factors for cardiovascular diseases. However, to make the case stronger, the aim was to also show that admixed and outbred individuals had experienced certain benefits from their presumed increased heterozygosity in comparison to inbred or autochtonous ("endogamous") populations. Therefore, the individuals included in the study were categorized according to their individual genetic history (IGH). One of the important theoretically predicted effects of heterosis on humans ("hybrid vigour"), which has already been shown in a large number of studies in animals and plants, is improved

fitness (which can be measured as the average number of children per "IGH" category in persons who are of reproductive age). Others include increased average height per category in adults, and even the improved chance of finding a spouse (which can be measured as the proportion of married persons within each "IGH" category).

In order to undertake this investigation, the sample of 910 non-immigrant individuals were divided into "IGH" categories according to the expectation of the level of their predicted heterozygosity as follows:

H1: inbred individuals (N=92) – genealogy revealed consanguinity, the category with the largest expected homozygosity;

H2: "endogamous" individuals (N=436) – all 4 grandparents were born in same village, but no signal of consanguinity in the genealogy;

H3: outbred: (N=312): mixture of heritage of different villages and outside world; expected heterozygosity greater than H1 and H2 categories, less than H4.

H4: "admixed" individuals (N=70): offspring of 2 (H2) individuals from different villages, and should have the maximal expected heterozygosity.

Table 3-2 shows that, when the hypothesized beneficial effects of heterosis were analysed, the H4 category (with highest presumed heterozygosity) was consistently showing the most favourable results with respect to fitness, mating success and height in both males and females. Furthermore, the lowest mean values of systolic blood pressure and triglycerides and the second lowest mean value of cholesterol were also noted in this group of admixed individuals. These effects (except height in males) were statistically significant across the groups, and the number of observed 1st ranks for this group (for six of seven measured variables) was significantly greater than expected under the null hypothesis ($p<0.05$).

Category H1 (92 inbred individuals), however, showed increased values of systolic blood pressure, triglycerides and cholesterol in comparison to the other three groups. The average height in this group of individuals was also lower than in other 3 groups for both sexes. Mean fitness was significantly lower than in the H3 and H4 categories, while no major difference was seen in mating success.

Table 3-5:
Comparison of the effects of inbreeding (H1 category), endogamy (H2), outbreeding (H3) and admixture (H4) on mean fitness, mating success, height, systolic blood pressure, cholesterol and triglycerides in 910 present-day examinees from the islands of Rab, Vis, Lastovo and Mljet. Ranks are assigned according to the positive outcome (e.g. greater fitness, mating success and height; lower blood pressure, triglycerides and cholesterol). *p*-value across groups (columns) tests if the value in any of the categories significantly differs from other categories; *p*-value across variables (columns) tests whether any rank is observed more frequently within the group than under the null-hypothesis.

	H1 category (N=92)	H2 category (N=436)	H3 category (N=312)	H4 category (N=70)	p-value*
MEAN FITNESS (average number of children in >=40 years old)	2.024 (3^{rd})	2.016 (4^{th})	2.193 (2^{nd})	2.196 (1^{st})	<0.01
MATING SUCCESS (% married in >=40 years old)	96.93 (2^{nd})	94.95 (4^{th})	95.10 (3^{rd})	97.80 (1^{st})	<0.05
AVERAGE HEIGHT (adults only, male gender, in cm)	174.7 (4^{th})	175.2 (3^{rd})	175.3 (2^{nd})	176.3 (1^{st})	N.S.
AVERAGE HEIGHT (adults only, female gender, in cm)	161.6 (4^{th})	162.8 (2^{nd})	162.4 (3^{rd})	165.0 (1^{st})	<0.001
AVERAGE SYSTOLIC BLOOD PRESSURE (standardized, mmHg)	148.0 (4^{th})	142.8 (3^{rd})	140.6 (2^{nd})	139.7 (1^{st})	<0.001
log (TRIGLYCERIDES) (regression- standardized, in mmol/L)	0.33 (4^{th})	0.21 (2^{nd})	0.32 (3^{rd})	0.20 (1^{st})	<0.001
log (CHOLESTEROL) (regression- standardized, in mmol/L)	1.83 (4^{th})	1.76 (1^{st})	1.80 (3^{rd})	1.77 (2^{nd})	<0.001
p-value**	<0.10	N.S.	N.S.	<0.05	

* The mean values were compared across 4 groups of examinees using ANOVA (N.S. – not significant);
** The ranks expected in each group (1^{st}, 2^{nd}, 3^{rd} or 4^{th}) were counted in each column (observed frequency) and compared to predicted frequency under null-hypothesis using chi-squared test.

When the two intermediate groups ("endogamous" vs. "outbred") with greatest sample sizes (436 vs. 312) were compared, the "outbred" group (H3) had significantly greater mean fitness, but other differences were not striking. We conclude that the H2 group and H3 group do not differ sufficiently in their heterozygosity to reveal considerable differences. However, when compared to the H2 and H3 categories, the inbred individuals seemed to fare worse in most aspects, while the admixed individuals fared better.

Although most of the effects indeed confirmed the *a priori* hypotheses, we again have to take these results with a degree of caution, as the sample sizes used to test the hypotheses are relatively small. Also, it is likely that some of the studied outcomes are correlated, such as male height and their mating success (**273**). The most convincingly demonstrated effect, however, was again the increased systolic blood pressure among the inbred individuals, and increased height in both sexes among the admixed individuals.

4. DISCUSSION

4.1. AIMS AND SCOPE

The aim of this work was to investigate the effects of inbreeding on human quantitative traits and complex diseases of late onset. In the first part of the introduction, the current understanding of the genetic basis of human diseases was reviewed. It was stated that one of the greatest current interests in biomedical sciences was to improve the understanding of the genetic basis of common human diseases of late-onset. This improved knowledge could then help the design of genetic epidemiological studies that are being undertaken by numerous groups. The success of genetic association studies could improve our understanding of disease pathogenesis and lead to the development of new methods of early diagnosis and treatment, which may become individualized and tailored to specific patients.

One of the approaches that could be helpful in improving our understanding of the genetics of susceptibility to common complex diseases of late onset are studies of inbreeding effects on the incidence and prevalence of such diseases. The measurement of those effects could provide some empirical evidence to support one of the two opposing (but not necessarily mutually exclusive) views on the genetic basis of complex late-onset diseases: the common disease/common variant hypothesis (CD/CV) and the common disease/rare variant hypothesis (CD/RV). Under the first hypothesis, major effects of inbreeding should not be expected, but under the second hypothesis, where it is presumed that the susceptibility variants are extremely numerous, rare in the population, partially recessive and likely to be of small effect, inbreeding could substantially affect disease prevalence.

In the second part of the introduction, the results of the extensive literature review on the documented effects of inbreeding on human health in the pre-reproductive and post-reproductive periods were presented. It was shown that the main focus of interest of a large majority of previous studies were diseases of early onset (pre-

reproductive), that have been selected against during the course of human history. However, it was surprising to note a paucity of investigations studying the effects on human late-onset traits and diseases.

There are several possible reasons why the effect of inbreeding on late-onset complex chronic diseases has not been widely evaluated to date. Firstly, in countries where inbreeding is prevalent in the population, life expectancy is generally considerably shorter than in western communities, and late-onset diseases did not historically represent research priorities. Therefore, the effects of inbreeding were mostly investigated in small isolated communities (geographic, cultural, linguistic, ethnic or religious isolates) in developed countries, which could be more easily reached. In most developed countries, however, the isolated human populations characterized by inbreeding often do not have the same level of access to health care as the general (especially urban) population where the public health sector is well developed and the majority of epidemiological studies are being undertaken. Therefore, the health status of human isolates is not easily evaluated from medical records or communicated with their local physicians. In addition, there are not many isolate populations world-wide with well-preserved parish registries from which reliable estimates of inbreeding coefficients can be determined, based on the familial relationships over at least several ancestral generations. Furthermore, there are usually multiple concerns over confounding factors, as it is often quite difficult to find a non-inbred control population which would match the studied inbred isolate in environmental exposures and differed significantly only in genetic structure. Contemporary isolates usually share specific climate and environment, as well as a multitude of common socio-cultural factors such as diet, lifestyle, religious practices and socio-economic status for which it is very difficult to control. In small and isolated communities the phenomena such as founder effect, genetic drift and inbreeding can significantly affect allele and genotype frequencies, which invalidates comparisons with control populations where genotype frequencies follow Hardy-Weinberg equilibrium (**274**). Khoury et al. concluded that, despite rare attempts, there is still hardly any study that would satisfactorily deal with all the usual concerns including

small sample sizes, unreliable inbreeding estimates, doubtful disease diagnoses and inappropriate control populations (**6**).

The analysis of the prevalence of inbreeding as a potential risk factor in the human population showed that up to 1 billion people globally (nearly 15% of the total population) are likely to be offspring of consanguineous unions and have an average inbreeding coefficient of about 3%. Given the number of 25,000 genes in the human genome, this would imply that those 15% of humans have about 800 more genes identical by descent than all other humans. It is therefore surprising that the possible effects of this have not been investigated, as the increase in relative risk of those people to develop a common complex disease of late onset could substantially affect the global burden of disease. As an example, an increase in the relative risk of cancer among recently inbred humans of about 2.0, similar to that found in this study, would increase the global burden of cancer (using the simple formula for potential impact fraction) by (RR-1) x 12% (which is the conservative estimate of the prevalence of inbred individuals globally), i.e. by 12%, assuming that those people would survive to the age of disease onset.

This consideration becomes even more worrying when the evidence from animal studies and evolutionary and population genetic arguments are considered. In the last part of the introduction, it was implied through several lines of theoretical and empirical evidence that the effects of inbreeding in post-reproductive age would be expected to be considerably larger than in pre-reproductive age. This is mainly due to the invisibility of genetic variants underlying late-onset traits to selection, leading to their accumulation in the genome as a consequence. Due to these reasons, the study into the effects of inbreeding on biologically relevant quantitative traits (blood pressure, cortical index, biochemical measurements), learning disability and complex diseases of late onset (coronary heart disease, stroke, diabetes type II, psychiatric disorders, epilepsy, gout, asthma, peptic ulcer, osteoporosis) was undertaken in this work.

4.2. PROPOSED THEORETICAL MECHANISMS OF INBREEDING EFFECTS ON HUMAN LATE-ONSET TRAITS AND COMMON DISEASES

The genetic architecture underlying late-onset diseases such as cardiovascular diseases, cancer, adult-onset diabetes and psychiatric disorders, which represent the major health burden globally, is still a matter of open debate (**3,10,18,44,72,78**). A genetic model that is finding increasing support from both animal experiments and human studies is one in which the genetic variants underlying complex chronic diseases are more likely to be rare rather than common in the population. They are also likely to be numerous (a highly polygenic architecture) and of a small individual effect (**3,78**). If this view of the genetic architecture of common complex diseases is correct then it would be important to consider the predicted effect of inbreeding. I will propose five reasons why inbreeding may have a considerable effect on post-reproductive human health.

4.2.1. Deleterious Effects of Inbreeding Depression on Quantitative (Endo)Phenotypes that Confer Risk for Late-Onset Diseases May Be Multiplicative

Inbreeding depression is a recognised phenomenon that is common to polygenic traits in all living organisms (**189**). It is thought to result from increased homozygosity of recessive alleles that act in the same direction at loci that influence the phenotype of interest ("directional dominance") (**222**). In an inbred individual, inbreeding depression is predicted to affect many polygenic endophenotypes (quantitative [patho]-physiological intermediates involved in physiological or disease processes). Many of

these are established risk factors for late-onset diseases, such as blood pressure, body mass index, cholesterol and glucose levels. A substantial effect of inbreeding acting to increase human blood pressure has been shown directly in at least four studies (**195-197**), and implied indirectly in several more studies (**191-194**). Effects on body mass index and cholesterol levels have also been implied (**191,192,194**). Similarly, the effects on various measures of intelligence have been consistently shown (**174,175**). It is known that the risk of e.g. increased blood pressure, body mass index and cholesterol levels on cardiovascular diseases is not threshold-dependent, but increases across the entire range of values observed in the population (**201**). Thus, one important consequence is that inbred individuals are expected to be at slightly increased risk relative to the outbred general population to develop a late-onset disease, regardless of the absolute measurement of their blood pressure, body mass index and cholesterol. Furthermore, even if the effect of inbreeding depression on each of those phenotypes individually was rather small, it is known that the concurrent presence of several risk factors for the same disease increases risks in a multiplicative rather than additive manner. Therefore, the joint effect of inbreeding depression on all the potential quantitative phenotypes that confer the risk of late-onset disease during a lifetime could be more substantial than is widely appreciated (**112,113**).

4.2.2. Effects of Inbreeding on Rare Variants With Large Effect in Post-Reproductive Adults ("Invisible Mendelian Diseases" of Late Onset)

Inbreeding is predicted to have larger effects on the population-attributable fraction of disease if the underlying variants are rare rather than common. This is because common recessive variants will occasionally become homozygous in the population by chance, without a need for inbreeding to bring them together. If the variants are very rare in the population, and inbreeding is almost the only realistic scenario under which they can become homozygous in an individual, then the fraction of

disease cases in the population who are the offspring of related parents will be much larger. This was shown to be the case with the population attributable fraction of early-onset monogenic (Mendelian) diseases in the presence of inbreeding: it has been shown that the prevalence of autosomal dominant Mendelian disorders is constant in all world populations, but the prevalence of autosomal recessive Mendelian disorders is increased by 3-4-fold in regions where inbreeding is prevalent (**102,107**) Therefore, the great majority of Mendelian disease that is caused by rare recessive variants of large effect and early age of onset is due to inbreeding in those countries. However, these diseases manifest in the pre-reproductive period, so they are "visible" to selection. Although these variants continuously arise through mutations, most of the affected cases never reproduce, so they are being effectively removed from the gene pool by selection. Thus, their overall public health burden is reasonably low. Bittles and Neel estimated that each human carries about 1.4 such recessive lethal mutations in the genome (**13**).

However, rare variants of large effect also act in Mendelian fashion to cause late-onset complex diseases. For nearly all late-onset diseases, clustering in families has been reported and, in some, rare high-penetrance variants have been found which are associated with an extremely increased lifetime risk of disease (**10,18,44**). Examples of this include variants in BRCA1 and BRCA2 and breast cancer (**69**), hMLH1 and hMSH and colorectal cancer (**275**), and GCK (glucokinase) and maturity-onset diabetes of the young (MODY) diabetes (**276**). In large outbred populations, it is estimated that up to 15% of disease cases such as cardiovascular, cancer, diabetes and psychiatric disorders cluster in families, while 85% or more are due to combined effects of polygenic susceptibility and cumulative environmental exposures (**67**). However, in countries where inbreeding is common, the prevalence of all recessively inherited monogenic forms of complex diseases could be expected to increase by at least as much as seen jointly for all recessive early-onset Mendelian diseases, i.e. 3-4-fold (**102**). In theory, this could greatly increase both the overall disease prevalence and the proportion of the late-onset disease burden caused by rare recessive variants.

It is also known that these rare variants can affect quantitative phenotypes, such as blood pressure or cholesterol (familial hypertension, familial hypercholesterolaemias) (**77,191**), which represent "invisible Mendelian diseases", as their phenotypes are not clinically apparent. Most of the Mendelian diseases were defined by their apparent clinical phenotype, but the number of human phenotypes is endless, and it is possible that some genetic variants affect traits that are not obvious during the visual inspection of a patient, such as blood pressure or cholesterol levels. These quantitative phenotypes are associated with increased morbidity and mortality from diseases of public health importance such as stroke and coronary heart disease. It is conceivable that there are many more such "invisible" Mendelian diseases affecting metabolic pathways at different levels and predisposing individuals to complex diseases in the post-reproductive period. The joint effect of inbreeding on all of these variants could be expected to be of a magnitude of at least that seen for early-onset Mendelian diseases. Indeed there are reasons to believe that the number of rare and recessive Mendelian variants with large effect on late-onset complex diseases is much larger than the estimated number with early effects. These rare variants have accumulated in the genome through mutations that are either neutral in early life, or even beneficial, but show deleterious effects in post-reproductive period ("antagonistic pleiotropy") (**112,113**). There is no known mechanism that would be expected to remove these mutations from the genome or act against their accumulation, as they become deleterious in the post-reproductive period and therefore should be invisible to selection. Thus there are cogent reasons why the effects of inbreeding on late-onset Mendelian diseases should be carefully considered in the same way as has been done for early-onset diseases (**13**).

4.2.3. Autozygosity in Many Rare Recessive Variants of Small Effect Could Result in Epistatic Effects that Could Jointly Impair Capacity to Compensate Against Environmental Risks

Modest levels of inbreeding observed in human populations are expected to have much larger effects on the population distributions of polygenic traits than on oligogenic traits and diseases. This is because an excess in autozygosity of 6.25% of the genes in human genome (i.e. about 2,000 genes), which would be expected in a child from a first-cousin marriage, will lead to autozygosity of rare recessive mutations of small effect in those 2,000 genes. In a polygenic trait, it is expected that some of the genes that determine its expression would be affected even with only 6.25% of the genome autozygous. We have argued that the genetic component of late-onset diseases may be due principally to large numbers of rare variants in numerous genes - the common disease/rare variant (CD/RV) hypothesis (**3**). Recent estimates (**277**) imply that each person carries, on average, 500-1200 slightly deleterious mutations, most of which are rare and present in heterozygous form. Those mutations are more likely to also be recessive, as dominant mutations are continuously being removed from the gene pool by selection, while the rare and recessive mutations remain invisible to selection unless exposed by inbreeding (**3**). In an offspring of first-cousin marriage, 30-75 of these variants would be expected to become homozygous, with uncertain effects (**278**). If the mutations are numerous and of small deleterious effect, their autozygosity throughout the genome might not lead to apparent syndromes of early onset, but may mildly impair the function of affected genes. As a result, the compensatory potential to oppose the harmful environmental stimuli would be non-specifically impaired. This impairment of homeostasis or repair capacity could lead to an earlier age at diagnosis of a late-onset complex disease. This is consistent with the model of these diseases arising over long periods of time, and becoming clinically apparent when the compensatory potential is exhausted (**113**,**189**).

This theoretical mechanism is difficult to study in humans, but has been clearly demonstrated in animals, where a greater sensitivity of homeostatic mechanisms to inbreeding in later life has been suggested (**113,189**). In an experiment with inbred and non-inbred mice strains, the two strains did not show large differences in survival when the animals were kept in the laboratories. However, when both groups were released into their natural habitat, the inbred mice strains had a dramatically reduced chance of survival in comparison to non-inbred group (**279**). The effects of inbreeding are therefore thought to be much greater in natural populations (exposed to a less uniform and more challenging environment) than in those studied in laboratories.

4.2.4. Heterozygote Advantage in Loci Under Balancing Selection is Expected to Reduce by Inbreeding

At some genomic loci, there may not be variants present in a population that are clearly deleterious, but the heterozygous genotype may be more favourable than either homozygous genotype. This effect is widely known as the "heterozygote advantage", "hybrid vigour" or simply "heterosis". The effects of heterosis usually act in an opposite way from those of inbreeding depression and they have been demonstrated in humans (**280,281**), and widely in animals and plants (**282,283**). The type of selection that tends to maintain more than one allele in the population at intermediate frequencies, thus maximising the frequency of heterozygous genotypes in a population, is known as "balancing selection" (**3**). It is clear that balancing selection probably has an important role in shaping diversity in the genes that are important for defence against unknown and unpredictable environmental risks, such as infectious diseases (**280**). Populations that are more genetically diverse are at less risk from diverse environmental threats (since it is more likely that someone would carry a rare protective variant) (**284**). It is therefore likely that inbreeding leading to autozygosity in several hundred genes will affect some of these genomic loci under balancing selection, thus reducing the beneficial effects of heterosis in those individuals. This mechanism could be more important than generally

thought, since recent evidence suggests that loci under balancing selection may be surprisingly common in the genome (**285-288**).

4.2.5. Empirical Evidence of Inbreeding Effects in Humans and Other Organisms

The most extensive research into the effects of inbreeding in general, and particularly on genetic variation related to senescence has been carried out in *Drosophila spp.* A review of 25 years of this research has concluded that deleterious alleles generated by mutation and kept at low frequency by selection contribute between 33% and 67% of the genetic variation in a typical fitness-related trait. This supports a polygenic model of genetic architecture of most phenotypes and suggests that the common disease / rare variant mechanism contributes to a substantial share of complex disease aetiology (**222,289**). A recently published experiment in *Drosophila spp.* (**112**) showed that genetic variation and inbreeding effects increase dramatically with age, supporting these hypotheses. Numerous recent studies of other animals, some of them performed in populations of large mammals, have also consistently reported that inbreeding negatively affected key components of fitness, resulting in increased morbidity and decreased life span (**290-293**). A meta-analysis (**294**) and a critical review and re-examination of these studies (**287**) have both concluded that, although unexpected and in some aspects against current understanding, their findings could not be easily dismissed on grounds of publication bias or apparent flaws in methodology.

Finally, my own intensive review of the human literature since 1965 has managed to identify very few case-control studies of late-onset diseases in which inbreeding status was not determined by self-reporting, and disease status determination was based on a clear diagnostic criterion which did not change during the period of study. These studies investigated the effects of inbreeding on coronary heart disease (**205,208**), cancer (**209, 210**), psychiatric disorders (**213**) and Alzheimer's disease (**216**).

All those studies reported considerable relative risks associated with inbreeding, typically between 2.0-5.0, which persisted after adjustment for known or suspected confounding factors. Although the available evidence is surprisingly sparse, it appears to support the hypothesis that inbreeding could have a considerable effect on human health and disease occurrence in post-reproductive age adults. However, it is also possible that large publication bias could be responsible for such conclusion: the early reports are, as a rule, more likely to over-estimate the effect size, and the subsequent lack of publications and scarcity of available data could also imply that those initial findings were difficult to repeat.

4.3. SUMMARY OF THE INVESTIGATED EFFECTS OF INBREEDING ON HUMAN QUANTITATIVE TRAITS AND LEARNING DISABILITY

4.3.1. Blood Pressure

It is widely recognised that essential hypertension is under considerable genetic influence. However, apart from isolated successes in mapping rare monogenic loci, which account for less than 5% of hypertension, no major progress has been made in defining the genetic basis of essential hypertension (**295**). A common, often implicit, assumption in mapping studies of such complex traits is that relatively few genetic loci of moderate to large effect account for a large component of the underlying genetic variance despite the paucity of empirical data to support this. This study has provided an insight into the possible highly complex genetic architecture of blood pressure variation, showing that the effects of recessive QTL on BP are likely to be widespread, accounting for 10-15% of the total variation in BP in this population. These effects are attributable to a very large number of loci (at least 300-600) which will almost certainly show a range of effects on BP.

The model used for the calculation of the estimated number of QTL that underlie blood pressure levels (see **Sections 1.2.5. and 2.2.4.**) makes several assumptions which may influence these estimates. Firstly, the inbreeding coefficient is based on measures of recent inbreeding (over 4-5 generations). We therefore calculated isonymy estimates for each village (**Table 1**), and found that their median value exceeded the median F value by a factor of 1.53. Since isonymy is widely recognised to over-estimate inbreeding (**259**), this represents an upper bound to the inbreeding estimate. This is plausible, as it has recently been shown by mathematical simulations that individual inbreeding coefficient estimates based on 4-5 ancestral generations are expected to capture 85-90%

of expected autozygosity, regardless of the possible ancestral relatedness earlier in the past (**287**). Inflating the F values by 153% in the model, which is certainly an overestimate of individual F values, would reduce the estimates of minimum QTL numbers by a factor of 1.53^2 (=2.34). On the other hand if, as seems likely, the ratio of dominance variance (V_D) and total phenotypic variance (V_P) is less than the assumed 100%, the effect would be to increase the estimates by a factor (V_P / V_D), which could be 2-4 times. Cavalli-Sforza and Bodmer (**296**), for example, estimated V_D/V_P to be 0.38 and 0.33 for systolic and diastolic blood pressure respectively, using the data of Miall and Oldham (**297**), which would double or even triple the estimates in this study.

Secondly, as in many genetic models, all loci were assumed to have equal effects, whereas both theory (**298**) and empirical data in animals (**299**) show that the QTL effects vary widely and may even be of opposite sign. Ignoring this fact would again result in under-estimating the true number of QTL. Thirdly, the model assumes bi-allelic loci. However, for multi-allelic loci the estimate of minimum QTL numbers is unaffected, except in the unlikely situation that high degrees of over-dominance occur at all, or most, loci (**189**). Fourthly, if a substantial proportion of the phenotypic variance is due to epistatic effects, additive and dominance variances may be upwardly biased, with an uncertain effect on the estimated number of QTL. However, many studies suggest that epistatic QT effects are uncommon (**299**), although some have recently expressed a different opinion, suggesting that epistatic effects are too often neglected (**278**). It is fair to say that the magnitude of epistatic effects is still uncertain, but regardless of its importance, it should not affect the overall conclusion of this study (i.e. that the genetic basis of blood pressure variation is polygenic and highly complex). Fifthly, this study assumed that recombination between adjacent loci was sufficiently frequent that the IBD status of any locus could be considered independent of its neighbours. In effect, the method treats tightly-linked loci as a single "super-locus" (**300**), leading to further underestimation of the true number of loci.

Finally, the model assumed random mating within the total population of 25 villages, although this criterion was certainly not met due to population subdivision. The studied population was in fact a "metapopulation", composed of a larger number of isolated small populations. However, village of residence showed a strong independent effect on BP suggesting either village-specific patterns of inbreeding or other local effects. Multiple regression performed separately in each village population revealed that the mean within-village slope (for the regression of F on BP) did not differ from the pooled-village slope for either systolic or diastolic BP. This would seem to support a common disease/common variant model (**60**) in which individual villages share a common genetic diathesis influencing BP. Random samples representing about 20% of each village did not seem to underestimate the total number of QTL within the whole population. An alternative explanation is that, due to the metapopulation nature of the investigated sample, each village had an excess frequency of at least a few recessive variants that are village-specific and otherwise rare, and had moderate to large effect on blood pressure. When "exposed" by inbreeding, those variants could be responsible for the observed increase of blood pressure values among inbred individuals within each village.

The magnitude of the inbreeding effect on BP is large (equivalent to a rise in systolic BP of ~20mm Hg and diastolic of ~12 mm Hg in the offspring of first cousin marriages; F=0.0625). This increase is, however, very similar to estimates of inbreeding effects on diastolic (**195**) and systolic (**196**) BP in other isolate populations. One of the explanations for the large observed effect may be because inbreeding has a greater influence on late-onset traits than on traits that are subject to early selection (**222**). It is also possible that low environmental variation, or underestimation of F due to individuals being related through multiple lines of descent, contribute to the size of inbreeding effect in this isolate population (**194-196,202**). Finally, the effect of inbreeding in a "metapopulation" may be much stronger than in a non-subdivided, outbred population, as greater allele frequencies of rare alleles of large effect and their expression are possible in each isolate through interaction of founder effect and

subsequent inbreeding and genetic drift (**283**). The mechanism is the same in each isolate, although it affects different variants: initially, a small group of founders introduces a number of very rare and deleterious variants from the general population, in which those variants were insignificant for the overall disease burden. However, in the newly created gene pool of an isolate population those variants become common, and they may even continue to increase in frequency over generations due to genetic drift. Their effects in a population are also considerably more pronounced because inbreeding, characteristic of human isolate populations, increases the chance for their homozygosity. Therefore, in population of such genetic structure (many rare alleles brought to common frequencies in individual villages) the effect of inbreeding on the population-attributable fraction of any trait or disease is predicted to be much stronger. All these mechanisms may help explain the observed effect size of inbreeding, which may be lesser in more environmentally diverse or outbred populations. The unidirectionality of the effect is also striking and consistent with a linear unidirectional effect seen in an S-Leut isolate population (**196**), but the mechanism is unclear. A change in BP with inbreeding is predicted as a consequence of recessive or partially recessive variants with the direction of change towards the value of the more recessive alleles. Physiological homeostasis may also act to support a directional change in BP, for example through selection against variants tending to reduce BP in order to maintain circulatory viability.

The estimate of several hundred QTL relevant to human hypertension is realistic and indeed conservatively low. Some additional theoretical work by Carothers (**Carothers A, personal communication**) showed that this general conclusion remains unchanged even under a wide variety of assumptions related to the distribution of size effects of QTL, e.g. even under the assumption of the presence of a number of loci with large effect. The conclusion of a highly polygenic genetic architecture of blood pressure regulation is consistent with a complex and genetically highly variable (**301**) system of blood pressure control mediated by cardiac output, blood vessel architecture, renal function and CNS integration, and requiring the interaction of homeostatic systems, including baroreceptors, natriuretic peptides, renin-angiotensin-aldosterone, kinin-

kallikrein, adrenergic receptors, and local vasodilator mechanisms (**295**). Furthermore, published work from animal models of hypertension support a polygenic rather than oligogenic basis for hypertension (**295**) and yet these probably underestimate the genetic complexity, since they are typically bred towards fixation of a small subset of the diversity found in wild populations (**300**). The greater genetic complexity of a diverse and outbred human population would seem to be self-evident, despite the fact that humans show less haplotype and polymorphic diversity than several other species, including other primates (**29**).

The present study demonstrates an important effect of inbreeding on the genetically complex late-onset disorder, hypertension, which appears to be mediated by a large number of recessive QTL alleles as a result of increased homozygosity. Several factors support the validity of the data and reinforce the conclusions. Firstly, the standard measurement procedures adopted and exclusion of known confounding factors, coupled with the fact that hypertension was untreated in this historic sample. Secondly, the consistency of findings in diverse populations (**194-196**). Thirdly, the linear increase in BP with increasing *F* (prevalence of hypertension rises by 10% for every increment in *F* of 0.01 up to F=0.06). Fourthly, the overall strength of the effect. Fifthly, the existence of biologically plausible mechanisms, all of which point to a causal relationship between inbreeding and hypertension. Moreover, the consistency of the observation in a random sample of individuals across 25 villages is not explicable by a kinship effect. This is to say, if the effect of inbreeding on hypertension was observed in a single isolate population, it could be argued that there is clustering of hypertension among the relatives, who are also more inbred, so the observed effect was that of familial clustering, and therefore possibly even environmental. However, my results showed that across 25 different villages on 3 different islands the inbred individuals are always those who are also more likely to be hypertensive, which cannot be attributed to familial clustering. In terms of health impact, the results show that 36% of the hypertension incidence in this population can be attributed to inbreeding (population attributable fraction). The population prevalence of hypertension among non-inbred individuals is

approximately 20%, similar to most outbred populations, but it increases steeply among 50 year olds as the inbreeding coefficient rises.

4.3.2. Cortical Index

One of the ways to support the conclusion that the increasing effect on blood pressure was indeed due to inbreeding was to investigate the inbreeding effect on another quantitative trait unrelated to blood pressure, using a similar design in the same population. Cortical index was chosen as that trait can be measured objectively, and it represents the most important intermediate phenotype for a common late-onset disease, i.e. osteoporosis. A subset of the same population (14 of 25 villages) used for the blood pressure study in which cortical index was also measured was investigated. The subdivision of the islanders of the eastern Adriatic into small villages and diversity of their attitudes towards inbreeding influenced by geographic isolation, political privileges in the past, as well as socio-cultural reasons, resulted in a range of inbreeding coefficients at both the individual and population level. This has already been shown as a favourable setting for the blood pressure study.

The bone X-rays were performed in local village health clinics during the field work in 1970s and early 1980s, using the same device and computed from the scans by the same assessor from the Institute of Anthropological Research in Zagreb. This was performed at the same time as blood pressure measurements were being taken. Therefore, the cortical index represents an objective measurement of an intermediate disease phenotype. An attempt was also made to avoid the significant confounding effects of genetic drift and founder effect in specific villages. This has been achieved by including several villages of similar inbreeding levels from different islands into the groups with "high", "moderate" and "low" inbreeding, as it is unlikely that the two population genetic phenomena would affect gene frequencies in the same direction in all villages.

The results showed a significant decreasing effect of inbreeding on cortical index (initially standardised by age) across 14 villages (**Figures 3-3** and **3-4**). Similarly, the prevalence of osteoporosis found in these villages appeared to increase from about 1.3% in villages with low inbreeding prevalence to nearly 4.0% in villages with average inbreeding coefficient close to F=0.05, i.e. a 3-fold increase (**Figures 3-5** and **3-6**). Analysis by village showed that the standardised prevalence of osteoporosis follows the increase in average inbreeding coefficients, although this correlation is not really linear as there are villages with an extremely increased prevalence in both sexes (e.g. Cara, with an OP prevalence of about 10%). In addition, not all villages with higher rates of inbreeding reveal a high prevalence of osteoporosis (e.g. Pupnat in males). These observations might be due to the specific genetic structure of those villages which are characterised by increased or decreased frequencies of rare alleles with large effects mediating susceptibility to osteoporosis, possibly due to the combined effect of genetic drift and founder effect.

If we accept the conclusion that the increase in average inbreeding coefficient of about 2-3% (from F=0.005 to 0.03) could be, to some extent, responsible for the observed increase in prevalence of osteoporosis, the central question becomes what does it tell us about the genetic basis of the disease? Similarly as in the analysis of effects of inbreeding on hypertension, the susceptibility seems to be controlled, at least partly, by recessive genetic variants that are more likely to be rare than common, as the inbreeding effects on rare variants are more apparent. Major effect genes arise after mutations that are considered to be very rare, as the probability of random mutation causing small effect is much greater. Therefore, even if such mutations were present in some of the studied villages, it is extremely unlikely that similar effects of inbreeding would be observed across several villages, as our results indicated. The results provide further evidence on a polygenic basis of susceptibility for osteoporosis with detectable effects of recessive genes. The finding of the significant effect on both blood pressure and cortical index, i.e. two seemingly unrelated quantitative traits, reinforced the conclusions that

inbreeding is indeed responsible for the observed effects. An analysis of the association between inbreeding and cortical index is planned at the individual level to further strengthen these findings.

4.3.3. Learning Disability

The study of the effects of the average inbreeding coefficient in 10 communities on the prevalence of learning disability further reinforced the findings of the effect on blood pressure and cortical index. The association between inbreeding coefficient and learning disability has been consistently demonstrated and repeated in more than 20 independent studies performed in very different populations and time periods, and the findings in the Croatian island isolates are in line with all previously reported results (**174,175**). The consistency between the effects observed on blood pressure, cortical index and learning disability both between themselves and with the results of other studies supports the hypothesis that the "metapopulation" of Croatian island isolates is particularly favourable for investigating the inbreeding effects on human quantitative traits and complex diseases.

4.4. SUMMARY OF THE INVESTIGATED EFFECTS OF INBREEDING ON HUMAN LATE-ONSET COMPLEX DISEASES

The extensive literature on the health effects of inbreeding has not examined late-onset phenotypes such as hypertension to the same extent as traits such as reproduction, childhood mortality or Mendelian disorders (**13,106**). In other species, an increase with age in the deleterious effects of inbreeding has been reported. It has been suggested that this may be due to increasing sensitivity to genetic effects with advancing age caused by loss of selective elimination of deleterious effects on homeostatic mechanisms (**113**). In this research, the studies conducted on blood pressure and cortical index indicated that inbreeding may affect biomedically relevant intermediate phenotypes that are considered major risk factors for late-onset complex diseases. It was intriguing to evaluate the inbreeding effects on the prevalence of most common complex diseases of late onset in Croatian island isolates.

The impact of inbreeding on reproduction, childhood mortality and Mendelian disorders is well-documented (**189,222**). In contrast, very little has been published on the effects of inbreeding on late-onset diseases. This is despite the fact that inbreeding may have a greater influence on late-onset traits than on traits that are subject to early selection (**112,113**), as discussed earlier. This study demonstrated an important effect of inbreeding on a number of genetically complex late-onset diseases that are of major public health importance. The results are consistent with the hypothesis that a major genetic influence on these disorders is mediated by numerous deleterious recessive alleles, suggesting that inbreeding increases disease risk as a result of increased homozygosity (**3**).

It is important to consider again whether these results can be explained by chance, bias or confounding. The effects of inbreeding were expected, as this was the main *a priori* hypothesis (based on four different theoretically possible mechanism and empirical evidence in other species, discussed in **Section 4.2.**). This taken together with the levels of statistical significance reported argues strongly against chance as an explanation for the observed findings that consistently showed a significant inbreeding effect on nearly all of the investigated late-onset diseases.

With respect to possible selection bias when inbreeding effects were investigated at the individual level, the 480 individuals on whom it was possible to obtain disease outcome data were a subset of the original cohort from the 1970s and 1980s. However, census data revealed that the major reason for loss to follow-up, other than deaths of cohort members, was emigration from the villages over the 20-year period. It is presumed that this emigration was random in terms of the genetic structure of the remaining population, so the selection of the sub-sample of the original population should not result in substantial bias.

Measurement bias is also unlikely, as disease outcomes were not ascertained or recorded differently among individuals who differed by inbreeding status. Explicit, standard clinical criteria were adopted by the two study doctors. They were both blind to the inbreeding status of all the individuals - in fact, they systematically scanned the complete medical records of all current village populations and did not have any information on who participated in the initial field work 20 years ago. Furthermore, the results cannot be explained by different diagnostic practices in different villages, as the village term was not found to be statistically significant in the multiple logistic regression analysis (except for coronary heart disease).

As chance and bias are unlikely explanations for the observed findings, the potential confounding effects also need to be addressed. A number of potential confounding factors (age, sex, smoking status, education level, general diet,

occupational group, height, weight) were measured and their effects adjusted for in the multivariate analysis. Although a degree of imprecision is inevitable in measuring some of these factors, we do not believe that confounding could have accounted for the large and consistent effects demonstrated.

There is also a concern that there could have been an interaction between age and inbreeding, and that older people in the population were more likely to be inbred and ill. However, the multiple regression model used in the analysis investigated this correlation and corrected for it in all cases. The correlation was generally not significant in the large majority of the analyses, showing that inbreeding was still prevalent in all age groups included in this historic cohort. Also, the lack of the statistical significance of the effect of age on some diseases where it would be expected, e.g. stroke, shows that there are other independent factors in this metapopulation that may contribute to earlier age of onset of stroke patients: inbreeding (through increase in blood pressure levels) and smoking.

Several factors support the validity of the data: firstly, the findings support our prior hypothesis and are consistent with similar findings on hypertension in the same population and with other reports of inbreeding effects on blood pressure (**191-196**). Secondly, the overall strength of the effect argues against bias or confounding. Thirdly, we have presented detailed arguments that biologically plausible mechanisms underpinning this effect are consistent with population genetic theory and observations in a wide range of organisms (see **Section 4.2.**). Finally, the data are consistent with the few other published reports of inbreeding effects on late-onset diseases in which inbreeding was measured rather than self-reported (**205-219,302,303**).

The magnitude of the inbreeding effect on disease prevalence was large. However, the effect on the prevalence of stroke and coronary heart disease, for example, is consistent with our estimate of a rise in diastolic BP by 2mmHg for an increase in F of 0.01. Both cohort studies and randomized trials show that an increase of 5mmHg

diastolic BP is associated with a 33% increase in stroke risk and a 20% increase in risk of ischaemic heart disease (**304**). The large effect may be due to the greater influence of inbreeding on late-onset traits than on traits that are subject to early selection (**112,113**). It is also possible that low environmental variation, underestimation of F due to individuals being related through multiple lines of descent, and "metapopulation" structure contribute to the size of inbreeding effect in these isolate populations, which has all been discussed earlier. Thus, the magnitude of the inbreeding effect relative to the overall variation may be smaller in outbred and more environmentally diverse populations.

The effect of inbreeding was shown in 7 of the other 8 diseases studied, with epilepsy and schizophrenia revealing too few cases for an analysis in a multiple regression model. The lack of observed effect of inbreeding on type 2 diabetes may reflect the lower heritability and stronger environmental influences involved in the aetiology of this condition (**189,222**). The observed effect of inbreeding on the prevalence of several different late-onset diseases is consistent with the presence of many deleterious recessive alleles, located throughout the genome. It is also consistent with a more general effect of inbreeding with increased homozygosity at these loci leading to an accumulation of small deleterious effects on homeostatic pathways, which cumulatively increase disease risk. This suggests a greater sensitivity of homeostatic mechanisms to inbreeding in later life, as predicted by findings in animals (**106,113**). Decay of homeostatic capacity would also be expected to lead to reduced capacity to respond appropriately to diverse stimuli. This is supported by the recently reported observation that the reduced survival found in inbred animals is greater in the natural habitat than in a controlled laboratory environment (**279**).

The inbreeding data do not allow an easy distinction to be made between the relative contributions of common versus rare variants but do inform two somewhat neglected aspects of the genetic architecture underlying complex diseases (**3**). Firstly, the results provide indirect evidence in support of a major polygenic component to

disease susceptibility. The inbreeding coefficient is shown to be a significant predictor of CHD, stroke, cancer, depression, asthma, gout and peptic ulcer with population attributable risks varying between 23% and 48%. An upward bias affecting these estimates is possible, as only a few covariates were available for this historic cohort, and inbreeding was among those few variables included in the regression model, which was likely to increase its overall importance.

Secondly, the recessive or partially recessive nature of complex disease susceptibility has received little emphasis. Both factors have implications for the identification of individual disease susceptibility alleles. If disease susceptibility is indeed highly polygenic then it implies the necessity to reduce the phenotypic complexity of "disease" by means of genetically simpler but contributory quantitative traits (QT) or disease subgroups. Those with extreme values of QT distributions or early disease age-of-onset will be those most likely to harbour susceptibility alleles of large effect and hence to provide a realistic possibility of gene identification. Secondly, a significant component of genetic susceptibility appears to result from variants that are recessive or partially recessive. This implies that the study of inbred populations would be advantageous since the increased gene dosage of such variants in inbred individuals will tend to amplify their phenotypic effects compared with outbred populations, where most alleles are present in heterozygotes.

The population attributable risk estimates from this study suggest that 23-48% of the incidence of the disorders showing an inbreeding effect in this population can be attributed to inbreeding. The study of inbreeding effects on blood pressure showed that 36% of hypertension incidence in this population could be attributed to inbreeding. Those estimates are likely to be biased upwardly, and should probably not be expected in general populations, for a number of reasons already discussed: (i) very few covariates were available for investigation from the historic sample, which is likely to overestimate the inbreeding effect; (ii) a "metapopulation" nature of the studied population, which has a strong genetic sub-structure and therefore the effects of

inbreeding are expected to be stronger; (iii) lack of understanding of a possible effect of "background inbreeding" in the most isolated and the smallest villages, specific to the entire village population, which would imply that the inbreeding coefficients in those villages are actually underestimates, and therefore the observed size effects of inbreeding are overestimates.

4.5. SUMMARY OF THE INVESTIGATED EFFECTS OF ADMIXTURE AND OUTBREEDING ON HUMAN HEALTH AND DISEASE

The global human population has experienced dramatic changes in population genetic structure on both regional levels and at the global scale relatively recently. After having been organized in small and sparsely scattered isolate communities tied to the land they harvested throughout their entire history, enormous changes have occurred within the last 5-6 generations. Primarily, measures to reduce childhood mortality (vaccination, antibiotic treatment of infections, improved nutrition and sanitation) have led to an unprecedented increase in population size, from about 1 billion (in 1850) to more than 6 billion (in 2000). This relaxing of selection that kept the human population size reasonably constant for centuries is predicted to help in retaining new rare genetic variants introduced through mutations, rather than efficiently eliminating them by genetic drift in sub-structured populations (**3**). Therefore, increase in population size is the likely cause of the recently increased genetic diversity of contemporary human populations in comparison to those that lived only two centuries ago. The effects of population expansion on the changes in current genetic diversity are expected to be largely due to the introduction of rare and recent mutations that are being retained in the absence of genetic drift.

Apart from large and sudden population expansion, admixture and outbreeding, there are other population genetic mechanisms that have occurred relatively recently and on a massive scale. The organization of humans in small rural communities limited mate choice, and there are several reasons to believe that inbreeding was common throughout human history. First, in our contemporary set of isolate populations used for the admixture and outbreeding analyses (the 10 village dataset), all of them reasonably large

in terms of population size (Npop=800-1500), it was apparent that about 10% of all individuals were inbred, although inbreeding was generally discouraged in these Roman Catholic communities. In smaller villages, such as those found throughout history, and with a deeper look into genealogies, it is certain that this percentage would be considerably greater. Strong positive selection (through high childhood mortality) of infectious diseases on rare variants that were shared by relatives in closed communities was causing a so-called "selective sweep", where genetic variation in the population was being decreased through the constant positive selection of relatives carrying rare protective alleles (**305**). Finally, up to year 1900 about 98% of the World's population lived in villages, and even today up to 2 billion people globally are living in areas with a considerable prevalence of consanguineous marriages (**102,306**). The process of urbanization that took place in both developed and developing countries suddenly shifted a substantial proportion of the human population from villages into the cities. At first, there is often tendency of the migrants to preferentially marry people from the same or very similar background, with so-called "chain migration" a very common phenomenon. However, eventually, their offspring will mate with individuals of a genetically different background, which is predicted to cause a massive outbreeding at the global scale.

Finally, international travel and large inter-continental migrations of people during the second half of 20th century led to the third effect on a global scale: "gene flow" and "admixture". The effects of urbanization at regional levels are now repeated through international migrations and mating between people from different origins at the global scale. Taking into account the relatively sudden occurrence and enormous magnitude of all four effects - increase in population size, outbreeding, gene flow and admixture - it is very surprising that the possible public health effects of changes in genetic structure of populations have hardly been addressed, let alone investigated. A careful review of the literature performed during 2003 implied only a handful of attempts to address various elements of this hypothesis (**307-309**).

The World Health Organisation recently defined major disease risk factors in the developed world that attribute most to the disease burden in the population (**190**). Apart from smoking, other major risks include increases in body mass index, blood pressure, cholesterol levels and blood glucose, all of them readily measurable quantitative traits. There is a plausible theoretical argument why inbreeding (decreased heterozygosity) and outbreeding (increased heterozygosity) should cause changes in mean population values of quantitative traits, as discussed in **Section 4.2.** (**3,189,310**). Associations between heterozygosity and biologically important quantitative traits and their cumulative effect on fitness have been convincingly demonstrated in a number of plant and animal populations (**279,291,292,311-315**). However, we could only identify a handful of studies in human populations that provided any data on these effects in post-reproductive age (**196,316,317**). The only important human quantitative trait that has been frequently associated with decreased heterozygosity is intelligence, with a reasonably large number of supportive studies from various parts of the world showing strikingly large effects on the trait (**174**).

Therefore, in this study our main aim was to investigate if changes in population structure could be associated with shifts in the population distribution of quantitative traits identified as major disease risk factors, such as blood pressure, cholesterol and triglycerides. However, an even more scientifically intriguing aim was to investigate if the effects of appreciable magnitude predicted from evolutionary and population genetic viewpoints could be observed in a present-day human metapopulation: effects of heterosis on fitness, mate choice and height. The study mainly suggested that inbred individuals generally had higher pathologic values of biomedically relevant QTs than the examinees of more diverse genetic background. This could be explained through the action of numerous rare and recessive variants of slightly deletorious effect that have not been selected against in childhood, as they mostly affect the variation in late-onset traits. Therefore, those variants could reach appreciable frequencies at numerous quantitative trait loci and show a non-specific pathologic effect on a number of QTs even in the presence of small levels of inbreeding, as it was in our study. Also, heterosis seemed to

positively influence height and increase fitness by 10%. Although seemingly small, this rate of increase in fitness tends to rapidly increase the population share of outbred individuals with each new generation, while keeping the absolute number of inbred and endogamous individuals constant.

These observations are likely to be genuine, as the standardization excluded a number of potential biases and confounding effects present at the level of between-population comparisons. Standardizing the individuals in each category (inbred, endogamous, admixed and outbred) by gender, age and village of residence eliminated potential biases such as seasonality, assessor-related differences, isolate-specific effects of environment, and any differences in data collection between populations. Methods to measure inbreeding from genomic information are currently being developed, to support the findings reported here with genomic measures of gene diversity (at population level) and average heterozygosity (at individual level) calculated from genome-wide scans using several hundred of microsatellite polymorphisms (**318**).

4.6. CRITICAL ASSESSMENT OF THE RELIABILITY OF THE FINDINGS

The potential sources of chance, bias and confounding specific to the design of each individual study on inbreeding effects presented in this thesis (i.e., hypertension, osteoporosis, learning disability, complex diseases of late onset and several quantitative traits) were discussed under appropriate headings. This is because study designs, selected samples and methods used in each one of those studies differed considerably, and they were all selected appropriately to the study question. However, there are also considerations related to the validity and reliability of the findings in this thesis that are more fundamental and therefore relevant to all five individual studies that were undertaken. They could all provide alternative explanations to the observed results, and therefore should be carefully considered.

Firstly, there is a concern over the causes of diversity of attitudes towards inbreeding in different villages. If this was systematically influenced by an unmeasured variable, that variable could perhaps also correlate with all the traits that were measured. One of such variables is, for example, clerical approval of the marriage between the close relatives. If some villages systematically differed from the others, that could point to the general role of the cultural factors, such as more intense role or practice of religion in some villages, contributing to better health over time. Still, we consider this factor highly unlikely, as the records showed that the priests generally migrated between the villages quite often, or were in charge of religious practices in a number of villages at the same time, that it is difficult to believe that some villages differed in inbreeding practices from the others consistently and during larger periods of time due to clerical (dis)approval of consanguineous unions.

To extend this argument, there is always a concern in any epidemiological study which is ecological in design on whether the chosen predictor variable (i.e., preferential and random inbreeding, due to village endogamy) is really the sole parameter under investigation, or does inbreeding act as some form of surrogate measure for a composite collection or other mainly non-genetic variables. For example, when comparing inbreeding levels and effects within a community, it is possible that there were differences between consanguineous and non-consanguineous families, especially in terms of disease prevalence. For example, certain families could more commonly marry their cousins because of recognized "frailty" that effectively excludes them from the wider marriage pool. This may indeed be the case, but the extent of its effect is very difficult to estimate. I believe that, under the specific circumstances (e.g., using the historic cohort sample), as much has been done in terms of identification, definition and inclusion of the covariates into the analyses as realistically possible. Another argument in favour of the inbreeding effects being real is the consistency of the observed effects on the wide range of unrelated outcomes using tailored study designs, varying samples and different methods, coupled with the support for these effects in the literature that studied the effects of inbreeding in animals and plants.

Then, there is a bias in this thesis towards only considering and discussing the negative health effects of inbreeding, although one may also anticipate the positive ones, assuming these exist. For example, it is unclear why should the increased levels of identity by descent apparently focus on the expression of genes with this increased, and not on reduced blood pressure. One explanation of this unidirectional effect could be the so-called antagonistic pleiotropy, e.g., that factors that contribute to greater reproductive success among inbred couples early in life also make them more susceptible to diseases later in life. I will try to explain why this was the case. However, I would argue that the explanation of the unidirectional and generally negative effects of inbreeding has to do with a metapopulation nature of the studied population, the intra-village rare gene specificity and the population sub-division, leading to the Wahlund effect. When the Wahlund effect is present, inbreeding values are affected by internal geographic

subdivision of the population, with social and cultural factors, such as avoidance of or preference for consanguineous marriages, being less important. Now, let us consider the effects of population subdivision and structuring: the founder effect will immediately become important, and gene frequencies will start to change in each particular isolate and to diverge from each other. However, the only net effect consistently present under this scenario is that all new mutations arising over time within each particular isolate will actually be increasing in frequency in all the isolates, due to effects of genetic drift in small population and the absence of gene flow. Those frequencies of newly introduced mutations will, in time, reach the levels that are orders of magnitude greater than it would b possible in large general populations with constant gene flow. Although the effects of those new mutations depends on the interaction with the particular environment, and some of them may be beneficial and others silent, it is nevertheless predicted that the large majority of those new mutations would still be slightly deleterious. When such metapopulations practice inbreeding, they "expose" the effects of otherwise silent recessive mutations, both "old" (founder) ones and "new" (isolate-specific) ones, the large majority of which is predicted to be deleterious. So, although the mutations leading to beneficial effects must also exist in these isolates, the majority of the observed effects on health-relevant traits across all the studied isolates are expected to be negative and unidirectional.

In considering the disease prevalence data, there is always a concern that none of them are mainly determined by either inbreeding or the environment, as there is always a possibility of the presence of unmeasured factors, such as e.g. possible influence of prenatal factors on the conditions investigated (also known as "the Barker hypothesis"). The Barker hypothesis states that there may be fetal origins of diseases that are important in adulthood. The hypothesis is primarily based on reported epidemiological associations between indicators of fetal malnutrition and mortality and morbidity recorded during the adulthood. The associations were reported between birth weight and coronary heart disease, stroke, hypertension, type 2 diabetes mellitus, insulin resistance and serum lipids. This association appears even stronger in developing countries of the

world (**319,320**). It is not possible to speculate about the Barker hypothesis as an alternative explanation of the findings in this thesis, but given that genomic imprinting was suggested as one of the possible mechanisms contributing to those effects, perhaps it would be an interesting line of research in the future to study how inbreeding affects genomic imprinting in fetal period and possibly affects the predictions under the Barker hypothesis.

There is a general concern about the structure of the emigrants and the immigrants that could affect the design of parts of this thesis. In studying the effects on late-onset diseases, the validity of the assumption that post-1980 migration from the islands was random is questionable, as it may be possible that the migrants were better educated, healthier, more ambitious or conversely less advantaged in their lives on the islands. In general terms, emigrants are very seldom truly representative of those who either elect or are forced by circumstances to remain sedentary. Therefore, the presented results could only be valid for a group of people that remained, and not necessarily for those who left the island. The similar concern extends to the selection of the 2^{nd} generation immigrants from the Croatian mainland as an outbred control group in the study of outbreeding effects on a number of quantitative traits.

Finally, a fundamental question remains: if outbreeding is a universally more successful mating strategy than inbreeding, then why does consanguinity continue to be favoured by many populations? I believe that the preference of inbreeding practices has strong cultural, social and economic roots, and that in the poor communities its benefits far outweigh the costs. Knowing that the health risks of inbreeding in pre-reproductive age were shown to be modest at the level of population, and negligible at the level of individual couples, it is not surprising that the harsh economic realities or deep cultural traditions still motivate large number of people globally to choose spouses among their close relatives. In those communities, the overall life expectancy is usually too low to even consider the possible risks of inbreeding in post-reproductive period. However, there is a balance of advantages and disadvantages offered by inbreeding under different

socioeconomic circumstances, and it is likely that with economic improvements and urbanization of the poor regions of the world, we may expect inbreeding practices to slowly become abandoned in the segments of the society that accept an active role in economic transition, as was the case in several countries in South America, Asia and north Africa, where inbreeding was relatively prevalent before the economic transition.

4.7. WIDER IMPLICATIONS OF CONDUCTED RESEARCH

4.7.1. Inbreeding and Global Burden of Disease

Inbreeding is generally decreasing among non-immigrant Western societies but it is still prevalent globally. In **Section 1.2.2.**, it has been conservatively estimated that about 12% of the total human population are the progeny of consanguineous marriages, defined as unions between individuals related as second cousins or closer (equivalent to $F \geq 0.0156$ in their progeny). Further analysis showed that the average F-value in those 12% of individuals is expected to be at least about 0.03. Based on the recently revised estimate of 25,000 genes in the human genome (**168**), this suggests that those individuals, in their respective populations, have an excess of nearly 800 autozygous genes. The effects of these genes on their health in older age are uncertain, and the theoretically predicted consequences were explained in **Section 4.1.**

Let us now consider the implications of the estimated inbreeding prevalence on the global burden of Mendelian diseases first. It is known that there are thousands of genes in which mutations of large effect can lead to Mendelian syndromes in early life. The joint effect of inbreeding on all of them is a 3-fold increase in the prevalence of autosomal recessive syndromes in World regions characterized by inbreeding (**102**), while the prevalence of all other early-onset genetic disorders is constant across all regions of the World. This is very informative, as the effect of inbreeding on highly polygenic late-onset disease could be viewed as an effect on many Mendelian variants underlying the disease and the effect on a larger predicted number of genes with modest effect size. It therefore seems plausible that for less heritable diseases the effects of inbreeding will be modest (e.g. relative risks of 1.2-1.5 in individuals with F=0.03), as shown in our study (e.g. for coronary heart disease or stroke). For other late onset

diseases that still may be polygenic, but more heritable (e.g. schizophrenia, epilepsy, some forms of cancer), this risk may be greater (e.g. 2.0-3.0). It is easy to compute that the relative risk of 2.0 present in 12% of individuals globally would increase the global burden of that particular disease by 12%, which is not a negligible effect, and it is derived from relatively conservative estimates. Therefore, further attention to this issue may be of wider general interest.

4.7.2. The Outbreeding Theory

Throughout this thesis, it was pointed that the human population was rather small, subdivided and inbred up to several generations ago, and that it has undergone massive changes in genetic structure at regional and global level due to expansion in size, urbanization and outbreeding, gene flow and admixture. The effects of those changes on public health have not been addressed, although a number of plausible theoretical hypotheses and empirical observations imply that they should be expected. It has been shown that a large portion of increase in life expectancy in both developed and developing world between the years 1930 and 1960 could not be explained by known environmental improvements and economic factors (**321,322**). Therefore, perhaps the dramatic changes in the genetic structure of human population could represent a "forgotten" variable that may be at least partly responsible for the observed trends. Global outbreeding could, theoretically, lead to increased fitness of the human race (possibly partially explaining the human population expansion), and an increase in height, overall health and cognitive performance, i.e. the trends which have all been observed and reported during the past 150 years.

The elements to support this theory rely on at least four independent premises that will need to be proven. The first one is to establish that the appreciable portion of the global increase in life expectancy has not been explained by social, economic and environmental causes. This premise has received some support in the papers of Preston

et al. (**321**) and Caldwell et al. (**322**). It is mainly based on the observation that the economic progress has not been nearly equal in all regions of the world over the past century, and yet the observed trends have been nearly equally observed in both developed and developing countries.

The second premise of the theory that will be particularly difficult to prove is that most human populations were considerably inbred until the past 150-200 years. This is consistent with a life in small, isolated communities attached to the land they harvested. In such communities it is predicted that 10-50% of persons were inbred, as this is still true for many rural areas in the developing world. In addition, there had been an unknown, but possibly considerable prevalence of incestuous relationships. Finally, a factor that could have contributed strongly to the inbreeding in historic populations was positive selection of relatives who shared rare resistance alleles against infectious diseases or starvation during historic disasters leading to population bottlenecks. All of these mechanisms could have contributed to inbreeding historically.

The third element of the theory that would be reasonably easy to prove is that the human population achieved sudden outbreeding through urbanization during 19^{th} and 20^{th} century in both the developed and developing world. The percentage of humans living in cities globally was negligible in the year 1850. However, in 2000 up to a half of all humans moved to cities (**323**). This demographic trend was simultaneously happening in both the developed and developing world, so it does not contradict the observed trends of increase in height, cognitive performance and lifespan in both rich and poor countries simultaneously. Furthermore, changes in inbreeding practices led to a decrease in recessively inherited monogenic disorders at the level of communities, regions and continents. It should be noted that the effects of urbanisation on health and fitness are expected after a "delay" of one average lifespan, until the inbred people who migrate to cities produce the admixed offspring in which the beneficial effects can be measured.

Finally, the fourth independent postulate of the theory would be to prove that inbreeding has a negative predicted overall effect on traits such as fitness, height, cognitive performance, post-reproductive health and life expectancy. The large number of experiments in animals living in natural populations have consistently showed that inbreeding reduces fitness. Also, in natural animal and plant populations inbreeding was shown to increase susceptibility to infectious diseases and other pathologic conditions, which has all been discussed earlier. In humans, inbreeding is consistently linked with an increase in early morbidity and mortality. Outbreeding was shown to increase height and improve scores for cognitive functioning, which were both documented to improve through the 20^{th} century. The latter was even showed to positively correlate with life expectancy (**324**), although it appears tha the initially negative correlation between life expectancy and moderate to severe intellectual disability, but through time a disproportionate increase in the mean life expectancy of people with intellectual disability. The research in the Croatian isolate resource is currently underway to test whether there is indeed any significant correlation between inbreeding, admixture and life expectancy in the appropriately chosen cohort from within these isolate populations.

4.8. CONCLUSION

I have argued in this work that there is a coherent theoretical basis for a role for inbreeding in late-onset diseases of public health importance in humans. The available data from animal and plant studies strongly support the proposed disease mechanisms put forward in this thesis and suggest that these phenomena may be common across species. The available data on the effects of inbreeding in humans has focused on assessing the risk of early mortality and morbidity due to rare recessive deleterious mutations. In an extensive review of inbreeding literature in humans, it was possible to find only a few publications on inbreeding effects on late onset traits or diseases. It is possible that this may be explained to some extent by the fact that in areas of the world where inbreeding is prevalent, late-onset diseases have not until recently represented the main public health problem (e.g. Mediterranean countries, parts of India and sub-Saharan Africa). In western societies, however, inbreeding is not prevalent enough to be studied in a large-scale epidemiological investigation, unless it is undertaken within a large group of immigrants from the countries where it is practiced widely (e.g. the Pakistani population of the United Kingdom). Nevertheless, this may be an important epidemiological risk to evaluate, as with improving life expectancy in large human populations where inbreeding is prevalent these effects could substantially contribute to disease burden and life expectancy.

In a present-day metapopulation of distinct genetic isolates that seems to be well-suited for inbreeding studies due to range of inbreeding values, accessible genealogical records and known demographic history, I have observed the effect of inbreeding on the prevalence of several different late-onset diseases and quantitative traits. The observed effect on blood pressure was strong and not easy to explain by any other variable but inbreeding. This observation has also been reported previously, using similar study design, by at least three more studies in three different continents (South America,

Brazil; North America, USA; and Asia, India). To support the hypothesis that the observed effect was indeed due to inbreeding, I undertook a separate analysis in the same study population using the same design, but investigating the effect on another quantitative trait. This trait, cortical index, can be very precisely measured and represents an intermediate phenotype for a common late-onset disease (osteoporosis), but should not be expected to correlate with blood pressure. This analysis again found the correlation between inbreeding and the quantitative trait of interest, supporting the hypothesis that both observations represent a genuine effect. Another observation in support of genuine inbreeding effects is the increasing prevalence of learning disability in more inbred communities, an effect that has been well documented in many societies previously.

The next step in this thesis was to investigate effect of inbreeding on late-onset diseases of complex etiology, which represent the greatest public health burden in developed countries and are increasing in importance in the developing world. The interest in studying this effect was justified, as there is a sound theoretical background for a potentially large effect of inbreeding on the occurrence of those diseases, which was reviewed. Due to the remarkably few published studies, hardly anything was known about the potential size of this effect in human populations, and the conducted studies seemed to agree about the presence of the effect that, in some instances, was fairly substantial. This work demonstrated significant effects of inbreeding on most of the studied diseases, with those presumed to more heritable and less environmentally influenced showing a greater effect of inbreeding on their prevalence. These findings were not unexpected, because if the uniformity with which inbreeding depression is observed in animals and plants is considered, it would be striking if those effects would not be observed in human populations. Still, there is a need for more studies of better design to be conducted in different world populations and to try to yield the objective estimates of an increase in complex disease prevalence attributable to inbreeding.

A substantial proportion of human populations in the world still practice consanguinity. With an expected increase in life expectancy and prevalent inbreeding, complex chronic diseases could become more common in those regions and thus significantly contribute to the global burden of disease. In addition, the findings of this work could perhaps help mounting further evidence that could show that outbreeding, admixture and heterosis achieved through urbanization during the 19th and 20th centuries could have been partly responsible for some of the secular trends reported globally, such as improvements in fitness, height, health, cognitive performance and adult life expectancy.

* * *

ACKNOWLEDGEMENTS

I am grateful to my mentor, Professor Harry Campbell, and Ms Rosa Bisset from Public Health Sciences, University of Edinburgh, for their continuing friendship, help and support during this work. I am indebted by Dr Andrew Carothers, Professor Alan Wright and Professor Nick Hastie from Human Genetics Unit, Medical Research Council, Edinburgh for their help and encouragement. I thank all the people in Croatia that substantially helped me conducting this research: Professor Pavao Rudan, Dr Branka Janicijevic, Dr Nina Smolej-Narancic, Dr Tatjana Skaric-Juric, Dr Lovorka Barac, Dr Marijana Pericic, Dr Irena Martinovic Klaric and Dr Diana Rudan from the Institute for Anthropological Research in Zagreb; Dr Branka Salzer from the laboratory center "Salzer" in Zagreb; Professor Bozidar Vojnikovic from Faculty of Medicine in Rijeka, Croatia; Dr Nela Mise and Dr Anita Jancic from Lastovo island, Croatia, Dr Magdalena Nardelli-Kovacic and Dr Svetlana Ancic from Mljet island, Croatia; Dr Zrinka Biloglav, Dr Ozren Polasek, Dr Ivana Kolcic and Dr Branko Kolaric from School of Public Health, Faculty of Medicine, University of Zagreb, Croatia.

I am also grateful to the funding bodies that supported this research during 2001-2004: The University of Edinburgh (PhD Scholarship), The Overseas Support Scheme (ORS, tuition fees covered), The British Council (Academic Linkage and Interchange Scheme grant), The Royal Society (International Collaboration Award), The Wellcome Trust (International Research Development Award), The Croatian Ministry of Science and Technology (Research Grant 0108330), and the Medical Research Council UK (Research Grant).

5. REFERENCES

1. World Health Organization: Preamble to the Constitution of the World Health Organization. Official Records of the World Health Organization No. 2, Geneva, 1948.
2. Bachmann W. Health and disease. Critical thoughts on the health concept of the World Health Organisation (WHO). MMW Munch Med Wochenschr 1977;119:349-352.
3. Wright A, Charlesworth B, Rudan I, Carothers A, Campbell H. A polygenic basis for late-onset disease. Trends Genet 2003; 19:97-106.
4. Hamosh A, Scott AF, Amberger J, Valle D, McKusick VA. Online Mendelian Inheritance in Man (OMIM). Hum Mutat 2000; 15:57-61.
5. Rothman KJ, Greenland S. Modern epidemiology, 2nd edition. Lippincott, Williams & Wilkins Publishers, New York, 1998.
6. Khoury MJ, Cohen BH, Beaty TH. Fundamentals of genetic epidemiology, 1st edition. Oxford University Press, Oxford, 1993.
7. Cardon LR, Bell JI. Association study designs for complex diseases. Nat Rev Genet 2001; 2:91-99.
8. Zondervan KT, Cardon LR. The complex interplay among factors that influence allelic association. Nat Rev Genet 2004; 5:89-100.
9. Freimer N, Sabatti C. The human phenome project. Nat Genet 2003; 34:15-21.
10. Merikangas KR, Risch N. Genomic priorities and public health. Science 2003; 302:599-601.
11. Boomsma D, Busjahn A, Peltonen L. Classical twin studies and beyond. Nat Rev Genet 2002; 3:872-882.
12. Barton NH, Keightley PD. Understanding quantitative genetic variation. Nat Rev Genet 2002; 3:11-21.

13. Bittles AH, Neel JV. The costs of human inbreeding and their implications for variations at the DNA level. Nat Genet 1994; 8:117-121.
14. Crow JF. The origins, patterns and implications of human spontaneous mutation. Nat Rev Genet 2000; 1:40-47.
15. Passarge, E. Color Atlas of Genetics. 2nd ed. Georg Thieme Verlag, Stuttgart, 2001.
16. Erlich HA, Gelfand D, Sninsky JJ. Recent advances in the polymerase chain reaction. Science 1991; 252: 1643-1651.
17. Housman D. Human DNA polymorphysm. N Engl J Med 1995; 332:318-320.
18. Reich DE, Lander ES. On the allelic spectrum of human disease. Trends Genet 2001; 17: 502-510.
19. Risch N, Merikangas K. The future of genetic studies of complex human diseases. Science 1996; 273:1516-1517.
20. Bentley DR. Genomes for medicine. Nature 2004; 429:440-445.
21. Bell J. Predicting disease using genomics. Nature 2004; 429: 453-456.
22. Lander ES et al. Initial sequencing and analysis of the human genome. Nature 2001; 409:860-921.
23. Venter JC et al. The sequence of the human genome. Science. 2001; 291:1304-51.
24. Subramanian G, Adams MD, Venter JC, Broder S. Implications of the human genome for understanding human biology and medicine. JAMA 2001; 286:2296-307.
25. Rubinstein DC et al.: Microsatellite evolution and evidence for directionality and variation in rate between species. Nat Genet 1995; 10: 337-343.
26. Collins FS, Guyer MS, Chakravarty A: Variations on a theme: cataloguing human DNA sequence variation. Science 1998; 282:682-689.
27. Brown TA: Genomes. Bios Scientific Publications, Oxford, 1999.
28. Sachidanandam R et al. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. Nature 2001; 409:928-933.
29. Reich DE, Cargill M, Bolk S, Ireland J, Sabeti PC, Richter DJ, Lavery T, Kouyoumjian R, Farhadian SF, Ward R, Lander ES: Linkage disequilibrium in the human genome. Nature 2001; 411: 199-204.

30. Ardlie KG, Kruglyak L, Seielstad M: Patterns of linkage disequilibrium in the human genome. Nat Rev Genet 2002; 3:299-309.

31. Wall JD, Pritchard JK: Haplotype blocks and linkage disequilibrium in the human genome. Nat Rev Genet 2003; 4:587-597.

32. Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, Higgins J, DeFelice M, Lochner A, Faggart M, Liu-Cordero SN, Rotimi C, Adeyemo A, Cooper R, Ward R, Lander ES, Daly MJ, Altshuler D: The structure of haplotype blocks in the human genome. Science 2002; 296:2225-2229.

33. Kruglyak L, Nickerson DA. Variation is the spice of life. Nat Genet 2001; 27:234-236.

34. Strachan TA, Read AP. Human Molecular Genetics, 2nd ed. Bios Scientific Publications, Oxford, 1999.

35. Marth G, Yeh R, Minton M, Donaldson R, Li Q, Duan S, Davenport R, Miller RD, Kwok PY. Single-nucleotide polymorphisms in the public domain: how useful are they? Nat Genet 2001; 27:371-372.

36. Carlson CS, Eberle MA, Rieder MJ, Smith JD, Kruglyak L, Nickerson DA: Additional SNPs and linkage-disequilibrium analyses are necessary for whole-genome association studies in humans. Nat Genet 2003; 33:518-521.

37. Terwilliger JD, Goring HH. Gene mapping in the 20th and 21st centuries: statistical methods, data analysis, and experimental design. Hum Biol 2000; 72:63-132.

38. Lander ES, Botstein D. Homozygosity mapping: a way to map human recessive traits with the DNA of inbred children. Science 1987; 236:1567-70.

39. Peltonen L, Palotie A, Lange K. Use of population isolates for mapping complex traits. Nat Rev Genet 2000; 1:182-190.

40. McVean GA, Myers SR, Hunt S, Deloukas P, Bentley DR, Donnelly P. The fine-scale structure of recombination rate variation in the human genome. Science 2004; 304:581-584.

41. Miyo T, Charlesworth B. Age-specific mortality rates of reproducing and non-reproducing males of Drosophila melanogaster. Proc R Soc Lond B Biol Sci. 2004; 271:2517-2522.

42. Neel JV. Diabetes mellitus: a "thrifty" genotype rendered detrimental by "progress"? Am J Hum Genet 1962; 14:353-362.

43. Online Mendelian Inheritance in Man, OMIM (TM). McKusick-Nathans Institute for Genetic Medicine, Johns Hopkins University (Baltimore, MD) and National Center for Biotechnology Information, National Library of Medicine (Bethesda, MD), 2000. World Wide Web URL: http://www.ncbi.nlm.nih.gov/omim/

44. Botstein D, Risch N. Discovering genotypes underlying human phenotypes: past successes for Mendelian disease, future approaches for complex disease. Nat Genet 2003; 33 Suppl:228-237.

45. Weatherall DJ. Phenotype-genotype relationships in monogenic disease: lessons from the thalassaemias. Nat Rev Genet 2001; 2:245-255.

46. Heckenlively JR, Daiger SP. In: Rimoin DL, Connor JM, Pyeritz RE (Eds): Emory and Rimoin's Principles and Practice of Medical Genetics, 3rd Edition. Churchill-Livingstone, Edinburgh, 1996.

47. Holschneider AM. Hirschprung's disease. Thieme-Stratton, New York, 1982.

48. Gabriel SB, Salomon R, Pelet A, Angrist M, Amiel J, Fornage M, Attie-Bitach T, Olson JM, Hofstra R, Buys C, Steffann J, Munnich A, Lyonnet S, Chakravarti A. Segregation at three loci explains familial and population risk in Hirschsprung disease. Nat Genet. 2002; 31:89-93.

49. Mira MT, Alcais A, Nguyen VT, Moraes MO, Di Flumeri C, Vu HT, Mai CP, Nguyen TH, Nguyen NB, Pham XK, Sarno EN, Alter A, Montpetit A, Moraes ME, Moraes JR, Dore C, Gallant CJ, Lepage P, Verner A, Van De Vosse E, Hudson TJ, Abel L, Schurr E. Susceptibility to leprosy is associated with PARK2 and PACRG. Nature 2004; 427:636-640.

50. Mira MT, Alcais A, Van Thuc N, Thai VH, Huong NT, Ba NN, Verner A, Hudson TJ, Abel L, Schurr E. Chromosome 6q25 is linked to susceptibility to leprosy in a Vietnamese population. Nat Genet. 2003; 33:412-415.

51. Siddiqui MR et al.: A major susceptibility locus for leprosy in India maps to chromosome 10p13. Nat Genet 2001; 27:439-441.

52. Zhang Y, Leaves NI, Anderson GG, Ponting CP, Broxholme J, Holt R, Edser P, Bhattacharyya S, Dunham A, Adcock IM, Pulleyn L, Barnes PJ, Harper JI, Abecasis G, Cardon L, White M, Burton J, Matthews L, Mott R, Ross M, Cox R, Moffatt MF, Cookson WO. Positional cloning of a quantitative trait locus on chromosome 13q14 that influences immunoglobulin E levels and asthma. Nat Genet 2003; 34:181-186.

53. Allen M, Heinzmann A, Noguchi E, Abecasis G, Broxholme J, Ponting CP, Bhattacharyya S, Tinsley J, Zhang Y, Holt R, Jones EY, Lench N, Carey A, Jones H, Dickens NJ, Dimon C, Nicholls R, Baker C, Xue L, Townsend E, Kabesch M, Weiland SK, Carr D, von Mutius E, Adcock IM, Barnes PJ, Lathrop GM, Edwards M, Moffatt MF, Cookson WO. Positional cloning of a novel gene influencing asthma from chromosome 2q14. Nat Genet 2003; 35:258-263.

54. Laitinen T, Polvi A, Rydman P, Vendelin J, Pulkkinen V, Salmikangas P, Makela S, Rehn M, Pirskanen A, Rautanen A, Zucchelli M, Gullsten H, Leino M, Alenius H, Petays T, Haahtela T, Laitinen A, Laprise C, Hudson TJ, Laitinen LA, Kere J. Characterization of a common susceptibility locus for asthma-related traits. Science 2004; 304:300-304.

55. Wills-Karp M, Ewart SL. Time to draw breath: asthma-susceptibility genes are identified. Nat Rev Genet 2004; 5:376-387.

56. Prokunina L, Castillejo-Lopez C, Oberg F, Gunnarsson I, Berg L, Magnusson V, Brookes AJ, Tentler D, Kristjansdottir H, Grondal G, Bolstad AI, Svenungsson E, Lundberg I, Sturfelt G, Jonssen A, Truedsson L, Lima G, Alcocer-Varela J, Jonsson R, Gyllensten UB, Harley JB, Alarcon-Segovia D, Steinsson K, Alarcon-Riquelme ME. A regulatory polymorphism in PDCD1 is associated with susceptibility to systemic lupus erythematosus in humans. Nat Genet. 2002; 32:666-669.

57. Helms C, Cao L, Krueger JG, Wijsman EM, Chamian F, Gordon D, Heffernan M, Daw JA, Robarge J, Ott J, Kwok PY, Menter A, Bowcock AM. A putative RUNX1 binding site variant between SLC9A3R1 and NAT9 is associated with susceptibility to psoriasis. Nat Genet 2003; 35:349-356.

58. International Psoriasis Genetics Consortium. The International Psoriasis Genetics Study: assessing linkage to 14 candidate susceptibility loci in a cohort of 942 affected sib pairs. Am J Hum Genet 2003; 73:430-437.

59. Wright AF, Hastie ND. Complex genetic diseases: controversy over the Croesus code. Genome Biol 2001; 2: 2007.1–8.

60. Lander ES. The new genomics: global views of biology. Science 1996; 274: 536-539.

61. Campbell H, Rudan I. Interpretation of genetic association studies in complex disease. Pharmacogenomics J 2002; 2:349-360.

62. Ioannidis JP, Ntzani EE, Trikalinos TA, Contopoulos-Ioannidis DG. Replication validity of genetic association studies. Nat Genet 2001; 29:306-309.

63. Dahlman I, Eaves IA, Kosoy R, Morrison VA, Heward J, Gough SC, Allahabadia A, Franklyn JA, Tuomilehto J, Tuomilehto-Wolf E, Cucca F, Guja C, Ionescu-Tirgoviste C, Stevens H, Carr P, Nutland S, McKinney P, Shield JP, Wang W, Cordell HJ, Walker N, Todd JA, Concannon P. Parameters for reliable results in genetic association studies in common disease. Nat Genet 2002; 30:149-150.

64. Ozaki K, Inoue K, Sato H, Iida A, Ohnishi Y, Sekine A, Sato H, Odashiro K, Nobuyoshi M, Hori M, Nakamura Y, Tanaka T. Functional variation in LGALS2 confers risk of myocardial infarction and regulates lymphotoxin-alpha secretion in vitro. Nature 2004; 429:72-75.

65. Gretarsdottir S, Thorleifsson G, Reynisdottir ST, Manolescu A, Jonsdottir S, Jonsdottir T, Gudmundsdottir T, Bjarnadottir SM, Einarsson OB, Gudjonsdottir HM, Hawkins M, Gudmundsson G, Gudmundsdottir H, Andrason H, Gudmundsdottir AS, Sigurdardottir M, Chou TT, Nahmias J, Goss S, Sveinbjornsdottir S, Valdimarsson EM, Jakobsson F, Agnarsson U, Gudnason V, Thorgeirsson G, Fingerle J, Gurney M, Gudbjartsson D, Frigge ML, Kong A, Stefansson K, Gulcher JR. The gene encoding phosphodiesterase 4D confers risk of ischemic stroke. Nat Genet. 2003; 35:131-138.

66. Helgadottir A, Manolescu A, Thorleifsson G, Gretarsdottir S, Jonsdottir H, Thorsteinsdottir U, Samani NJ, Gudmundsson G, Grant SF, Thorgeirsson G,

Sveinbjornsdottir S, Valdimarsson EM, Matthiasson SE, Johannsson H, Gudmundsdottir O, Gurney ME, Sainz J, Thorhallsdottir M, Andresdottir M, Frigge ML, Topol EJ, Kong A, Gudnason V, Hakonarson H, Gulcher JR, Stefansson K. The gene encoding 5-lipoxygenase activating protein confers risk of myocardial infarction and stroke. Nat Genet 2004; 36:233-239.

67. Balmain A, Gray J, Ponder B. The genetics and genomics of cancer. Nat Genet 2003; 33 Suppl:238-244.
68. Editorial. Intersecting paths to cancer. Nat Genet 2004; 36:313.
69. Pharoah PD, Antoniou A, Bobrow M, Zimmern RL, Easton DF, Ponder BA. Polygenic susceptibility to breast cancer and implications for prevention. Nat Genet 2002; 31:33-36.
70. Lengauer C, Kinzler KW, Vogelstein B. Genetic instabilities in human cancers. Nature 1998; 396:643-649.
71. Reich DE, Schaffner SF, Daly MJ, McVean G, Mullikin JC, Higgins JM, Richter DJ, Lander ES, Altshuler D. Human genome sequence variation and the influence of gene history, mutation and recombination. Nat Genet 2002; 32:135-142.
72. Lohmueller KE, Pearce CL, Pike M, Lander ES, Hirschhorn JN: Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. Nat Genet 2003; 33:177-182.
73. Freimer N, Sabatti C. The use of pedigree, sib-pair and association studies of common diseases for genetic mapping and epidemiology. Nat Genet 2004; 10:1045-1051.
74. Carlson CS, Eberle MA, Kruglyak L, Nickerson DA. Mapping complex disease loci in whole-genome association studies. Nature 2004; 429:446-452.
75. Larsen CE, Alper CA. The genetics of HLA-associated disease. Curr Opin Immunol 2004; 16:660-667.
76. Denver DR, Morris K, Lynch M, Thomas WK. High mutation rate and predominance of insertions in the *Caenorhabditis elegans* nuclear genome. Nature 2004; 430: 679-682.

77. Cohen JC, Kiss RS, Pertsemlidis A, Marcel YL, McPherson R, Hobbs HH: Multiple rare alleles contribute to low plasma levels of HDL cholesterol. Science. 2004; 305:869-872.

78. Pritchard JK, Cox NJ: The allelic architecture of human disease genes: common disease-common variant...or not? Hum Mol Genet 2002; 11:2417-2423.

79. Wright AF, Carothers AD, Campbell H. Gene-environment interactions--the BioBank UK study. Pharmacogenomics J 2002; 2:75-82.

80. Collins FS. The case for a US prospective cohort study of genes and environment. Nature. 2004; 429:475-7.

81. The International HapMap Consortium. The International HapMap Project. Nature. 2003; 426:789-796.

82. Johnson GC, Esposito L, Barratt BJ, Smith AN, Heward J, Di Genova G, Ueda H, Cordell HJ, Eaves IA, Dudbridge F, Twells RC, Payne F, Hughes W, Nutland S, Stevens H, Carr P, Tuomilehto-Wolf E, Tuomilehto J, Gough SC, Clayton DG, Todd JA. Haplotype tagging for the identification of common disease genes. Nat Genet 2001; 29:233-237.

83. Wright AF, Carothers AD, Pirastu M. Population choice in mapping genes for complex diseases. Nat Genet. 1999; 23:397-404.

84. Rahman P, Jones A, Curtis J, Bartlett S, Peddle L, Fernandez BA, Freimer NB. The Newfoundland population: a unique resource for genetic investigation of complex diseases. Hum Mol Genet 2004; 13:1287.

85. Kaessmann H, Zollner S, Gustafsson AC, Wiebe V, Laan M, Lundeberg J, Uhlen M, Paabo S. Extensive linkage disequilibrium in small human populations in Eurasia. Am J Hum Genet 2002; 70:673-685.

86. Tenesa A, Wright AF, Knott SA, Carothers AD, Hayward C, Angius A, Persico I, Maestrale G, Hastie ND, Pirastu M, Visscher PM. Extent of linkage disequilibrium in a Sardinian sub-isolate: sampling and methodological considerations. Hum Mol Genet 2004; 13:25-33.

87. Shifman S, Kuypers J, Kokoris M, Yakir B, Darvasi A. Linkage disequilibrium patterns of the human genome across populations. Hum Mol Genet 2003; 12:771-776.

88. Aulchenko YS, Heutink P, Mackay I, Bertoli-Avella AM, Pullen J, Vaessen N, Rademaker TA, Sandkuijl LA, Cardon L, Oostra B, van Duijn CM. Linkage disequilibrium in young genetically isolated Dutch population. Eur J Hum Genet 2004; 12:527-534.

89. Rudan I, Campbell H, Rudan P: Genetic epidemiological studies of eastern Adriatic island isolates, Croatia: Objectives and strategies. Coll Antropol, 23: 531-546, 1999.

90. Bulaeva KB, Pavlova TA, Kurbanov RM, Leal S, Bulaev OA. Genetic and epidemiological studies in Dagestan highland isolates. Genetika 2003; 39:413-422.

91. Freedman ML, Reich D, Penney KL, McDonald GJ, Mignault AA, Patterson N, Gabriel SB, Topol EJ, Smoller JW, Pato CN, Pato MT, Petryshen TL, Kolonel LN, Lander ES, Sklar P, Henderson B, Hirschhorn JN, Altshuler D. Assessing the impact of population stratification on genetic association studies. Nat Genet 2004; 36:388-393.

92. Lazarus R, Raby BA, Lange C, Silverman EK, Kwiatkowski DJ, Vercelli D, Klimecki WJ, Martinez FD, Weiss ST. TOLL-like receptor 10 genetic variation is associated with asthma in two independent samples. Am J Respir Crit Care Med 2004;170:594-600.

93. Bonifati V, Rizzu P, Baren J van, Schaap O, Breedveld GJ, Krieger E, Dekker DCJ, Squitieri F, Ibanez P, Joosse M, Dongen JW van, Vanacore N, Swieten JC van, Brice A, Meco G, Duijn CM van, Oostra BA, Heutink P. Mutations in the DJ-1 gene associated with autosomal recessive early-onset parkinsonism. Science 2003; 299:256-259.

94. Haga SB, Khoury MJ, Burke W. Genomic profiling to promote a healthy lifestyle: not ready for prime time. Nat Genet 2003; 34: 347-350.

95. Egger G, Liang G, Aparicio A, Jones PA. Epigenetics in human disease and prospects for epigenetic therapy. Nature 2004; 429:457-463.

96. Evans WE, Relling MV. Moving towards individualized medicine with pharmacogenomics. Nature 2004; 429:464-468.

97. Druker BJ. Imatinib alone and in combination for chronic myeloid leukaemia. Semin Hematol 2003; 40: 50-58.

98. Fornier M, Risio M, Van Poznak C, Siedman A. HER2 testing and correlation with efficacy of trastuzumab therapy. Oncology 2002; 16: 1340-1358.

99. Merikangas KR, Risch N. Genomic priorities and public health. Science 2003; 302: 599-601.

100. Weatherall DJ. Genomics and global health: Time for a reapprisal. Science 2003; 302: 597-598.

101. Bittles AH, Savithri HS, Murthy G, Baskaran G, Wang W, Cahill J, Rao NA. Consanguinity: a familiar story full of surprises, In *Health and Ethnicity* edited by H. Macbeth and P. Shetty. Taylor and Francis, London, 2001.

102. Modell B, Darr A. Science and society: genetic counseling and customary consanguineous marriage. Nat Rev Genet 2002; 3:225-229.

103. Editorial. Consanguinity and Health. Lancet 1991; 338:85-86.

104. Gibbons A. The risks of inbreeding. Science 1993; 259:1252.

105. Bittles AH. Consanguineous marriage and childhood health. Dev Med Child Neurol 2003; 45:571-576.

106. Bittles AH, Mason WM, Greene J, Rao NA. Reproductive behavior and health in consanguineous marriages. Science. 1991; 252:789-794.

107. World Health Organization. Primary health care approaches for the prevention and control of congenital and genetic disorders. WHO Meeting Report, Cairo, Egypt, 6-8 December 1999. WHO/HGN/WG/00.1. WHO, Geneva, 2000.

108. Zlotogora J. What is the birth defect risk associated with consanguineous marriages? Am J Med Genet 2002; 109:70-71.

109. Jorde LB: Consanguinity and prereproductive mortality in Utah Mormon population. Hum Hered 2001; 52:61-65.

110. Schull WJ, Neel JV. The effects of parental consanguinity and inbreeding in Hirado, Japan. V. Summary and interpretation. Am J Hum Genet 1972; 24:425-453.

111. Khlat M, Khoury M. Inbreeding and diseases: Demographic, genetic and epidemiologic perspectives. Epidemiol Rev 1991; 13:28-41.

112. Hughes KA, Alipaz JA, Drnevich JM, Reynolds RM. A test of evolutionary theories of aging. Proc Natl Acad Sci USA 2002; 99:14286-14291.

113. Charlesworth B, Hughes KA. Age-specific inbreeding depression and components of genetic variance in relation to the evolution of senescence. Proc Natl Acad Sci USA 1996; 93: 6140-6145.

114. Bittles AH. (1988) Empirical estimates of the global prevalence of consanguineous marriage in contemporary societies, in *Morrison Institute for Population and Resource Studies Working Report 74*. Stanford University, Stanford.

115. Freire-Maia N. Inbreeding levels in different countries. Soc Biol 1982; 29:69-81.

116. Benallègue A, Kedji F. Consanguinité et santé publique: Etude algérienne. Archives Francaises de Pediatrie 1984; 41:435-440.

117. Scott-Emuakpor AB. The mutation load in an African population. I. An analysis of consanguineous marriages in Nigeria. Am J Hum Genet 1974; 26:674-682.

118. Stevenson AC, Johnston HA, Stewart MIP, Golding DR. Congenital malformations: a report of a study of series of consecutive births in 24 centres. Bull World Health Organ 1966; 34 (supplement): 1-125.

119. Freire-Maia N. Inbreeding levels in American and Canadian populations: a comparison with Latin America. Eugen Q 1968; 15:22-33.

120. Farhud DD, Mahmoudi M, Kamali MS, Marzban M, Andonian L, Saffart R. Consanguinity in Iran. Iranian J Public Health 1991; 20:1-13.

121. Khoury SA, Massad D. Consanguineous marriage in Jordan. Am J Med Genet 1992; 43:769-775.

122. Al-Awadi SA, Naguib KK, Moussa MA, Farag TI, Teebi AS, El-Khalifa MY. The effect of consanguineous marriages on reproductive wastage. Clinical Genetics 1986; 29:384-388.

123. Khlat M. Consanguineous marriage and reproduction in Beirut, Lebanon. Am J Hum Genet 1988; 43:188-196.

124. Rajab A, Patton M. A study of consanguinity in the Sultanate of Oman. Annals of Human Biology 2000; 27:321-326.

125. El-Hazmi MAF, Al-Swailem AR, Warsy AS, Al-Swailem AM, Sulaimani R, Al-Meshari AA. Consanguinity among the Saudi Arabian population. J Med Genet 1995; 52:623-626.

126. Riou SE, Younsi C, Chaabouni H. Consanguinité dans la population du Nord de la Tunisie. La Tunisie Médicale 1989; 67:167-172.

127. Al-Gazali LI, Bener A, Abdulrazzaq YM, Micallef R, Al-Khayat AI, Gaber T. Consanguineous marriages in the United Arab Emirates. J Biosoc Sci 1997; 29:491-497.

128. Habib Z, Böök JA. Consanguinity and the incidence of thalassaemia in Egypt. Hereditas 1983; 99:215-217.

129. Hafez M, El-Tahan H, Awadalla M, El-Khayat H, Abdel-Gafar A, Ghoneim M. Consanguineous matings in the Egyptian population. J Med Genet 1983; 20:58-60.

130. Al-Hamamy H, Al-Bayati N, Al-Kubaisy W. Consanguineous matings in the Iraqi urban population and the effect on pregnancy outcome and infant mortality. Iraqi Med J 1986; 34:75-80.

131. Hamamy HA, Al-Hakkak ZS. Consanguinity and reproductive health in Iraq. Hum Hered 1989; 39:271-275.

132. Shami SA, Hussain SB. Consanguinity in the population of Gujrat (Punjab), Pakistan. Biologia 1984; 30:93-109.

133. Shami SA, Minhas IB. Effects of consanguineous marriages on offspring mortality in the City of Jhelum (Punjab), Pakistan. Biologia 1984; 30:153-165.

134. Shami SA, Siddiqui H. The effects of parental consanguinity in Rawalpindi City (Punjab), Pakistan. Biologia 1984; 30:189-200.

135. Bittles AH, Grant JC, Shami SA. An evaluation of consanguinity as a determinant of reproductive behaviour and mortality in Pakistan. Int J Epidemiol 1993; 22:463-467.

136. Ahmed T, Ali SM, Aliaga A, Arnold F, Ayub M, Bhatti MH, et al. Pakistan Demographic and Health Survey 1990/91. Islamabad and Columbia, MD: Pakistan National Institute of Population Studies and Macro International, 1992.

137. Saha N, El Sheikh FS. Inbreeding levels in Khartoum. J Biosoc Sci 1988; 20:333-336.

138. Jurdi R, Saxena PC. The prevalence and correlates of consanguineous marriages in Yemen: similarities and contrasts with other Arab countries. J Biosoc Sci 2003; 35:1-13.

139. McCullough JM, O'Rourke DH. Geographic distribution of consanguinity in Europe. Ann Hum Biol 1986; 13:359-367.

140. Jaber L, Bailey-Wilson JE, Haj-Yehia M, Hernandez J, Shohat M. Consanguineous matings in an Israeli-Arab community. Arch Pediatr Adolesc Med 1994; 148:412-415.

141. Freundlich E, Hino N. Consanguineous marriage among rural Arabs in Israel. Israel J Med Sci 1984; 20:1035-1038.

142. Vardi-Saliternik R, Friedlander Y, Cohen T. Consanguinity in a population sample of Israeli Muslim Arabs, Christian Arabs and Druze. Ann Hum Biol 2002; 29:422-431.

143. Basaran N, Sayli BS, Basaran A, Solak M, Artan S, Stevenson JD. Consanguineous marriages in the Turkish population. Clinical Genetics 1988; 34:339-341.

144. Tunçbilek E, Ulusoy M. Consanguinity in Turkey in 1988. Turk J Popul Stud 1989; 11:35-46.

145. Glinka J. Consanguineous marriage. In: Gesch PF (Ed): Culture, Gospel and Church. Madang, Divine Word Mission, 1994, pp. 92-99.

146. Reid RM. Effects of consanguineous marriage and inbreeding on couple fertility and offspring mortality in rural Sri Lanka. Hum Biol 1976; 48:139-146.

147. Khan et al. (unpublished material)

148. Bittles AH, Hussain R. An analysis of consanguineous marriage in the Muslim population of India at regional and state levels. Ann Hum Biol 2000; 27:163-171.
149. Goswami HK. Frequency of consanguineous marriages in Madhya Pradesh. Acta Genet Med Gemellol 1970; 19:486-490.
150. Afzal M. Consequences of consanguinity on cognitive behavior. Behav Genet 1988; 18:583-594.
151. Basu SK. Effect of consanguinity among North Indian Muslims. J Popul Res 1975; 2:57-68.
152. Sanghvi LD. Inbreeding in India. Eugen Q 1966; 13:291-301.
153. Reddy BM. Inbreeding effects on reproductive performance: a study based on a large sample from the endogamous Vadde of Kolleru Lake, Andhra Pradesh, India. Human Biology 1992; 64:659-682.
154. Rao PSS, Inbaraj SG, Kaliaperumal VG. An epidemiological study of consanguinity in a large South Indian town. Ind J Med Res 1971; 59:294-301.
155. Rao PSS, Inbaraj SG. Inbreeding effects on human reproduction in Tamil Nadu of South India. Ann Hum Genet 1977; 41:87-98.
156. Sanghvi LD, Varde DS, Master HR. Frequency of consanguineous marriages in twelve endogamous groups in Bombay. Acta Genet Stat Med 1956; 6:41-49.
157. Roychoudhury AK. Inbreeding in Lakshadweep. Man in India 1977; 57:328-330.
158. Nelson J, Smith M, Bittles AH. Consanguineous marriage and its clinical consequences in migrants to Australia. Clin Genet 1997; 52:142-146.
159. Bittles AH (ed): www.consang.net/global_prevalence/tables/ (accessed Jan 7, 2005)
160. Imaizumi Y, Shinozaki N, Aoki H. Inbreeding in Japan: results of a nation-wide study. Japan J Hum Genet 1975; 20, 91-107.
161. Imaizumi Y. A recent survey of consanguineous marriages in Japan. Clin Genet 1986; 30:230-233.
162. Du R-B, Zhao Z-L, Xu L-J, Wang Y-F, Cui W-Y, Mao Z-R, et al. Percentage and types of consanguineous marriages of different nationalities and regions in China. Natl Med J China 1981; 61:723-728. [In Chinese].

163. Wu L. Investigation of consanguineous marriages among 30 Chinese ethnic groups. Heredity Dis 1987; 4:163-166. [In Chinese].

164. Zhan J, Qin W, Zhou Y, Chen K, Yan W, Yu W. Effects of consanguineous marriages on hereditary diseases: a study of the Han ethnic group in different geographic districts of Zejiang province. Natl Med J China 1992; 172: 674-676. [In Chinese].

165. Wang W, Qian C, Bittles AH. Consanguineous marriage in PR China: a study in rural Man (Manchu) communities. Ann Hum Biol 2002; 29:685-690.

166. Ober C, Hyslop T, Hauck WW. Inbreeding effects on fertility in humans: evidence for reproductive compensation. Am J Hum Genet 1999; 64:225-231.

167. http:/www.who.int/whr2001/2001/archives/2000/en/pdf/Statitical Annex.pdf (Accessed on October 10th, 2004).

168. International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. Nature 2004; 431:931-945.

169. Denver DR, Morris K, Lynch M, Thomas WK. High mutation rate and predominance of insertions in the *Caenorhabditis elegans* nuclear genome. Nature 2004; 430:679-682.

170. Broman KW, Weber JL. Long homozygous chromosomal segments in reference families from the centre d'Etude du polymorphisme humain. Am J Hum Genet 1999; 65:1493-1500.

171. Rohde DL, Olson S, Chang JT. Modelling the recent common ancestry of all living humans. Nature 2004; 431:562-566.

172. Cavalli-Sforza LL, Moroni A, Zei G. Consanguinity, inbreeding and genetic drift in Italy. Princeton University Press, USA, 2004.

173. Freire-Maia N, Elisbao T. Inbreeding effect on morbidity: III. A review of the world literature. Am J Med Genet 1984; 18:391-400

174. Morton NE. Effect of inbreeding on IQ and mental retardation. Proc Natl Acad Sci U S A 1978; 75:3906-3908.

175. Rudan I, Rudan D, Campbell H, Biloglav Z, Sibbett L, Janicijevic B, Smolej-Narancic N, Rudan P. Inbreeding and learning disability in Croatian island isolates. Coll Antropol 2002; 26:421-428.

176. Edmond M, De Braekeleer M. Inbreeding effects on fertility and sterility: a case-control study in Saguenay-Lac-Saint-Jean (Quebec, Canada) based on a population registry 1838-1971. Ann Hum Biol 1993; 20:545-555.

177. Shami SA, Schmitt LH, Bittles AH. Consanguinity, spousal age at marriage and fertility in seven Pakistani Punjab cities. Ann Hum Biol 1990; 17:97-105.

178. Schull WJ, Furusho T, Yamamoto M, Nagano H, Komatsu I. The effect of parental consanguinity and inbreeding in Hirado, Japan. IV. Fertility and reproductive compensation. Humangenetik 1970; 9:294-315.

179. Werren JH, Hatcher MJ. Maternal-zygotic gene conflict over sex determination: effects of inbreeding. Genetics 2000; 155:1469-1479.

180. Bittles AH, Devi AR, Rao NA. Consanguinity, twinning and secondary sex ratio in the population of Karnataka, south India. Ann Hum Biol 1988; 15:455-460.

181. Goswami HK. Twinning and inbreeding in India: the fraternal component. Acta Genet Med Gemellol (Roma) 1987; 36:343-347.

182. Philippe P. Genetic epidemiology of twinning: a population-based study. Am J Med Genet 1985; 20:97-105.

183. Markow TA, Martin JF. Inbreeding and developmental stability in a small human population. Ann Hum Biol 1993; 20:389-394.

184. Renuka Nair R, Murty JS. ABO blood group incompatibility and inbreeding effects: evidence for an interaction. Hum Genet 1985; 69:147-150.

185. Hook EB. A method for the detection of the contribution of X heterosis to female survival by comparison of sex ratios of offspring from consanguineous matings. Am J Hum Genet 1969; 21:290-293.

186. Curie-Cohen M. The frequency of consanguineous matings due to multiple use of donors in artificial insemination. Am J Hum Genet 1980; 32:589-600.

187. Mingroni MA: The secular rise in IQ: Giving heterosis a closer look. Intelligence 2004; 32: 65-83.

188. Falconer DS, Mackay TFC. Introduction to quantitative genetics, 4th edition. Longman, Harlow, UK, 1996.

189. Charlesworth B, Charlesworth D. The genetic basis of inbreeding depression. Genet. Res. 1999; 74:329-340.

190. World Health Organization. The World Health Report 2002: Reducing risks, promoting healthy life. WHO, Geneva, 2002.

191. Hurwich BJ, Rosner B, Nubani N, Kass EH, Lewitter FI. Familial aggregation of blood pressure in a highly inbred community, Abu Ghosh, Israel. Am J Epidemiol 1982; 115: 646-656.

192. Thomas JD, Doucette MM, Thomas DC, Stoeckle JD. Disease, lifestyle and consanguinity in 58 American gypsies. Lancet 1987; 2: 377-379.

193. Wahid Saeed AA, al Shammary FJ, Khoja TA, Hashim TJ, Anokute CC, Khan SB. Prevalence of hypertension and socio-demographic characteristics of adult hypertensives in Riyadh City, Saudi Arabia. J Hum Hypertens 1996; 10: 583-587.

194. Halberstein RA. Blood pressure in the Caribbean. Hum Biol 1999; 71: 659–684.

195. Krieger H. Inbreeding effects on metrical traits in Northeastern Brazil. Am J Hum Genet 1969; 21:537-546.

196. Martin AO, Kurczynski TW, Steinberg AG. Familial studies of medical and anthropometric variables in a human isolate. Am J Hum Genet 1973; 25:581-593.

197. Baddaruddoza. Inbreeding effects on metrical phenotypes among North Indian children. Coll Antropol 2004; 28 Suppl 2: 311-320, 2004.

198. Schreider E. Body-height and inbreeding in France. Am J Phys Anthropol 1967; 26:1-3.

199. Little BB, Malina RM. Gene flow and variation in stature and craniofacial dimensions among indigenous populations of southern Mexico, Guatemala, and Honduras. Am J Phys Anthropol 1986; 70:505-512.

200. Ferak V, Lichardova Z, Bojnova V. Endogamy, exogamy, and stature. Eugen Q; 15:273-276.

201. Rose G. The strategy of preventive medicine. Oxford University Press, Oxford, UK, 1992.

202. Abney M, McPeek MS, Ober C. Broad and narrow heritabilities of quantitative traits in a founder population. Am J Hum Genet 2001; 68:1302-1307.

203. Abney M, McPeek MS, Ober C. Estimation of variance components of quantitative traits in inbred populations. Am J Hum Genet 2000; 66:629-650.

204. Blangero J. Localization and identification of human quantitative trait loci: king harvest has surely come. Curr Opin Genet Dev. 2004; 14:233-240.

205. Shami SA, Qaisar R, Bittles AH. Consanguinity and adult morbidity in Pakistan. Lancet 1991; 338:954.

206. Puzyrev VP, Lemza SV, Nazarenko LP, Panphilov VI. Influence of genetic and demographic factors on etiology and pathogenesis of chronic disease in north Siberian aborigines. Arctic Med Res 1992; 51:136–142.

207. Jaber L, Shohat T, Rotter JI, Shohat M. Consanguinity and common adult diseases in Israeli Arab communities. Am J Med Genet 1997; 70:346–348.

208. Ismail J, Jafar TH, Jafary FH, White F, Faruqui AM, Chaturvedi N. Risk factors for non-fatal myocardial infarction in young South Asian adults. Heart 2004; 90:259-263.

209. Simpson JL, Martin AO, Elias S, Sarto GE, Dunn JK. Cancers of the breast and female genital system: search for recessive genetic factors through analysis of human isolate. Am J Obstet Gynecol. 1981; 141:629-636.

210. Lebel RR, Gallagher WB. Wisconsin consanguinity studies. II: Familial adenocarcinomatosis. Am J Med Genet1989; 33:1–6.

211. Rudan I. Inbreeding and cancer incidence in human isolates. Hum Biol 1999; 71:173–187.

212. Denic S, Bener A. Consanguinity decreases risk of breast cancer--cervical cancer unaffected. Br J Cancer 2001; 85:1675–1679.

213. Abaskuliev AA, Skoblo GV. Inbreeding, endogamy and exogamy among relatives of schizophrenia patients. Genetika 1975; 11:145–148.

214. Saugstad L, Odegard O. Inbreeding and schizophrenia. Clin Genet 1986; 30:261–275.

215. Gindilis VM, Gainullin RG, Shmaonova LM. Genetico-demographic patterns of the prevalence of various forms of endogenous psychoses. Genetika 1989; 25:734-743.

216. Vezina H, Heyer E, Fortier I, Ouellette G, Robitaille Y, Gauvreau D. A genealogical study of Alzheimer disease in the Saguenay region of Quebec. Genet Epidemiol. 1999; 16:412-425.

217. Roberts DF, Roberts MJ, Poskanzer DC. Genetic analysis of multiple sclerosis in Orkney. J Epidemiol Comm Health 1979; 33:229–35

218. Roberts DF, Roberts MJ, Poskanzer DC. Genetic analysis of multiple sclerosis in Shetland. J Epidemiol Comm Health 1983; 37:281–285.

219. Ombra MN, Forabosco P, Casula S, Angius A, Maestrale G, Petretto E, Casu G, Colussi G, Usai E, Melis P, Pirastu M.. Identification of a new candidate locus for uric acid nephrolithiasis. Am J Hum Genet 2001; 68:1119–1129.

220. Partridge L, Gems D. Mechanisms of ageing: Public or private? Nat Rev Genet 2002; 3: 165-175.

221. Crow JF. Mutation, mean fitness and genetic load. Oxf Surv Evol Biol 1993; 9:3-42.

222. Charlesworth B, Hughes KA. The maintenance of genetic variation in life-history traits. In: Singh RS, Krimbas CB (Eds): Evolutionary genetics: From molecules to Morphology, Vol 1. Cambridge University Press, Cambridge, UK, 1999, pp. 369-392.

223. Hayes B, Goddard ME. The distribution of the effects of genes affecting quantitative traits in livestock. Genet Sel Evol 2001; 33:209-230.

224. Bennett LA, Angel JL, Roberts DF, Rudan P. Joint Study of Biological and Cultural Variation in Dalmatian Village Populations: Project Description. Coll Antropol 1983; 7:195-198.

225. Rudan P, Sujoldžić A, Šimić D, Bennett LA, Roberts DF. Population structure in the eastern Adriatic: the influence of historical processes, migration patterns,

isolation and ecological pressures, and their interaction. In: Roberts DF, Fujiki N, Torizuka K (Eds.). Isolation, Migration and Health. Cambridge Univ. Press, SSHB 33:204-218, 1992.

226. Martinović-Klarić I. Population Structure of the Rural Communities on the Island of Krk (Croatia): A Comparison of Genetics, Cultural and Geographic Data. Am J Hum Biol 2000; 12:509-525.
227. Martinović Klarić I, Barać L, Buković D, et al. STR polymorphisms in the population of the island of Brač, Croatia. Homo 2000; 51:141-150.
228. Martinović I, Barać L, Furač I, et al. STR polymorphisms in the population of the island of Hvar. Hum Biol 1999; 71:341-352.
229. Martinović Klarić I, Barać L, Buković D, et al. STR variation in eight village populations of the island of Korčula (Croatia). Ann Hum Biol 2001; 28:281-294.
230. Martinović I, Mastana S, Janićijević B, et al. VNTR DNA variation in the population of the island of Hvar, Croatia. Ann Hum Biol 1998; 25:489-499.
231. Barać L, Peričić M, Martinović-Klarić I, et al. Y chromosomal heritage of Croatian population and its island isolates. Eur J Hum Genet 2003; 11:535-542.
232. Tolk HV, Peričić M, Barać L, et al. MtDNA haplogroups in the populations of Croatian Adriatic islands. Coll Antropol 2000; 24:267-280.
233. Torroni A, Bandelt H-J, Macaulay V, et al. A signal, from human mtDNA, of postglacial recolonization in Europe. Am J Hum Genet 2001; 69:844-852.
234. Martinović I, Bakran M, Chaventre A, et al. Application of HLA Class II polymorphism analysis to the study of the population structure of the Island of Krk, Croatia. Hum Biol 1997; 69:819-829.
235. Grubić Z, Žunec R, Čečuk-Jeličić E, et al. HLA Class II Gene and Haplotype Diversity in the Population of the Island of Hvar, Croatia. Coll Antropol 1998; 22:157-168.
236. Cambon-Thomsen A, Sommer E, Sevin A, et al. HLA polymorphism in Olib and Silba populations. Coll Antropol 1989; 13:311-322.

237. Borot N, Dugoujon JM, Janićijević B, Rudan P, Chaventre A. Gm and Km Immunoglobulin allotypes in populations on the Island of Pag. Coll Antropol 1991; 15:247-255.

238. Roberts DF, Noor ZM, Papiha SS, Rudan P. Genetic variation in Brač, Croatia. Ann Hum Biol 1992; 19:539-557.

239. Janićijević B, Bakran M, Papiha SS, Chaventre A, Roberts DF. Serogenetic analysis in the study of the population structure of the Eastern Adriatic (Croatia). Hum Biol 1994; 66:991-1003.

240. Janićijević B. Genetic structure and differentiation between populations of Korčula island and Pelješac peninsula. Coll Antropol 1988; 12:369-376.

241. Arnaud J, Borot N, Chaventre A, et al. Haematological Study of the Populations of Silba and Olib: Red Blood Cell Markers. Coll Antropol 1989; 13:281-290.

242. Kopajtić B, Dujmović M, Kolacio Z, Kogoj-Bakić V: Enclaves of hereditary dwarfism on the island of Krk, Croatia. Coll Antropol 1995; 19: 365-371.

243. Kržišnik C, Kolacio Z, Battelino T, Brown M, Parks JS, Laron Z. The "little people" of the island of Krk – revisited. Etiology of hypopituitarism revealed. J Endocr Genet 1999; 1:9-19.

244. Zergollern L. A follow-up on Hanhart's dwarfs of Krk. Birth Defects Orig Artic Ser 1971; 7:28-32.

245. Bohaček N. Tristan da Cunha and Susak. Lijec Vjesn 1964; 86:1412-1416.

246. Maričević A. Incidence of congenital hip dislocation in Lastovo 1885-1993. Lijec Vjesn 1995; 117:126-129.

247. Rudan I. Ancestral kinship and cancer in Lastovo Island, Croatia. Hum Biol 2001; 73:871-884.

248. Rudan I. Inbreeding and cancer incidence in human isolates. Hum Biol 1999; 71:173-187.

249. Bakija-Konsuo A, Basta-Juzbašić A, Rudan I, et al. Mal de Meleda: Genetic haplotype analysis and clinicopathological findings in cases originating from the island of Mljet (Meleda), Croatia. Dermatology 2002; 205:32-39.

250. Turčinov D, Krishnamoorthy R, Janićijević B, et al. Anthropological analysis of abnormal human Alpha-Globin gene cluster arrangement on chromosome 16. Coll Antropol 2000; 24:295-302.

251. Borot N, Arnaud J, Rudan P, Chaventre A, Sevin J. Phosphoglucomutase-1 subtypes in two populations in Adriatic islands: presence of PGM1*W3 (PGM1*7+) Allele1. Hum Hered 1991; 41:309-315.

252. Tolk H-V, Barać L, Peričić M, et al. The evidence of mtDNA haplogroup F in a European population and its ethnohistoric implications. Eur J Hum Genet 2001; 9:717-723.

253. Forenbaher S. The earliest islanders of the eastern Adriatic. Coll Antropol 1999; 23:521-530.

254. Roguljic D, Rudan I, Rudan P: Estimation of inbreeding, kinship, genetic distances and population structure from surnames: Example from the island of Hvar, Croatia. Am J Hum Biol 1997; 9:595-608.

255. Rudan D, Rudan I, Sujoldzic A, Chaventre A, Janicijevic B, Macan T, Rudan P. Analysis of changes in genetic structure from surnames. A study of three generations of Pag island isolates, Croatia. Homo 2000; 51: 110-131.

256. Rudan I, Rudan P: Comparison between coefficients of inbreeding computed from deficit of heterozygotes for codominant autosomal genetic polymorphisms and from isonymy data: A study of Hvar island isolates, Croatia. In: Susanne C, Bodszar EB (Eds): Human population genetics in Europe. Biennial Book of European Anthropological Association, Vol. 1, Budapest, 2000, pp. 117-128.

257. Barac L, Smolej-Narancic N, Macan T. Assortative mate choice in the population of Brac, Croatia. Homo 1999; 50:183-194.

258. Wright S. Coefficients of inbreeding and relationship. Am. Nat. 1922; 56:330-338.

259. Tay YS, Yip WC. The estimation of inbreeding from isonymy: relationship to the average inbreeding coefficient. Ann Hum Genet 1984; 48:185-194.

260. Weiner JS, Lourie JA. Human Biology - A guide to field methods. Blackwell, Oxford, 1969.

261. Barnett E, Nordin BE. The radiological diagnosis of osteoporosis: a new approach. Clin Radiol 1960; 11:166-174.

262. Ginsburg E, Skaric-Juric T, Kobyliansky E, Karasik D, Malkin I, Rudan P. Evidence on major gene control of cortical index in pedigree data from Middle Dalmatia, Croatia. Am J Hum Biol 2001; 13:398-408.

263. Croatian Ministry of Health and Croatian Ministry of Education and Culture: Regulations on enrolling the children in elementary schools. Narodne novine, Zagreb, 1991, Vol. 13, pp. 425-427.

264. Freire-Maia N. Five landmarks in inbreeding studies. Am J Med Genet 1990; 25:118-120.

265. http://www.rab-croatia.com/history/arhiva (accessed on Dec 20th, 2004).

266. Skreblin L, Simicic L, Sujoldzic A. Ethnohistorical processes, demographic structure and linguistic determinants of the Island of Vis. Coll Antropol 2002; 26:333-350.

267. Rose GA. The diagnosis of ischaemic heart pain and intermittent claudication in field surveys. Bull World Health Organ 1962; 27:645-658.

268. Leng GC, Fowkes FG. The Edinburgh Claudication Questionnaire: an improved version of the WHO/Rose Questionnaire for use in epidemiological surveys. J Clin Epidemiol 1992; 45:1101-1109.

269. Pardell H, Roure E, Drygas W, Morava E, Nussel E, Puska P, Uhanov M, Laaksonen M, Tresserras R, Salto E, Salleras L. East-west differences in reported preventive practices. A comparative study of six European areas of the WHO-CINDI programme. Eur J Public Health 2001; 11:393-396.

270. Bellia V, Pistelli F, Giannini D, Scichilone N, Catalano F, Spatafora M, Hopps R, Carrozzi L, Baldacci S, Di Pede F, Paggiaro P, Viegi G. Questionnaires, spirometry and PEF monitoring in epidemiological studies on elderly respiratory patients. Eur Respir J Suppl 2003; 40:21s-27s.

271. Ware JE Jr, Sherbourne CD: The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Med Care 1992; 30:473-483.

272. Tambs K, Moum T, Eaves LJ, Neale MC, Midthjell K, Lund-Larsen PG, Naess S. Genetic and environmental contributions to the variance of body height in a sample of first and second degree relatives. Am J Phys Anthropol 1992; 88:285-294.

273. Pawlowski B, Dunbar RI, Lipowicz A. Tall men have more reproductive success.Nature 2000; 403:156.

274. Rudan I, Padovan M, Rudan D, Campbell H, Biloglav Z, Janićijević B, Smolej-Narančić N, Rudan P. Inbreeding and nephrolithiasis in Croatian island isolates. Coll Antropol 2002; 26:11-21.

275. Curtis CD, Goggins M. DNA methylation analysis in human cancer. Methods Mol Med 2004; 103:123-136.

276. Velho G, Froguel P, Clement K, Pueyo ME, Rakotoambinina B, Zouali H, Passa P, Cohen D, Robert JJ. Primary pancreatic beta-cell secretory defect caused by mutations in glucokinase gene in kindreds of maturity onset diabetes of the young. Lancet 1992; 340:444-448.

277. Fay JC, Wyckoffand GJ, Wu CI. Positive and negative selection on the human genome. Genetics 2001; 158:1227-1234.

278. Carlborg O, Haley C. Epistasis: too often neglected in complex trait studies? Nat Rev Genet 2004; 5:618-625.

279. Jimenez JA, Hughes KA, Alaks G, Graham L, Lacy RC. An experimental study of inbreeding depression in a natural habitat. Science 1994; 266:271-273.

280. Penn DJ, Damjanovich K, Potts WK. MHC heterozygosity confers a selective advantage against multiple-strain infections. Proc Natl Acad Sci USA 2002; 99:11260-11264.

281. Wolanski N, Jarosz E, Pyzuk M. Heterosis in man: growth in offspring and distance between parents' birthplaces. Soc Biol 1970; 17:1-16.

282. Ebert D, Haag C, Kirkpatrick M, Riek M, Hottinger JW, Pajunen VI. A selective advantage to immigrant genes in a Daphnia metapopulation. Science 2002; 295:485-488.

283. Ives AR, Whitlock MC. Ecology. Inbreeding and metapopulations. Science 2002; 295:454-455.

284. Spielman D, Brook BW, Frankham R. Most species are not driven to extinction before genetic factors impact them. Proc Natl Acad Sci USA 2004; 101:15261-15264.

285. Merila J, Sheldon B, Griffith S. Heterotic effects on fitness in a wild bird population. Ann Zool Fenn 2003; 40:269-280.

286. Van Oosterhut C, Van Heuven MK, Brakefield PM. On the neutrality of molecular genetic markers: pedigree analysis of genetic variation in fragmented populations. Mol Ecol 2004; 13: 1025-1034.

287. Balloux F, Amos W, Coulson T. Does heterozygosity estimate inbreeding in real populations? Mol Ecol 2004; 13:3021-3031.

288. Hansson B, Westerdahl H, Hasselquist D, Akesson M, Bensch S. Does linkage disequilibrium generate heterozygosity fitness correlations in great reed warblers. Evolution 2004; 58:870-879.

289. Mackay TF. The genetic architecture of quantitative traits: lessons from Drosophila. Curr Opin Genet Dev 2004; 14:253-257.

290. Coulston T, Albon S, Slate J, Pemberton J. Microsatellite loci reveal sex-dependent responses to inbreeding and outbreeding in red deer calves. Evolution 1999; 53:1951-1960.

291. Acevedo-Whitehouse K, Gulland F, Greig D, Amos W. Inbreeding: Disease susceptibility in California sea lions. Nature 2003; 422: 35.

292. Foerster K, Delhey K, Johnsen A, Lifjeld JT, Kempenaers B. Females increase offspring heterozygosity and fitness through extra-pair matings. Nature 2003; 425: 714-717.

293. Keller L, Waller D. Inbreeding effects in wild populations. Trends Ecol Evol 2002; 17:230-241.

294. Coltman D, Slate J. Microsatellite measures of inbreeding: a meta-analysis. Evolution 2003; 57:971-983.

295. Lifton RP, Gharavi AG, Geller DS. Molecular mechanisms of human hypertension. Cell 2001; 104:545-546.

296. Cavalli-Sforza LL, Bodmer WF. The Genetics of Human Populations. WH Freeman and Company, San Francisco, 1971, pp 534-536.

297. Miall WE, Oldham PD. The hereditary factor in arterial blood pressure. Brit Med J 1963; 19:75-80.

298. Brink A (ed). Heritage from Mendel. Univ Wisconsin, Madison, Wisconsin, 1967.

299. Mackay TF. Quantitative trait loci in Drosophila. Nat Rev Genet 2001; 2:11-20.

300. Flint J, Mott R. Finding the molecular basis of quantitative traits: successes and pitfalls. Nature Rev Genet 2001; 2:437-445.

301. Halushka MK, Fan JB, Bentley K, Hsie L, Shen N, Weder A, Cooper R, Lipshutz R, Chakravarti A. Patterns of single-nucleotide polymorphisms in candidate genes for blood-pressure homeostasis. Nat Genet 1999; 22:239-247.

302. Nair RR, Thomas SV. Genetic liability to epilepsy in Kerala State, India. Epilepsy Res 2004; 62:163-170.

303. Gilani GM, Kamal S. Risk factors for breast cancer in Pakistani women aged less than 45 years. Ann Hum Biol 2004; 31:398-407.

304. Law MR, Wald NJ. Risk factor thresholds: their existence under scrutiny. BMJ 2002; 324:1570–1576.

305. Bamshad M, Wooding SP. Signatures of natural selection in the human genome. Nat Rev Genet 2003; 4:99-111.

306. World Population Prospects: the 2000 revision. United Nations, New York, 2001.

307. Roche AF. Secular trends in human growth, maturation, and development. Monogr Soc Res Child Dev 1979; 44:1-120.

308. Altukhov YuP, Sheremet'eva VA, Rychkov YuG. Heterosis as the cause of the secular trend in humans. Dokl Biol Sci 2000; 370: 43-46.

309. Rudan I, Campbell H. Five reasons why inbreeding may have considerable effect on post-reproductive human health. Coll Antropol 2004; 28: 943-950.

310. Glemin S, Ronfort J, Bataillon T. Patterns of inbreeding depression and architecture of the load in subdivided populations. Genetics 2003; 165: 2193-2212.

311. Joron M, Brakefield PM. Captivity masks inbreeding effects on male mating success in butterflies. Nature 2003; 424:191-194.

312. Briskie JV, Mackintosh M. Hatching failure increases with severity of population bottlenecks in birds. Proc Natl Acad Sci U S A 2004; 101:558-561.

313. Carr DE, Dudash MR. Recent approaches into the genetic basis of inbreeding depression in plants. Philos Trans R Soc Lond B Biol Sci 2003; 358:1071-1084.

314. Paran I, Zamir D. Quantitative traits in plants: beyond the QTL. Trends Genet 2003; 19:303-306.

315. Sternicki T, Szablewski P, Szwaczkowski T. Inbreeding effects on lifetime in David's deer (Elaphurus davidianus, Milne Edwards 1866) population. J Appl Genet 2003; 44:175-183.

316. Kobyliansky E, Livshits G. Relationship between levels of biochemical heterozygosity and morphological variability in human populations. Ann Hum Genet 1983; 47:215-223.

317. Eldon BJ, Axelsson J, Sigurdsson SB, Arnason E: Cardiovascular risk factors and relatedness in an Icelandic population. Int J Circumpolar Health 2001; 60:499-502.

318. Leutenegger AL, Prum B, Genin E, Verny C, Lemainque A, Clerget-Darpoux F, Thompson EA. Estimation of the inbreeding coefficient through use of genomic data. Am J Hum Genet 2003; 73:516-523.

319. Wit JM. Implications of the Barker hypothesis for general practitioners. Ned Tijdschr Geneeskd 2000; 144:2491-5.

320. Young LE. Imprinting of genes and the Barker hypothesis. Twin Res 2001; 4: 307-317.

321. Preston SH. The changing relation between mortality and the level of economic development. Population Studies 1975; 2:231-248.

322. Caldwell JC. Mortality in relation to economic development. Bull WHO 2003; 81:831-832.

323. Clark D. Urban world, global city. Routledge, UK, 2004.

324. Bittles AH, Petterson BA, Sullivan SG, Hussain R, Glasson EJ, Montgomery PD. The influence of intellectual disability on life expectancy. J Gerontol A Biol Sci Med Sci 2002; 57:M470-472.

APPENDIX 1

Sveučilište u Zagrebu
Medicinski fakultet

To:

Dr Harry Campbell,
Department of Public Health Sciences, Medical School, University of Edinburgh, Edinburgh EH8 9AG, Scotland, UK

and

Dr Igor Rudan,
Department of Medical Statistics, Epidemiology and Medical Informatics, School of Public Health "Andrija Stampar", University Medical School, Rockefellerova 4, 10000 Zagreb, Croatia

Zagreb, 7 May 2001

Re:

Genetics of Complex Diseases and Autochtonous Mendelian Disorders in Eastern Adriatic Islands, Croatia: Preliminary Investigation of a Genetic Isolate Population
(An application for "The Wellcome Trust" International Research Development Award)

The Ethics Committee of the Faculty of Medicine, University of Zagreb, Croatia has reviewed in full the application of the research project named above proposed by two researchers addressed above on 8 May 2001.
Committee members in attendance were: prof.dr.sc. Jadranka Mustajbegović, doc.dr.sc. Vlasta Bradamante, ptof.dr.sc. Draško Šerman, doc.dr.sc. Slobodan Lang, prof.dr.sc. Đurđica Grbeša i prof.dr.sc. Josip Škavić

The reviewed documents included:

1. The full application for the International Research Development Award to the Wellcome Trust (Ref. No. 018057, pages i-ii and 1-18);
2. The annexes to the application (*Annex 1:* Research grants held by sponsor; *Annex 2:* Papers on the eastern Adriatic islands populations published in English in peer-reviewed international journals by the Institute of Anthropological Research in Zagreb, Croatia; *Annex 3:* Preliminary prevalence data on selected complex diseases from 4 villages; *Annex 4:* Details of proposed methods for the study of genetic diversity, linkage disequilibrium and admixture)
3. Information sheet describing details of the proposed research to the prospective examinees.
4. The informed consent form intended for signature of randomly selected examinees for purpose of obtaining DNA sample and assessing genetic diversity in particular villages of interest.

Conclusions: After considering all the aforementioned documents, the Ethics Committee of the Faculty of Medicine, University of Zagreb, Croatia has concluded the following:

(1) From the proposal, it is clear that the obtained personal medical data will be used solely to calculate indicators of disease frequency in chosen populations (such as incidence and prevalence), and no reports on particular individuals and their disease status will be made or distributed in any way;
(2) It is also clear that any other obtained personal information (including genealogical information) will be used solely to calculate indicators of average inbreeding and kinship in chosen populations, and no reports on particular individuals, their personal or genealogical information will be made or distributed in any way;
(3) From the *Annex 4* ("Details of proposed methods for the study of genetic diversity, linkage disequilibrium and admixture") it is clear that the necessary biological (blood) samples will be obtained from randomly selected individuals in chosen populations with previously obtained informed consent to determine genetic diversity of those populations by applying appropriate sequence of molecular genetic studies and no reports on particular individuals and their personal DNA information will be made or distributed in any way.
(4) It is clear that both proposers understand their responsibility and are committed to obtain written informed consent from the examinees, ensure the secure storage of any confidential information and guarantee limited access to data during and after the study period.

Decision: The Ethics Committee of the Faculty of Medicine, University of Zagreb, Croatia has *unanimously decided to grant ethical approval for the proposed International Research Development Award to Dr. Harry Campbell and Dr. Igor Rudan*.

Conditions of the approval are:

(1) The Committee expects to be promptly informed on any new circumstances that may arise during the research and that would affect the aforementioned premises for this decision;
(2) The Committee expects to recieve the copy of the final report on the achievements of the research project upon its completion.

Authority and Compliance: The Ethics Committee of the Faculty of Medicine, University of Zagreb, Croatia confirms that it has the authority to grant the approval for this particular research project. Furthermore, the Committee is fully compliant to the principles, declarations and resolutions of the World Medical Association (The Helsinki Declaration, Recommendations Guiding Medical Doctors in Biomedical Research Involving Human Subjects, Helsinki, 1964, revised in Tokyo, 1975, Venice, 1983 and Hong Kong, 1989), The International Code of Medical Ethics of the World Medical Association (London, 1949, ammended in Sydney, 1968 and Venice, 1983) and The Ethical Codex of the Croatian Medical Association (Zagreb, 1993).

On behalf of the members of the Ethics Committee,

Yours sincerely,

Professor Jadranka Mustajbegovic, MD, PhD
Chairperson
Ethics Committee of the Faculty of Medicine, University of Zagreb, Croatia
Andrija Stampar School of Public Health
Rockefellerova 4, HR-10000 Zagreb, Croatia

Multi-Centre Research Ethics
Committee for Scotland

Secretariat
Deaconess House
148 Pleasance
Edinburgh
EH8 9RS
Telephone 0131 536 9026
Fax 0131 536 9346
www.corec.org.uk

NHS
SCOTLAND

Dr Harry Campbell
Reader in Genetic Epidemiology
Public Health Sciences
University of Edinburgh
Teviot Place
Edinburgh
EH8 9AG

Date: 24 October 2001
Your Ref.:
Our Ref.: MREC/01/0/71

Enquiries to: Walter Hunter
Extension: 89026
Direct Line: 0131 536 9026
Email: walter.hunter@lhb.scot.nhs.uk

Dear Dr Campbell

MREC/1/0/71: Genetics of complex diseases and autochthonous mendelian disorders in Eastern Adriatic Islands, Croatia: Preliminary Investigation of a Genetic Isolate Population.

Thank you for your letter 1 October 2001.

The Chairman of the Multi-Centre Research Ethics Committee for Scotland has considered the amendments submitted in response to the Committee's earlier review of your application on 13 September 2001 as set out in our letter dated 26 September 2001. The document considered was as follows:

Letter of response dated 1 October 2001

The Chairman, acting under delegated authority, is satisfied that these accord with the decision of the Committee and has agreed that there is no objection on ethical grounds to the proposed study. I am, therefore, happy to give you our approval on the understanding that you will follow the conditions of approval set out below. A full record of the review undertaken by the Committee is contained in the attached MREC Response Form. The project must be started within three years of the date on which the Committee's approval is given.

Conditions of Approval

- You must follow the protocol agreed and any changes to the protocol will require prior approval from the Committee.

- If projects are approved before funding is received, the Committee must see, and approve, any major changes made by the funding body. The Committee would expect to see a copy of the

NHS
SCOTLAND

final questionnaire before it is used.

- You must promptly inform the Committee of:
- deviations from or changes to the protocol which are made to eliminate immediate hazards to the research subjects;
- any changes that increase the risk to subjects and/or affect significantly the conduct of the research;
- all adverse drug reactions that are both serious and unexpected;
- new information that may affect adversely the safety of the subjects or the conduct of the trial.

- You must complete and return the standard progress report form to the Committee one year from the date on this letter and thereafter on an annual basis. This form should also be used to notify the Committee when your research is completed.

ICH GCP Compliance

The Committee is fully compliant with the International Conference on Harmonisation/Good Clinical Practice (ICH GCP) Guidelines for the Conduct of Trials Involving the Participation of Human Subjects as they relate to the responsibilities, composition, function, operations and records of an Independent Ethics Committee/Independent Review Board. To this end it undertakes to adhere as far as is consistent with its Constitution, to the relevant clauses of the ICH Harmonised Tripartite Guideline for Good Clinical Practice, adopted by the Commission of the European Union on 17 January 1997. The Standing Orders and a Statement of Compliance were included on the computer disk containing the guidelines and application form and are available on request or on the Internet at http://dspace.dial.pipex.com/mrec.

Yours sincerely

WALTER HUNTER
MREC Administrator
cc: Dr Hans Hagen
Wellcome Trust
183 Euston Road
London
NW1 2BE

Multi-Centre Research Ethics
Committee for Scotland

NHS
SCOTLAND

RESPONSE FORM

DETAILS OF APPLICANT:

1. **Name and address of Principal Researcher:**

Dr Harry Campbell
Reader in Genetic Epidemiology
Public Health Sciences
University of Edinburgh
Teviot Place
Edinburgh
EH8 9AG

2. **Title of project:**

Genetics of complex diseases and autochthonous mendelian disorders in Eastern Adriatic Islands, Croatia: Preliminary Investigation of a Genetic Isolate Population

3. **Name and address of Sponsor:**

Dr Hans Hagen
Wellcome Trust
183 Euston Road
London
NW1 2BE

DETAILS OF MREC:

4. **MREC for Scotland**
Deaconess House
148 Pleasance
Edinburgh
EH8 9RS

5. **MREC Reference Number:** MREC/01/0/71

---

6. **Listed below is a complete record of the review undertaken by the Committee with the decisions made, dates of decisions and the requirements at each stage of the review:**

Date of review: 13 September 2001

Committee members in attendance:

Professor C Gillis (Chairman)(Consultant in Public Health Medicine)
Dr N Anderson (Statistician)
Dr B Holland (Consultant Paediatrician)
Mrs C Horne (Lay)
Professor K Lees (Consultant Physician/Clinical Pharmacologist)
Dr G Masterton (Consultant Psychiatrist)
Mrs H Millar (Lay)
Mrs J Munro (Professions Allied to Medicine)
Professor P Peattie (Nurse)
Professor G Raab (Statistician)
Mr P Rogers (Consultant Surgeon)
Mr I Smith (Lay)
Dr J Webster (Consultant Physician/Clinical Pharmacologist)

Documents reviewed:

Application form signed 2 August 2001
Wellcome Trust application (ref. 018057): submitted December 2000
Subject information sheet: version 1 dated 25.7.2001
Consent form: version 1 dated 25.7.2001
GP letter: version 1 dated 25.7.2001
GP reply: version 1 dated 25.7.2001
Letter to participants: version 1 dated 25.7.2001
Letter of ethical approval from the Ethics Committee, Faculty of Medicine, University of Zagreb
CV

Outcome of review: Approved in principle

Amendments/Information requested:

- that the handling of samples etc held in the UK would be in accordance with UK guidelines.

Documents reviewed by Chairman:

Letter of response dated 1 October 2001

Date approved by Chairman: 10 October 2001

7. THE FINAL DOCUMENTS AND ARRANGEMENTS APPROVED BY THE MREC

The following items have been approved by the Multi-Centre Research Ethics Committee for Scotland:

Subject information sheet: version 1 dated 25.7.2001
Consent form: version 1 dated 25.7.2001
GP letter: version 1 dated 25.7.2001
GP reply: version 1 dated 25.7.2001
Letter to participants: version 1 dated 25.7.2001
Methods of initial recruitment to study
Compensation arrangements for subjects
Payments to researcher
Provision of expenses for subjects

Walter Hunter
Administrator
Multi-Centre Research Ethics Committee for Scotland
Date: 24 October 2001

(FRONT PAGE)

STUDY OF RISK FACTORS FOR COMMON DISEASES IN KOMIZA VILLAGE

Andrija Stampar School of Public Health
Faculty of Medicine, University of Zagreb, 2003

Information Booklet

(PAGE 2)

Dear Sir/Madam,

We are informing you that a team from the **Andrija Stampar School of Public Health, Faculty of Medicine, University of Zagreb,** will be investigating the population of Komiza between **February 15th, 2002 and February 19th, 2002**.

Men from the population are invited to take part on odd dates, and women on even dates, anytime they find it convenient within the period above.

This booklet has been issued to answer the most frequently asked questions about the investigation. If you require any additional information, you can ask:

(a) your local general practitioner, or
(b) contact persons at the Andrija Stampar School of Public Health:

Dr. Igor Rudan	(01)-3011-575, (091)-3011-575
Dr. Zrinka Biloglav	(091)-5088-313

(PAGE 3)

1. Why has Komiza been chosen for this project?

In the first phase of the project in 2002, a health status of several village populations in islands of Rab, Vis, Lastovo, Korcula and Susak was analysed by the team of Andrija Stampar School of Public Health.

Komiza was chosen after a comparative review of health data showed that it would benefit most from the continuation of the project, as the prevalence of health problems such as elevated plasma lipids and blood pressure is the highest.

It was also decided that the history of the population in Komiza and its structure were particularly favourable to allow us to study causes of important causes of disease in the Croatian islands.

This study is carried within the project "Genetic, social and behavioural determinants of health and disease" of the Ministry of Science and Technology of Croatia.

Another important reason was also that the first part of the study in Komiza in February 2002 was among the most successfully completed in all islands, with a good turnout of the population and a great hospitality received.

(PAGE 4)

2. What are the aims of the project?

The project aims to continue to study the causes of the some of the most important causes of disease in the Croatian islands. It will do this by gathering information on health behaviours such as diet and physical activity, by making some health measurement such as blood pressure and taking blood for biochemical and genetic analysis.

By studying all these aspects in detail it is hoped that causes of disease can be found. The long term hope is that this may lead in future to the development of better ways of treating these diseases. The project will also study the population structure of Komiza and compare the frequency of health hazards with several other regions of Croatia.

(PAGE 5)

3. Who will lead this project?

The work in Komiza will be carried out by a team from the Andrija Stampar School of Public Health of the Medical Faculty in Zagreb and Croatian Institute of Public Health.

This research team has very extensive experience carrying out similar work in these islands.

This group will collaborate with doctors and researchers from Scotland from the University of Edinburgh and from the UK Medical Research Council (the main government research agency in the UK). The project is funded jointly by the Ministry of Science and Technology of the Republic of Croatia and the UK Medical Research Council.

(PAGE 6)

4. What will I be asked to do?

We hope to study about 1,000 adults aged 18 years or more in Komiza. This will be entirely on a voluntary basis and with written consent. The research team will make

contact with village authorities to discuss the project and allow questions about the project to be discussed.

You will be given time to make up your mind after reading this information sheet and you will be able to approach team members to discuss any questions you may have before deciding to take part.

If you agree to take part, just sign a consent form at the end of this booklet and bring the booklet with you to investigation any day you find convenient. A team of 24 Croatian researchers (medical doctors, biologists) will then arrange to examine men on odd dates and women on even dates.

(PAGE 7)

5. What examinations will be involved?

The examination will take about 60 minutes and will take place in the mornings. It will involve:

1. completion of a health questionnaire based on standard questionnaires recommended by international bodies such as the World Health Organisation and covering general health (GHQ), disease symptoms and psychological factors (EPQ-R);
2. gathering information of about you (such as age, family members), dietary habits, physical activity, health behaviours (e.g. smoking)
3. Measurement of height, weight and other body measurements such as waist and chest measurements;
4. Measurement of blood pressure (at elbow and ankle) and lung function
5. Obtaining a blood sample for biochemical and genetic tests

---

(PAGE 8)

6. What results will I be given?

The research team will give you details of your height, weight, blood pressure and lung function and details of a large number of biochemical tests that will be carried out in Zagreb: cholesterol, tryglicerides, blood glucose, insulin, haemoglobin, fibrinogen, creatinine, uric acid and calcium.

Results of these tests will be given to your doctor free of charge within one month of your examination.

This project will be carried out in conjunction with your village doctor and any abnormalities found can be discussed with him so that appropriate treatment can be given. Any genetic tests will be solely for research purposes. The members of the research team would keep them in confidence according to European Union Data Protection Act related to biomedical research, which has also been accepted by the University of Zagreb Medical School, so they will not be shared with anyone.

(PAGE 9)

7. How will I benefit from this study?

You will be given a "health check" as described under question 5, and results of your blood pressure and atherosclerosis degree assessment, lung function examinations and biochemical tests in value of at least 350 kn per person will be made available to your personal general practitioner free of charge.

Any new information about the history of your village that we discover will be made available to the village authorities. Your blood sample will be received as a gift and stored in strict confidence, and if used for research purposes it would be in accordance with European Union Data Protection Act, which has also been accepted by the University of Zagreb Medical School.

The data collected in this study may only be used for scientific research into causes of disease in your village. In case of substantial discoveries, the collaboration with other research groups is possible, and even with industry if the information from your village may help in the future lead to the development of new treatments. However, in all these eventualities it may be only under the rules of the European Union Data Protection Act, which guarantees you anonymity of your personal data.

(PAGE 10)

8. What if I already took part in the study last year? And are their plans for further research work in Komiza?

Those who already took part last year are especially invited to repeat their health check-up from 2 reasons. Firstly, we will be able to see if their health status improved or not over 1 year. Secondly, this time we are offering to do more biochemical tests then in the initial round, such as haemoglobin, insulin and calcium tests.

It is important to note that if this study is successful again as the first one, we plan to approach the village authorities again to conduct a third study during the fall of 2003. This will involve:

1. Detailed tests of eyesight and hearing (visual acuity, hearing threshold, corneal tonometry and eye photograph);
2. Biochemical analysis of saliva and urine samples;
3. Measuring heart function and predisposition to heart disease (ECG and measures of blood vessel stiffness using ultrasound waves)

In case of a successful second phase, we will issue a new information booklet on these tests, similar to this one, and ask for consent again later this year. Once again people will be entirely free to choose whether or not to take part.

(PAGE 11)

9. Has the project been given ethical approval and will all information be confidential?

The project has been granted ethical approval by appropriate Ethics Committees in both Croatia and Great Britain. All data will be held on computer in an anonymous fashion so that no names are kept together with personal or health information.

In Croatia, all data will be held securely at the Andrija Stampar School of Public Health in Zagreb and kept in strictest confidence. With your permission your data will be made available to your personal general practitioner who will inform you of the results upon your request.

Hoping that you will decide to help us carry this project similarly successfully as the last time,

Yours sincerely,

Dr Igor Rudan
Andrija Stampar School of Public Health
Faculty of Medicine, University of Zagreb
Rockefellerova 4, 10000 Zagreb, Croatia

(PAGE 12):

I declare that I read and understood the information from this information booklet, and that I'm willing to take part in the investigation.

Name and surname:

_____

Signature:

_____

(Please do not forget to bring this booklet with filled genealogical form attached to the investigations. As our investigators can only study up to 50 persons per day, we ask for your understanding if in case of overcrowding we have to offer you the next most convenient date for your participation).

Consent Form

STUDY OF RISK FACTORS FOR COMMON DISEASES IN KOMIZA
Andrija Stampar School of Public Health
Faculty of Medicine, University of Zagreb, 2002

Principal Investigators in Croatia:

Igor Rudan MD, MPH
Andrija Stampar School of Public Health
Faculty of Medicine, University of Zagreb
Rockefellerova 4, 10000 Zagreb, Croatia

Contact Details for research team:

Igor Rudan
Tel: 01/3011-575, 091/3011-575
Fax: 01/4813-777
E-mail: irudan@mef.hr

Explanation of consent procedure

After you have had time to read the information sheet (version x/xx/2003) discuss the study with the member of our research team and consider whether you might wish to take part in the study, we would be grateful if you would complete this consent form. All individual results will be confidential and only made available to the research team. Blood samples will be coded to remove identifying information.

Please tick all boxes

1 I have read this consent form and information sheet (version x/xx/2003, and have had the opportunity to ask questions about them. ☐

2 I understand that taking part is voluntary and that a decision not to take part will not alter treatment that I would normally receive. ☐

3 I agree to take part in this study:
- a) by completing questionnaires about lifestyle and diet ☐
- b) by giving a family history ☐
- c) by giving a blood sample ☐
- d) by agreeing to storage of that sample for medical research in the future ☐
- e) ☐

4 I give permission for someone from the research team to look at my medical records. I understand that the information will be kept confidential by the research team. ☐

5 I agree that the sample I give and the information gathered about me can be stored by the research team for possible uses in future research, as described in the information sheet. I understand some of these projects may be carried out by other medical research groups (with appropriate ethical approvals first being obtained) including researchers working for commercial companies. ☐

6 I understand that my General Practitioner will be notified about my participation in the study. ☐

7 I understand that I am free to withdraw from the study at any time without giving a reason. ☐

8 I understand that the results of the genetic investigations in this study will have no direct medical benefit to me but rather will help in the overall understanding of the causes of diseases in Croatia. Other results from the medical examination or from blood test that may be of medical benefit to me will be given to my general practitioner within one month for discussion with me. ☐

9 I understand that I will not benefit financially if this research leads to the development of a new treatment or medical test ☐

.....	.....	.....	.....
Name of patient (please print)	**Signature**	**Date**	**Date of Birth**
.....	.....	.....	
Name of researcher	Signature	Date	

Thank you for agreeing to take part in this research.

APPENDIX 2

SVEUČILIŠTE U ZAGREBU
MEDICINSKI FAKULTET
ŠKOLA NARODNOG ZDRAVLJA
"ANDRIJA ŠTAMPAR"
Katedra za Med. Statistiku, Epidemiologiju Med. Informatiku, Katedra za Zdravstvenu Ekologiju i Medicinu Rada, i Katedra za Medicinsku Sociologiju

UNIVERSITY OF ZAGREB
MEDICAL SCHOOL
SCHOOL OF PUBLIC HEALTH
"ANDRIJA ŠTAMPAR"
Department of Med. Statistics, Epidemiology and Med. Informatics, Department of Environmental and Occupational Health and Department of Medical Sociology

FIELD WORK QUESTIONNAIRE:

A. Personal Information 2

B. Basic Health Information 2

C. Genealogical Information 3

D. WHO Angina Questionnaire 4

E. WHO Claudication Questionnaire 5

F. EU Respiratory Health Survey 6

G. WHO Diabetes and Chronic NCD Questionnaire 7

H. Indicators of the Socio-Economic Status 8

I. Indicators of Nutrition 10

J. GHQ-12 15

K. EPQ-R 16

L. Phenotype Measurements 17

M. Biochemical Measurements 18

)ate: _____ Surveyor: _____ Code: _____

. Personal Information:

urname: _____ Name: _____

Iaiden name: _____ Sex: _____

)ate of birth: _____ Place of birth: _____

Iedical record number: _____ Physician: _____

ddress: _____

)ccupation: _____

Iarital status: _____ Number of children: _____

. Basic Health Information:

amily history of diseases:

_____

ersonal disease status (*from medical records – please see criteria in the appendix to this uestionnaire - if positive, please enter the year of diagnosis!*):

levated blood pressure:	_____	Gout:	_____
oronary heart disease:	_____	Glaucoma:	_____
troke:	_____	Rheumatic disease:	_____
chizophrenia:	_____	Kidney disease:	_____
Manic)/Depressive psych.:	_____	Neoplasms:	_____
iabetes:	_____	Peptic ulcer:	_____

ther medical problems: _____

Vhich medication do you take (for diabetes, please specify: diet / insulin – oral / injected):

_____

_____

_____

ave you ever been hospitalised? When exactly and what was the reason?

Genealogical Information

bject: me: ce of birth: ar of birth:		Subject's father: Name: Place of birth: Year of birth:		Subject's paternal grandfather: Name: Place of birth: Year of birth:
bject's spouse: me: ce of birth: ar of birth:		Subject's mother: Name: Place of birth: Year of birth:		Subject's paternal grandmother: Name: Place of birth: Year of birth:
bject's child #1: me: ce of birth: ar of birth:		Subject's sib #1: Name: Place of birth: Year of birth:		Subject's maternal grandfather: Name: Place of birth: Year of birth:
bject's child #2: me: ce of birth: ar of birth:		Subject's sib #2: Name: Place of birth: Year of birth:		Subject's maternal grandmother: Name: Place of birth: Year of birth:
bject's child #3: me: ce of birth: ar of birth:		Subject's sib #3: Name: Place of birth: Year of birth:		Subject's other marriages or sibs (please list the name, place and year of birth for spouse and children):□□
bject's child #4: me: ce of birth: ar of birth:		Subject's sib #4: Name: Place of birth: Year of birth:		
bject's child #5: me: ce of birth: ar of birth:		Subject's sib #5: Name: Place of birth: Year of birth:		

D. WHO Angina Questionnaire:

1. Do you ever feel pain or discomfort in your chest?

YES ___ NO ___

(*if "NO", please proceed to question 8*)

2. Does this pain or discomfort develop when you walk uphill or very fast?

YES ___ NO ___

(*if "NO", please proceed to question 8*)

3. Does this pain or discomfort develop even when you are walking in your normal rhythm on the flat surface?

YES ___ NO ___

4. What do you do when you feel the pain or discomfort in your chest?

You stand still ___ You slow down ___ You keep walking at the same pace ___

5. Does this pain or discomfort go away once you stand still or sit down?

YES ___ NO ___

6. If yes, after how much time?

More than 10 minutes ___ 10 minutes or less ___

7. Can you mark on the chest diagram below where exactly does the pain appear?

8. Have you ever had a large pain across the entire front chest which lasted longer than 30 minutes?

YES ___ NO ___

9. If yes, what was the cause? _____

(Chest diagram for question 7)

WHO Claudication Questionnaire:

Do you ever feel the pain in one or both legs while walking?

YES ___ NO ___

f NO, please proceed to Question 8)

Does this pain ever begin when you stand still or sit?

YES ___ NO ___

Do you ever feel the pain in your lower leg muscles?

YES ___ NO ___

Do you ever feel this pain while walking at a regular pace?

YES ___ NO ___

Does this pain ever go away while you walk?

YES ___ NO ___

What do you do when you feel this pain while walking?

ou stand still ___ You slow down ___ You keep walking at the same pace ___

If you stand still, what happens?

isually lasts longer than 10 minutes ___ it usually goes away in 10 minutes or less ___

Have you ever had a surgery on the arteries or nerves of your legs, with the exception of the nous varices?

YES ___ NO ___

es, what was the reason: _____

_____

Have you ever had a surgery involving the amputation of:

toe (or several of them)

YES ___ NO ___

art of the leg below knee

YES ___ NO ___

art of the leg above knee

YES ___ NO ___

European Union Respiratory Health Survey:

Have you, over the past 12 months, ever had "heavy breathing" or wheeze in your chest?

YES ___ NO ___

f NO, please proceed to Question 4)

When you heard the wheeze, would you also be short of breath?

YES ___ NO ___

Do you have "heavy breathing" or wheeze when free of a cold or the inflammation airways?

YES ___ NO ___

Did you ever wake up due to the feeling that your airways are "narrowed" during the last 12 onths?

YES ___ NO ___

Did you ever wake up due to shortness of breath during the last 12 months?

YES ___ NO ___

Did you ever wake up due to an attack of cough during the last 12 months?

YES ___ NO ___

Did you have an asthma attack during the last 12 months?

YES ___ NO ___

Do you currently take any asthma-related medications (including inhalators, aerosols or tablets)?

YES ___ NO ___

Do you have any allergies of upper respiratory tract or "hay fever"?

YES ___ NO ___

. WHO Diabetes and Chronic Non-Communicable Diseases Questionnaire

Do you smoke?

Yes ___ No ___ Ex-smoker ___

If yes, what do you smoke:

Cigarettes ___ Pipe ___ Cigars ___

How many per day? _____

Over how many years? _____

Have you ever smoked?

YES ___ NO ___

If yes, for how many years and how many per day?

_____

How many years ago did you stop smoking? _____

How many alcohol units do you consume per week? _____

Physical activity during your everyday's work:

sitting ___ light ___ moderate ___ hard ___

. Physical activity during the remainder of the day:

sitting ___ light ___ moderate ___ hard ___

. The female health questionnaire:

Age at menarche: _____
Do you still have regular periods? _____
f not, at what age did they stop? _____
How many deliveries have you had? _____
How many miscarriages? _____
How many spontaneous abortions? _____
How many artificial abortions? _____

. Indicators of the Socio-Economic Status

The highest degree of education:

) no school, incomplete elementary school
) complete elementary school
) craftswork or industrial school
) high school
) faculty (includes higher school, art academy)
) postgraduate study, doctorate
) other (specify: _____)

Career status:

) employed
) self-employed
) unemployed
) pension
) housewife
) student
) supported person
) other (specify: _____)

Job category:

) agriculture
) non-qualified / semiqualified worker
) qualified worker
) clerk
) professional
director, politician
) other (specify: _____)

How would you assess the economic status of your household?

) Much worse than average
) Somewhat worse than average
Same as average
) Somewhat better than average
Much better than average

What is the surface area of your household?

) up to 40 m^2
) 41 - 50 m^2
) 51 - 60 m^2
) 61 - 70 m^2
) 71 - 80 m^2
) 81 - 100 m^2
) 101 - 150 m^2
) more than 150 m^2

Do you live in:

) rented flat
) own flat
) own house
) elsewhere (specify: _____)

The number of distinct rooms in your household (including the living room)?

2 3 4 5 6 and more

Do you have a balcony, terrace or a garden?

) No
) Yes

Which of these do you posess in your household?

ıter pipeline	1	2 TV's	1
lett that can be flushed	1	dishwasher	1
throom	1	computer	1
s/central heating	1	more than 100 books	1
ooden floors	1	art paintings/pottery	1
ephone	1	car	1
CR	1	cottage/2nd appartment	1
ezer	1	boat	1

)CIO-ECONOMIC INDEX (SES) = the sum total of these 16 factors;
w SES (bottom 25% of the population); Medium SES (25-75%); High SES (top 25%)

Indicators of Nutrition:

thin the last month, how many portions of the following foods did you consume in each week
ease use the attached album with photographs to choose correct portion size):

od	Portion Size	Number of Portions
sh		
hite fish fried		
hite fish grilled, poached or boiled		
ly fish fried		
ly fish grilled, poached or boiled		
awns, crabs, etc		
ussels, oysters, cockles, scallops		
uid, octopus fried		
uid, octopus grilled, poached or pickled		
eat & poultry		
ef		
rk or lamb (not cured/salted)		
mmon/bacon (i.e. cured & salted)		
ver		
dney/offal?		
ami/sausage		
rkey/chicken		
ck/goose		
rbohydrate		
tatoes boiled or baked		
tatoes fried		
ce - white		
ce - brown		
sta or couscous		
odles		
ead - white		
ead - brown		
ead containing butter/ oil e.g. croissant?		
za		
sps		
ke - plain		
ke with icing, cream etc		
cuits		
ackers - salt		
iry products		
lk – unpasteurised & bovine		
lk – pasteurised & bovine		
lk – pasteurised, bovine, low fat		
lk - goat		
ghurt		
eam		
eese -store		

eese - homemade		
tter		
gs		
-cream		
her fats & oils		
argarine		
nflower oil		
ive oil		
getables & fruit		
ot exc. potato (carrots, parsnips etc)		
afy (cabbage, spinach, etc.)		
ans, peas		
her green vegetable (broccoli, leeks etc)		
ions, garlic		
eet peppers		
sh fruit (apples, bananas, oranges, etc.)		
ied fruit (raisins, figs etc)		
ts - fresh		
ts – salted & roasted		
inks		
ater from the public pipeline		
ater from the own cistern		
ffee		
a		
uit juice		
rbonated drink (Coca-Cola etc)		
ne		
er		
irits		
iscellaneous		
eets		
ocolate		
tamin/mineral supplements (pharmacy)		
rbal remedy/tonic		
ded salt at table		
ded sugar/honey at table		
ghly salted food		
ghly sweetened food		
me-produced food		
lustry-produced food		

GHQ-12:

e are interested to know if you have had any medical complaints,a nd how your health has been general over the past few weeks. We want to know about present and recent complaints, and not ose that you had in the past. Have you recently:

been able to concentrate on whatever you're doing?

Better than usual | Same as usual | Less than usual | Much less than usual

lost much sleep over worry?

Not at all | No more than usual | Rather more than usual | Much more than usual

felt that you are playing a useful part in things?

More so than usual | Same as usual | Less useful than usual | Much less useful

felt capable of making decisions about things?

More so than usual | Same as usual | Less so than usual | Much less than usual

felt constantly under strain?

Not at all | No more than usual | Rather more than usual | Much more than usual

felt you couldn't overcome your difficulties?

Not at all | No more than usual | Rather more than usual | Much more than usual

been able to enjoy your normal day-to-day activities?

More so than usual | Same as usual | Less so than usual | Much less than usual

been able to face up to your problems?

More so than usual | Same as usual | Less so than usual | Much less able

been feeling unhappy and depressed?

Not at all | No more than usual | Rather more than usual | Much more than usual

. been losing confidence in yourself?

Not at all | No more than usual | Rather more than usual | Much more than usual

. been thinking of yourself as a worthless person?

Not at all | No more than usual | Rather more than usual | Much more than usual

. been feeling reasonably happy, all things considered?

More so than usual | About same as usual | Less so than usual | Much less than usual

EPQ-R:

ease, answer all the questions circling the answer you feel best describes you. Answer the estions honestly and do not spend too much time thinking about them.:

	Yes	No
Does your mood often go up and down?	____	____
Do you take much notice of what people think?	____	____
Are you a talkative person?	____	____
If you say you will do something, do you keep promise no matter how incon.?	____	____
Do you ever feel "just miserable" for no reason?	____	____
Would being in debt worry you?	____	____
Are you rather lively?	____	____
Were you ever greedy by helping yourself to more than your share of anyth.?	____	____
Are you an irritable person?	____	____
. Would you take drugs which may have strange or dangerous effects?	____	____
. Do you enjoy meeting new people?	____	____
. Have you ever blamed someone for doing sth. you knew was your fault?	____	____
. Are your feelings easily hurt?	____	____
. Do you prefer to go your own way rather than act by the rules?	____	____
. Can you usually let yourself go and enjoy yourself at a lively party?	____	____
. Are all your habits good and desirable ones?	____	____
. Do you often feel "fed-up"?	____	____
. Do good manners and cleanliness matter much to you?	____	____
. Do you usually take initiative in making new friends?	____	____
. Have you ever taken anything (pin, button) that belonged to someone else?	____	____
. Would you call yourself a nervous person?	____	____
. Do you think marriage is old-fashioned and should be done away with?	____	____
. Can you easily get some life into a rather dull party?	____	____
. Have you ever broken or lost something belonging to someone else?	____	____
. Are you a worrier?	____	____
. Do you enjoy co-operating with others?	____	____
. Do you tend to keep in the background of social occasions?	____	____
. Does it worry you if you know there are mistakes in your work?	____	____
. Have you ever said anything bad or nasty about anyone?	____	____
. Would you call yourself tense or "highly-strung"?	____	____
. Do you think people spend too much time safeguarding future with sav./ins.?	____	____
. Do you like mixing with people?	____	____
. As a child were you ever cheeky to your parents?	____	____
. Do you worry too long after an embarassing experience?	____	____
. Do you try not to be rude to people?	____	____
. Do you like plenty of bustle and excitement around you?	____	____
. Have you ever cheated at a game?	____	____
. Do you suffer from "nerves"?	____	____
. Would you like other people to be afraid of you?	____	____
. Have you ever taken advantage of someone?	____	____
. Are you mostly quiet when you are with other people?	____	____
. Do you often feel lonely?	____	____
. Is it better to follow society's rules than go your own way?	____	____
. Do other people think of you as being very lively?	____	____
. Do you always prectice what you preach?	____	____
. Are you often troubled about feelings or guilt?	____	____
. Do you sometimes put off until tomorrow what you ought to do today?	____	____
. Can you get a party going?	____	____

L. Measurements of phenotypes:

1. Height: _____ (mm)
2. Weight: _____ (dag)
3. BMI: _____ (kg/cm2)
4. Systolic blood pressure (1st measurement): _____ (mmHg)
5. Diastolic blood pressure (1st measurement): _____ (mmHg)
6. Systolic blood pressure (2nd measurement): _____ (mmHg)
7. Diastolic blood pressure (2nd measurement): _____ (mmHg)
8. Forced vital capacity (FVC): _____
9. Forced expiratory volume (FEV1): _____
10. Skinfold thickness 1: _____ (mm)
11. Skinfold thickness 2: _____ (mm)
12. Waist circumference: _____ (mm)
13. Waist to hip circumference ratio: _____
14. Bioimpedance: _____
15. Ankle brachial pressure index (ABPI): _____

I. Biochemical Measurements:

1. Total cholesterol: _____
2. HDL cholesterol: _____
3. LDL cholesterol: _____
4. Triglycerides: _____
5. Glucose: _____
6. Insulin: _____
7. Hb A1c: _____
8. Fibrinogen: _____
9. Creatinine: _____
0. Calcium: _____
1. Uric acid: _____

APPENDIX 3

Coll. Antropol. **28** (2004) 2: 943–950
UDC 575.222.5-053.88
Review

Five Reasons why Inbreeding May Have Considerable Effect on Post-Reproductive Human Health

Igor Rudan[1,2] and Harry Campbell[2]

[1] Department of Medical Statistics, Epidemiology and Medical Informatics, School of Public Health »Andrija Štampar«, School of Medicine, University of Zagreb, Zagreb, Croatia
[2] Department of Public Health Sciences, The University of Edinburgh Medical School, Edinburgh, Scotland, UK

ABSTRACT

As the genetic architecture of common complex diseases of late onset is emerging through intensive research, it is intriguing to assess the predicted effect of inbreeding on those diseases. In this paper, we propose five reasons why we believe inbreeding may have a considerable effect on post-reproductive human health. (i) The joint effect of inbreeding depression on all polygenic quantitative phenotypes that confer risk for late-onset diseases is predicted to be multiplicative rather than additive. (ii) The »genetic load« of rare »Mendelian« variants with large deleterious effects in post-reproductive adults is unknown, but could be much greater than expected as these variants were invisible to selection through human history. (iii) Deleterious effects resulting from autozygosity in hundreds of affected rare recessive variants of small effect under common disease/rare variant (CD/RV) hypothesis could result in epistatic effects that could jointly impair capacity to compensate against environmental risks. (iv) Heterozygote advantage in loci under balancing selection could be reduced by inbreeding. (v) Published empirical evidence in animals and humans consistently report large inbreeding effects on late-onset traits. Since inbreeding is common in many populations and the effects of inbreeding depression could substantially contribute to disease burden and reduced life expectancy we believe there is now a clear need for further genetic epidemiological research in humans to investigate this issue.

***Key words**: inbreeding, consanguinity, late-onset diseases, complex diseases, post-reproductive age, inbreeding depression, genetic load, balancing selection, heterosis*

Introduction

In many parts of the developing World and in many communities within the developed World large proportions of all marriages are still among close relatives.

Received for publication September 3, 2004

The reasons for this include geographic, tribal, cultural or religious isolation or socio-economic motivation such as preservation of property, particularly land rights[1,2]. The degree of inbreeding in offspring of such marriages can be measured by the genetic term »inbreeding coefficient« (F), which indicates the proportion of the autosomal genome which is expected to be homozygous through inheritance of identical genes from common ancestors (i.e. proportion of alleles identical by descent (IBD) or »autozygosity«). The F value is calculated from genealogical information and it amounts to about 6% in the offspring of first cousin parents and 25% in the offspring of incestuous unions of first-degree relatives[1,3]. The apparent risk in the individuals with a considerable proportion of their genes homozygous for identical allelic variants is the occurrence of »Mendelian« (monogenic) diseases caused by rare and recessive deleterious autosomal mutations of large effect[4–9].

Research on the effects of inbreeding on human health has historically focused on early-onset diseases, mainly recessively inherited monogenic (Mendelian) diseases, birth defects, decreased fertility and early mortality. This was due to the widespread recognition that consanguineous unions were more likely to result in genetic diseases of children, most of which had a distinctive phenotype that was readily identifiable. Therefore, the great majority of research on inbreeding effects had been focused on pre-reproductive health problems, and the risks have been thoroughly evaluated by numerous groups and individual authors[3–12].

However, the genetic architecture underlying late-onset diseases such as cardiovascular diseases, cancer, adult-onset diabetes and psychiatric disorders, which represent the major health burden globally, is still a matter of open debate[13–18]. A genetic model that is finding increasing support from both animal experiments and human studies is one in which the genetic variants underlying complex chronic diseases are more likely to be rare rather than common in the population. They are also likely to be numerous (highly polygenic architecture) and of a small individual effect[13,18]. If this view of the genetic architecture of common complex diseases is correct then it would be important to consider the predicted effect of inbreeding. In this paper, we put forward and discuss five reasons why we believe inbreeding may have a considerable effect on post-reproductive human health.

Reason #1: The deleterious effects of inbreeding depression on quantitative (endo)phenotypes that confer risk for late-onset diseases may be multiplicative

Inbreeding depression is a recognised phenomenon that is common to polygenic traits in all living organisms[19]. It is thought to result from increased homozygosity of recessive alleles that act in the same direction at loci that influence the phenotype of interest (»directional dominance«)[20]. In an inbred individual, inbreeding depression is predicted to affect many polygenic endophenotypes (quantitative [patho]-physiological intermediates involved in physiological or disease processes). Many of these are established risk factors for late-onset diseases, such as blood pressure, body mass index, cholesterol and glucose levels and bone mineral density. A substantial effect of inbreeding acting to increase human blood pressure has been shown directly in at least four studies[21–24], and implied indirectly in several more studies[25–28]. Effects on body mass index and cholesterol levels have also been implied[25,26,28]. Similarly, effects on various measures of intelligence have been consistently shown[29–31], and in this issue we report on an effect on cortical index, a

predictor of susceptibility to osteoporosis[32]. It is known that the risk of e.g. increased blood pressure, body mass index and cholesterol levels on cardiovascular diseases is not threshold-dependent, but is increasing across the entire range of values observed in the population[33]. Thus, one important consequence is that inbred individuals are expected to be at slightly increased risk relative to the outbred general population to develop a late- onset disease, regardless of the absolute measurement of their blood pressure, body mass index and cholesterol. Furthermore, even if the effect of inbreeding depression on each of those phenotypes individually was rather small, it is known that the concurrent presence of several risk factors for the same disease increases risks in a multiplicative rather than additive manner. Therefore, the joint effect of inbreeding depression on all the potential quantitative phenotypes that confer risk to late-onset disease during lifetime could be more substantial than widely appreciated[34,35].

Reason #2:
Effects of inbreeding on rare variants with large effect in post-reproductive adults (»invisible Mendelian diseases« of late onset)

Inbreeding is predicted to have larger effects on the population-attributable fraction of disease if the underlying variants are rare rather than common. This is because common recessive variants will occasionally become homozygous in the population by chance, without a need for inbreeding to bring them together. If the variants are very rare in the population, and inbreeding is almost the only realistic scenario under which they can become homozygous in an individual, then the fraction of disease cases in the population who are the offspring of related parents will be much larger. This was shown to be the case with population attributable fraction of early-onset monogenic (Mendelian) diseases in the presence of inbreeding: it has been shown that the prevalence of autosomal dominant Mendelian disorders is constant in all world populations, but the prevalence of autosomal recessive Mendelian disorders is increased by 3–4-fold in regions where inbreeding is prevalent.[2,7] Therefore, the great majority of Mendelian disease that is caused by rare recessive variants of large effect and early age of onset is due to inbreeding in those countries. However, these diseases manifest in pre-reproductive period, so they are »visible« to selection. Although these variants continuously arise through mutations, most of the affected cases never reproduce, so they are being effectively removed from the gene pool by selection. Thus their overall public health burden is reasonably low, and Bittles and Neel estimated that each human carries about 1.4 such recessive lethal mutations in the genome[12].

However, rare variants of large effect also act in Mendelian fashion to cause late-onset complex diseases. For nearly all late--onset diseases, clustering in families has been reported and, in some, rare high- penetrance variants have been found which are associated with an extremely increased lifetime risk of disease[14–16]. Examples of this include variants in BRCA1 and BRCA2 and breast cancer[36], hMLH1 and hMSH and colorectal cancer[37], and GCK (glucokinase) and maturity-onset diabetes of the young (MODY) diabetes[38]. In large outbred populations, it is estimated that up to 15% of disease cases such as cardiovascular, cancer, diabetes and psychiatric disorders cluster in families, while 85% or more are due to combined effects of polygenic susceptibility and cumulative environmental exposures[39]. However, in countries where inbreeding is common, the prevalence of all recessively inherited monogenic forms of complex diseases could

be expected to increase by at least as much as seen jointly for all recessive early-onset Mendelian diseases, i.e. 3–4-fold. In theory, this could greatly increase both the overall disease prevalence and the proportion the late-onset disease burden caused by rare recessive variants.

It is also known that these rare variants can affect quantitative phenotypes, such as blood pressure or cholesterol (familial hypertension, familial hypercholesterolaemias)[25,40], which represent »invisible Mendelian diseases« as their phenotypes are not clinically apparent. These quantitative phenotypes are associated with increased morbidity and mortality from diseases of public health importance such as stroke and coronary heart disease. It is conceivable that there are many more such »invisible« Mendelian diseases affecting metabolic pathways at different levels and predisposing individuals to complex diseases in the post-reproductive period. The joint effect of inbreeding on all these variants could be expected to be of a magnitude of at least that seen for early-onset Mendelian diseases. Indeed there are reasons to believe that the number of rare and recessive Mendelian variants with large effect on late-onset complex diseases much larger than the estimated number with early effects. These rare variants have accumulated in the genome through mutations that are either neutral in early life, or even beneficial, but show deleterious effects in post-reproductive period (»antagonistic pleiotropy«)[34,35]. There is no known mechanism that would be expected to remove these mutations from the genome or act against their accumulation, as they are invisible to selection. Thus there are cogent reasons why the effects of inbreeding on late-onset Mendelian diseases should be carefully considered in the same way as has been done for early-onset diseases[12].

Reason #3:
Autozygosity in many rare recessive variants of small effect could result in epistatic effects that could jointly impair capacity to compensate against environmental risks.

Modest levels of inbreeding observed in human populations are expected to have much larger effects on the population distributions of polygenic traits than on oligogenic traits and diseases. This is because an excess in autozygosity of 6.25% of the genes in human genome (i.e. about 2,000 genes), which would be expected in a child from a first-cousin marriage, will lead to autozygosity of rare recessive mutations of small effect in those 2,000 genes. In a polygenic trait, it is expected that some of the genes that determine its expression would be affected even with only 6.25% of the genome autozygous. We have argued that the genetic component of late-onset diseases may be due principally to large numbers of rare variants in numerous genes – the common disease/rare variant (CD/RV) hypothesis.[13] Recent estimates[41] imply that each person carries, on average, 500–1200 slightly deleterious mutations, most of which are rare and present in heterozygous form. In an offspring of first-cousin marriage, 30–75 of these variants would be expected to become homozygous, with uncertain effects[42]. If the mutations are numerous and of small deleterious effect, their autozygosity throughout the genome might not lead to apparent syndromes of early onset, but may mildly impair the function of affected genes. As a result, the compensatory potential to oppose the harmful environmental stimuli would be non-specifically impaired. This impairment of homeostasis or repair capacity could lead to an earlier age at diagnosis of a late-onset complex disease. This is consistent with the model of these diseases arising over long periods of time, and becoming clini-

cally apparent when the compensatory potential is exhausted[19,35].

This theoretical mechanism is difficult to study in humans, but has been clearly demonstrated in animals, where a greater sensitivity of homeostatic mechanisms to inbreeding in later life has been suggested[19,35]. In an experiment with inbred and non-inbred mice strains, the two strains did not show large differences in survival when the animals were kept in the laboratories. However, when both groups were released into their natural habitat, the inbred mice strains had a dramatically reduced chance of survival in comparison to non-inbred group[43]. The effects of inbreeding are therefore thought to be much greater in natural populations (exposed to a less uniform and more challenging environment) than in those studied in laboratories.

Reason #4: Heterozygote advantage in loci under balancing selection is expected to reduce by inbreeding

At some genomic loci, there may not be variants present in a population that are clearly deleterious, but the heterozygous genotype may be more favourable than either homozygous genotype. This effect is widely known as the »heterozygote advantage«, »hybrid vigour« or simply »heterosis«. The effects of heterosis usually act in an opposite way from those of inbreeding depression and they have been demonstrated in humans[44,45], and widely in animals and plants[46,47]. The type of selection that tends to maintain more than one allele in the population at intermediate frequencies, thus maximising the frequency of heterozygous genotypes in a population, is known as »balancing selection«[13]. It is clear that balancing selection probably has important role in shaping gene diversity in the genes that are important for defence against unknown and unpredictable environmental risks, such as infectious diseases[44]. Populations that are more genetically diverse are at less risk from diverse environmental threats (since it is more likely that someone would carry a rare protective variant)[48]. It is therefore likely that inbreeding leading to autozygosity in several hundred genes will affect some of these genomic loci under balancing selection, thus reducing the beneficial effects of heterosis in those individuals. This mechanism could be more important than generally thought, since recent evidence suggests that loci under balancing selection may be surprisingly common in the genome[49–52].

Reason #5: Empirical evidence of inbreeding effects in humans and other organisms

The most extensive research into the effects of inbreeding in general, and particularly on genetic variation related to senescence has been carried out in *Drosophila spp*. A review of 25 years of this research has concluded that deleterious alleles generated by mutation and kept at low frequency by selection contribute between 33% and 67% of the genetic variation in a typical trait. This supports a polygenic model of genetic architecture of most phenotypes and suggests that the common disease / rare variant mechanism contributes to a substantial share of complex disease aetiology[20,53]. A recently published experiment in *Drosophila spp*[34] showed that genetic variation and inbreeding effects increase dramatically with age, supporting these hypotheses. Numerous recent studies of other animals, some of them performed in populations of large mammals, have also consistently reported that inbreeding negatively affected key components of fitness, resulting in increased morbidity and decreased life span[54–57]. A meta-analysis[58] and a critical

review and re-examination of these studies[51] have both concluded that, although unexpected and in some aspects against current understanding, their findings could not be easily dismissed on grounds of publication bias or apparent flaws in methodology.

Finally, our own intensive review of human literature since 1965 has managed to identify very few case-control studies of late-onset diseases in which inbreeding status was not determined by self-reporting, and disease status determination was based on a clear diagnostic criteria which did not change during the period of study. These studies investigated the effects of inbreeding on coronary heart disease[59,60], cancer[61,62], psychiatric disorders[63] and Alzheimer's disease[64]. There was only one longitudinal epidemiological study investigating the effects of inbreeding on 10 complex late-onset diseases[65]. All seven studies reported considerable relative risks associated with inbreeding, typically between 2.0–5.0, which persisted after adjustment for known or suspected confounding factors. Although the available evidence is surprisingly sparse, it appears to support the hypothesis that inbreeding could have a considerable effect on human health and disease occurrence in post-reproductive age adults.

Conclusion

We have argued that there is a coherent theoretical basis for a role for inbreeding in diseases of public health importance in humans. Available data from animal and plant studies strongly support the disease mechanisms put forward in this paper and suggest that these phenomena may be common across species. Available data on the effects of inbreeding in humans has focused on assessing the risk of early mortality due to rare recessive deleterious mutations. In an extensive review of inbreeding in the PubMed database from 1965 to date (nearly 10,000 references), we were able to find very few publications on inbreeding effects on late onset traits or diseases in humans. It is possible that this may be explained to some extent by the fact that in areas of the world where inbreeding is prevalent, late-onset diseases have not until recently represented the main public health problem (e.g. Mediterranean countries, parts of India and sub-Saharan Africa). In western societies, however, inbreeding is not prevalent enough to be studied in a large-scale epidemiological investigation. Nevertheless, we call for more genetic epidemiological research in humans to address this potential problem, and invite related papers from all regions of the world where this issue can be studied. We believe this to be an important epidemiological risk to evaluate, as with improving life expectancy in large human populations where inbreeding is prevalent these effects could substantially contribute to disease burden and life expectancy.

Acknowledgements

This paper was partly supported through research grants from The Wellcome Trust (IRDA) to H.C. and I.R., The Royal Society UK to H.C. and I.R., and the Croatian Ministry of Science, Education and Sport to I.R. (0108330). I.R. gratefully acknowledges support from the University of Edinburgh, Medical Research Council UK and Overseas Research Scheme UK.

REFERENCES

1. BITTLES, A. H., H. S. SAVITHRI, G. MURTHY, G. BASKARAN, W. WANG, J. CAHILL, N. A. RAO, In: MACBETH, H., P. SHETTY: Health and Ethnicity. (Taylor and Francis, London, 2001). — 2. MODELL, B., A. DARR, Nat. Rev. Genet., 3 (2002) 225. — 3. EDITORIAL, Lancet, 338 (1991) 85. — 4. GIBBONS, A., Science, 259 (1993) 1252. — 5. BITTLES, A. H., Dev. Med. Child. Neurol., 45 (2003) 571. — 6. BITTLES, A. H., W. M. MASON, J. GREENE, N. A. RAO, Science, 252 (1991) 789. — 7. WORLD HEALTH ORGANIZATION: Primary health care approaches for the prevention and control of congenital and genetic disorders. WHO Meeting Report, Cairo, Egypt, 6–8 December 1999. WHO/HGN/WG/00.1. (WHO, Geneva, 2000). — 8. ZLOTOGORA, J., Am. J. Med. Genet., 109 (2002) 70. — 9. JORDE, L. B., Hum. Hered., 52 (2001) 61. — 10. SCHULL, W. J., J. V. NEEL, Am. J. Hum. Genet., 24 (1972) 425. — 11. KHLAT, M., M. KHOURY, Epidemiol. Rev., 13 (1991) 28. — 12. BITTLES, A. H., J. V. NEEL, Nat. Genet., 8 (1994) 117. — 13. WRIGHT, A., B. CHARLESWORTH, I. RUDAN, A. CAROTHERS, H. CAMPBELL, Trends Genet., 19 (2003) 97. — 14. MERIKANGAS, K. R., N. RISCH, Science, 302 (2003) 599. — 15. REICH, D. E., E. S. LANDER, Trends Genet., 17 (2001) 502. — 16. BOTSTEIN, D., N. RISCH, Nat. Genet., 33 Suppl (2003) 228. — 17. LOHMUELLER, K. E., C. L. PEARCE, M. PIKE, E. S. LANDER, J. N. HIRSCHHORN, Nat. Genet., 33 (2003) 177. — 18. PRITCHARD, J. K., N. J. COX, Hum. Mol. Genet., 11 (2002) 2417. — 19. CHARLESWORTH, B., D. CHARLESWORTH, Genet. Res., 74 (1999) 329. — 20. CHARLESWORTH, B., K. A. HUGHES, In: SINGH, R. S., C. B. KRIMBAS (Eds): Evolutionary genetics: From molecules to Morphology, Vol 1. (Cambridge University Press, Cambridge, UK, 1999, pp. 369–392). — 21. KRIEGER, H., Am. J. Hum. Genet., 21 (1969) 537. — 22. MARTIN, A. O., T. W. KURCZYNSKI, A. G. STEINBERG, Am. J. Hum. Genet., 25 (1973) 581. — 23. RUDAN, I., N. SMOLEJ-NARANČIĆ, H. CAMPBELL, A. CAROTHERS, A. WRIGHT, B. JANIĆIJEVIĆ, P. RUDAN, Genetics, 163 (2003) 1011. — 24. BADDARUDDOZA, Coll. Antropol., 28 Suppl 2 (2004) 311. — 25. HURWICH, B. J., B. ROSNER, N. NUBANI, E. H. KASS, F. I. LEWITTER, Am. J. Epidemiol., 115 (1982) 646. — 26. THOMAS, J. D., M. M. DOUCETTE, D. C. THOMAS, J. D. STOECKLE, Lancet, 2 (1987) 377. — 27. WAHID SAEED, A. A., F. J. AL SHAMMARY, T. A. KHOJA, T. J. HASHIM, C. C. ANOKUTE, S. B. KHAN, J. Hum. Hypertens., 10 (1996) 583. — 28. HALBERSTEIN, R. A., Hum. Biol., 71 (1999) 659. — 29. MORTON, N. E., Proc. Natl. Acad. Sci. USA, 75 (1978) 3906. — 30. RUDAN, I., D. RUDAN, H. CAMPBELL, Z. BILOGLAV, R. UREK, M. PADOVAN, L. SIBBETT, B. JANIĆIJEVIĆ, N. S. NARANČIĆ, P. RUDAN, Coll. Antropol., 26 (2002) 421. — 31. BASHI, J., Nature, 266 (1977) 440. — 32. RUDAN, I., T. ŠKARIĆ-JURIĆ, N. SMOLEJ-NARANČIĆ, B. JANIĆIJEVIĆ, D. RUDAN, I. MARTINOVIĆ KLARIĆ, L. BARAĆ, M. PERIČIĆ, R. GALIC, M. LETHBRIDGE-CEJKU, P. RUDAN, Coll. Antropol., 28 (2004) (in press). — 33. ROSE, G.: The strategy of preventive medicine. (Oxford University Press, Oxford, UK, 1992). — 34. HUGHES, K. A., J. A. ALIPAZ, J. M. DRNEVICH, R. M. REYNOLDS, Proc. Natl. Acad. Sci. USA, 99 (2002) 14286. — 35. CHARLESWORTH, B., K. A. HUGHES, Proc. Natl. Acad. Sci. USA, 93 (1996) 6140. — 36. PHAROAH, P. D., A. ANTONIOU, M. BOBROW, R. L. ZIMMERN, D. F. EASTON, B. A. PONDER, Nat. Genet., 31 (2002) 33. — 37. CURTIS, C. D., M. GOGGINS, Methods Mol. Med., 103 (2004) 123. — 38. VELHO, G., P. FROGUEL, K. CLEMENT, M. E. PUEYO, B. RAKOTOAMBININA, H. ZOUALI, P. PASSA, D. COHEN, J. J. ROBERT, Lancet, 340 (1992) 444. — 39. BALMAIN, A., J. GRAY, B. PONDER, Nat. Genet., 33 Suppl (2003) 238. — 40. COHEN, J. C., R. S. KISS, A. PERTSEMLIDIS, Y. L. MARCEL, R. McPHERSON, H. H. HOBBS, Science, 305 (2004) 869. — 41. FAY, J. C., G. J. WYCKOFFAND, C. I. WU, Genetics, 158 (2001) 1227. — 42. CARLBORG, O., C. HALEY, Nat. Rev. Genet., 5 (2004) 618. — 43. JIMENEZ, J. A., K. A. HUGHES, G. ALAKS, L. GRAHAM, R. C. LACY, Science, 266 (1994) 271. — 44. PENN, D. J., K. DAMJANOVICH, W. K. POTTS, Proc. Natl. Acad. Sci. USA, 99 (2002) 11260. — 45. WOLANSKI, N., E. JAROSZ, M. PYZUK, Soc. Biol., 17 (1970) 1. — 46. EBERT, D., C. HAAG, M. KIRKPATRICK, M. RIEK, J. W. HOTTINGER, V. I. PAJUNEN, Science, 295 (2002) 485. — 47. IVES, A. R., M. C. WHITLOCK, Science, 295 (2002) 454. — 48. SPIELMAN, D., B. W. BROOK, R. FRANKHAM, Proc. Natl. Acad. Sci. USA, 101 (2004) 15261. — 49. MERILA, J., B. SHELDON, S. GRIFFITH, Ann. Zool. Fenn., 40 (2003) 269. — 50. VAN OOSTERHUT, C., M. K. VAN HEUVEN, P. M. BRAKEFIELD, Mol. Ecol., 13 (2004) 1025. — 51. BALLOUX, F., W. AMOS, T. COULSON, Mol. Ecol., 13 (2004) 3021. — 52. HANSSON, B., H. WESTERDAHL, D: HASSELQUIST, M. AKESSON, S. BENSCH, Evolution, 58 (2004) 870. — 53. MACKAY, T. F., Curr. Opin. Genet. Dev., 14 (2004) 253. — 54. COULSTON, T., S. ALBON, J. SLATE, J. PEMBERTON, Evolution, 53 (1999) 1951. — 55. ACEVEDO-WHITEHOUSE, K., F. GULLAND, D. GREIG, W. AMOS, Nature, 422 (2003) 35. — 56. FOERSTER, K., K. DELHEZ, A. JOHNSON, J. LIFJELD, B. KEMPENAERS, Nature, 425 (2003) 714. — 57. KELLER, L., D. WALLER, Trends Ecol. Evol., 17 (2002) 230. — 58. COLTMAN, D., J. SLATE, Evolution, 57 (2003) 971. — 59. SHAMI, S. A., R. QAISAR, A. H. BITTLES, Lancet, 338 (1991) 954. — 60. ISMAIL, J., T. H. JAFAR, F. H. JAFARY, F. WHITE, A. M. FARUQUI, N. CHATURVEDI, Heart, 90 (2004) 259. — 61. SIMPSON, J. L., A. O. MARTIN, S. ELIAS, G. E. SARTO, J. K. DUNN, Am. J. Obstet. Gynecol., 141 (1981) 629. — 62. LEBEL, R. R., W. B. GALLAGHER, Am. J. Med. Genet., 33 (1989) 1. — 63. ABASKULIEV, A. A., G. V. SKOBLO, Genetika, 11 (1975) 145. — 64. VEZINA, H., E. HEYER, I. FORTIER, G. OUELLETTE, Y. ROBITAILLE, D. GAUVREAU, Genet. Epidemiol., 16 (1999) 412. — 65. RUDAN, I., D. RUDAN, H.

CAMPBELL, A. CAROTHERS, A. WRIGHT, N. SMOLEJ-NARANČIĆ, B. JANIĆIJEVIĆ, L. JIN, R. CHAKRABORTY, R. DEKA, P. RUDAN, J. Med. Genet., 40 (2003) 925.

I. Rudan

Department of Medical Statistics, Epidemiology and Medical Informatics, School of Public Health »Andrija Štampar«, School of Medicine, University of Zagreb, Rockefellerova 4, 10000 Zagreb, Croatia
irudan@hotmail.com

PET RAZLOGA ZBOG KOJIH BI SRODIVANJE MOGLO ZNAČAJNO UTJECATI NA LJUDSKO ZDRAVLJE NAKON ZAVRŠETKA GENERATIVNOG RAZDOBLJA

SAŽETAK

Kako intenzivnim istraživanjima počinjemo nazirati genetsku arhitekturu kompleksnih bolesti starije dobi, zanimljivo je razmotriti kako bi srodivanje trebalo utjecati na pojavnost tih bolesti. U ovom članku, predlažemo pet razloga zašto bi srodivanje moglo imati znatan utjecaj na ljudsko zdravlje nakon završetka generativnog razdoblja. (i) Ukupan učinak »depresije srodivanjem« na sve poligenski određene kontinuirane ljudske fenotipove koji se povezuju s rizikom za bolesti starije dobi trebao bi biti multiplikativan, a ne aditivan. (ii) »Genetski teret« rijetkih (tzv. Mendelskih) alela s jakim negativnim učinkom na zdravlje u post-generacijskom razdoblju nije poznat, no mogao bi biti znatno veći od očekivanog jer su te varijante bile nevidljive utjecajima selekcije tijekom ljudske povijesti. (iii) Nepoželjni učinci kao rezultat autozigotnosti u stotinama zahvaćenih rijetkih recesivnih genetskih varijanti malog učinka u okviru hipoteze »česta bolest/rijetka varijanta«, gdje bi rezultirajući epistatski učinci mogli zajednički umanjiti sposobnost kompenziranja protiv okolišnih čimbenika rizika. (iv) Srodivanje bi moglo utjecati na gubitak povoljnih učinaka heterozigotnosti na lokusima koji su pod utjecajem balansirajuće selekcije. (v) Sistematski pregled rijetkih empirijskih dokaza u literaturi u eksperimentalnih životinja i ljudi konzistentno upućuje na snažne učinke srodivanja na svojstva karakteristična za stariju dob. Pozivamo na dodatna genetičko-epidemiološka istraživanja u ljudskim populacijama kako bi se ovaj problem istražio, jer rastom očekivanog trajanja života u velikim područjima svijeta gdje je srodivanje učestalo, navedeni bi učinci mogli značajno pridonijeti ukupnom pobolu i umiranju od bolesti starije dobi.

Inbreeding and the Genetic Complexity of Human Hypertension

Igor Rudan,*,† Nina Smolej-Narancic,‡ Harry Campbell,*,1 Andrew Carothers,§ Alan Wright,§ Branka Janicijevic‡ and Pavao Rudan‡

Department of Community Health Sciences, University of Edinburgh Medical School, Edinburgh EH8 9AG, Scotland, United Kingdom, School of Public Health "Andrija Stampar," University Medical School, 10000 Zagreb, Croatia, ‡Institute for Anthropological Research, 10000 Zagreb, Croatia and §MRC Human Genetics Unit, Western General Hospital, Edinburgh EH4 2XU, Scotland, United Kingdom

Manuscript received June 28, 2002
Accepted for publication November 19, 2002

ABSTRACT

Considerable uncertainty exists regarding the genetic architecture underlying common late-onset human diseases. In particular, the contribution of deleterious recessive alleles has been predicted to be greater for late-onset than for early-onset traits. We have investigated the contribution of recessive alleles to human hypertension by examining the effects of inbreeding on blood pressure (BP) as a quantitative trait in 2760 adult individuals from 25 villages within Croatian island isolates. We found a strong linear relationship between the inbreeding coefficient (F) and both systolic and diastolic BP, indicating that recessive or partially recessive quantitative trait locus (QTL) alleles account for 10–15% of the total variation in BP in this population. An increase in F of 0.01 corresponded to an increase of ~3 mm Hg in systolic and 2 mm Hg in diastolic BP. Regression of F on BP indicated that at least several hundred (300–600) recessive QTL contribute to BP variability. A model of the distribution of locus effects suggests that the 8–16 QTL of largest effect together account for a maximum of 25% of the dominance variation, while the remaining 75% of the variation is mediated by QTL of very small effect, unlikely to be detectable using current technologies and sample sizes. We infer that recent inbreeding accounts for 36% of all hypertension in this population. The global impact of inbreeding on hypertension may be substantial since, although inbreeding is declining in Western societies, an estimated 1 billion people globally show rates of consanguineous marriages >20%.

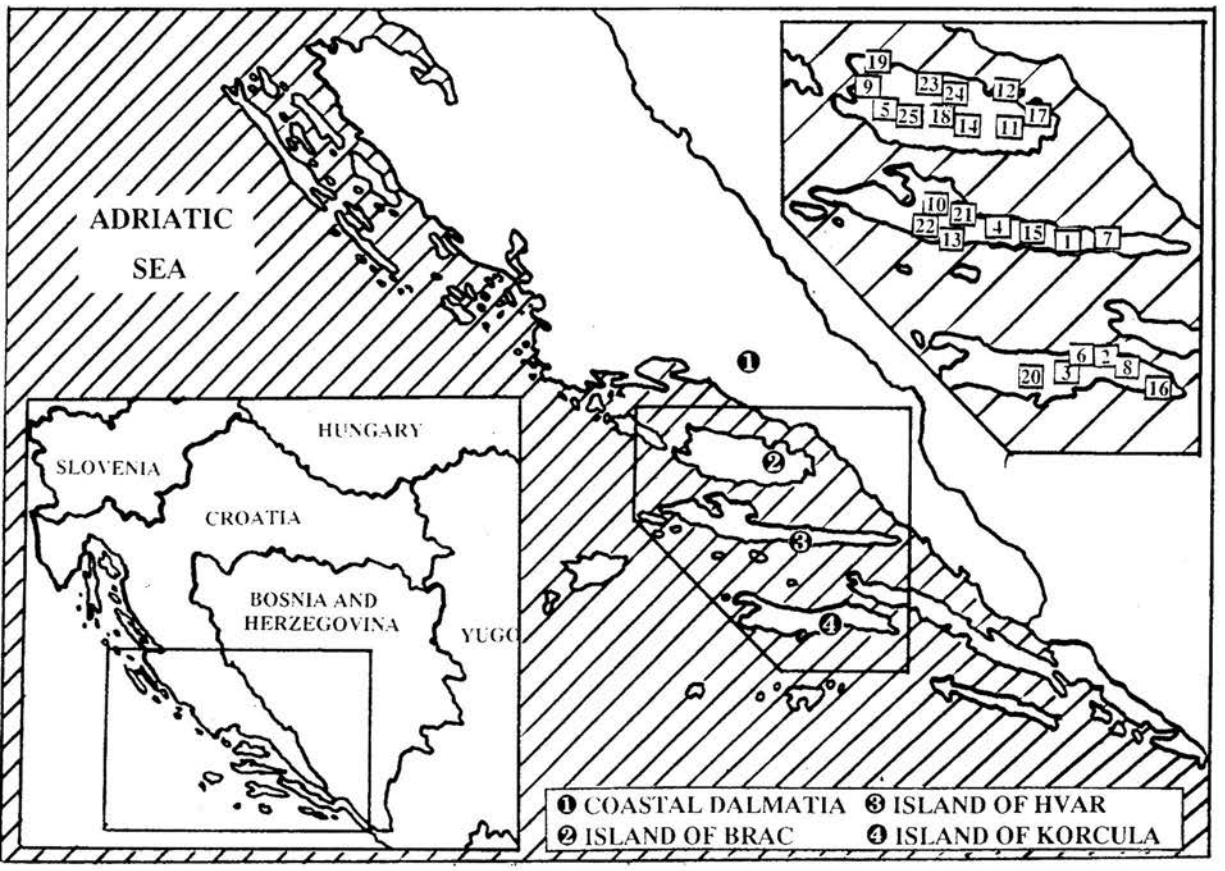
THE extensive literature on the health effects of inbreeding has largely focused on its impact on roduction, childhood mortality, and Mendelian dis-ers (BITTLES *et al.* 1991; BITTLES and NEEL 1994). narkably little has been published on the effects of reeding on genetically complex late-onset disorders t account for most of the public health burden of ease. This is despite the observation in other species t the deleterious effects of inbreeding may increase ı age, suggesting greater sensitivity of homeostatic chanisms to inbreeding in later life (CHARLESWORTH l HUGHES 1996; CHARLESWORTH and CHARLES-RTH 1999).

We postulated that the quantitative trait, blood pres-e (BP), and the related late-onset disorder, essential ertension, might be mediated by recessive and par-ly recessive quantitative trait locus (QTL) alleles, ch would be influenced by the increased homozygos-found in inbred individuals. In support of this hy-hesis, several studies of small inbred communities ldwide have reported an increased prevalence of ertension (KRIEGER 1968; MARTIN *et al.* 1973; HUR-H *et al.* 1982; THOMAS *et al.* 1987; WAHID SAEED *et al.* 1996; HALBERSTEIN 1999). In addition, analogous observations have come from experiments in inbred ("spontaneous") and engineered animal models of hypertension (STOLL *et al.* 2000). To investigate the relationship between inbreeding and BP we studied a large population sample from well-characterized genetic isolates from the Dalmatian islands, Croatia (Figure 1).

Corresponding author: Department of Community Health Sciences, versity of Edinburgh Medical School, Teviot Pl., Edinburgh EH8 , Scotland, UK. E-mail: harry.campbell@ed.ac.uk

SUBJECTS AND METHODS

Study population: The village populations of three neighboring islands in the eastern Adriatic, Middle Dalmatia, Croatia (Brac, Hvar, and Korcula—see Figure 1) represent well-characterized genetic isolates. Over 100 publications describe the ethnohistory, migration patterns, genealogical reconstruction, biological trait measurements, disease prevalence, and environmental and sociocultural characteristics of this population (RUDAN *et al.* 1987, 1992, 1999; WADDLE *et al.* 1998). Population genetic characteristics of the study population based on a number of serogenetic polymorphisms were reported by ROBERTS *et al.* (1992) and JANICIJEVIC *et al.* (1994). Subsequent analyses of variable number of tandem repeat and short tandem repeat DNA polymorphisms and mtDNA characterized genetic variation in specific islands (MARTINOVIC *et al.* 1998, 1999; KLARIC *et al.* 2001a,b; TOLK *et al.* 2001). The results indicated that



URE 1.—Map of middle Dalmatia, Croatia, showing villages and islands.

e populations in these islands have preserved sepa- characteristics over the course of history to the nt day. Measures of genetic kinship and genetic ıces revealed isolation of individual villages or vil- groups from each other and from the mainland. fic village clustering was noted on Brac and Hvar ls, which coincided with known historic processes. ppreciable degree of genetic homogeneity within udied villages has been noted, which is especially for the most geographically isolated villages. The ages chosen for this study were founded during one ee periods: the BC era (by admixture of Illyrians, s, and succeeding Romans), the 7th century AD roats who immigrated from Asia), and the 16th th centuries AD (by Croats who left the Balkan sula fearing Ottoman expansion). The subse- tendency toward inbreeding in each village has influenced by geographic isolation, political ("*Pas-*") privileges given to residents of certain communi- nd by sociocultural factors (RUDAN *et al.* 1992). island populations present a range of inbreeding rns at both individual and subpopulation levels, as nented in previous studies reporting endogamy, ny, mating choice, genealogical information, and ic marker distributions. High inbreeding levels been implicated in at least three Mendelian disor- ders characterized in neighboring island populations: Mal de Meleda in Mljet (FISCHER *et al.* 2001), hereditary dwarfism in Krk (KOPAJTIC *et al.* 1995), and hereditary mental retardation in Susak (BOHACEK 1964). Measures of diet and lifestyle factors show restricted variation in this population, suggesting its suitability for genetic studies of hypertension (RUDAN *et al.* 1992).

Blood pressure and other measurements: We measured blood pressure, height, and weight between 1979 and 1981 in 2760 adult individuals selected at random from voting lists from 25 isolate villages on three islands (Brac, Hvar, and Korcula) in middle Dalmatia, Croatia (representing a 20% sample of the village populations). In addition, we collected data on body mass index, diet, education level, occupation, and smoking status. This was carried out with the informed consent of participants by the Institute for Anthropological Research in Zagreb, Croatia, in collaboration with the Smithsonian Institute in Washington, DC. None of the examinees had ever received antihypertensive treatment. Blood pressure was measured by a single observer in local health centers and dispensaries between 6 AM and 12 noon following standard procedures as described by Weiner (WEINER and LOURIE 1969). BP values were adjusted for the major determinants of BP (age, height, weight, and smoking status in the analyses) and were reported separately in

les and females. Hypertension was defined as systolic ≥160 or diastolic BP ≥95 mm Hg.

Computation of individual inbreeding coefficients: A gle researcher (I. Rudan) computed individual in-eding coefficients independently and blind to BP us for each study participant on the basis of pedigree ormation on four ancestral generations (five genera-ns where these occurred over a similar time frame) orded during the initial field work and supple-nted by study of parish registries stored in local rches during 1997–2000. The individual inbreeding fficients (F) were then computed according to ight's path method,

$$F = \sum_{i=1}^{c} \left(\frac{1}{2}\right)^{m_i + n_i + 1}$$

RIGHT 1922), where m_i and n_i refer to the number paths from the ith common ancestor, and c refers to number of common ancestors. The genealogical reeding coefficient for each village was then com-ed as the average of all individual F values. To further port these estimates, F was calculated from isonymy proposed by CROW (1980),

$$F = \frac{S}{4} + \left(1 - \frac{S}{4}\right)\frac{P - S}{4(1 - S)},$$

ere $S = \Sigma p_k q_k$, p_k, q_k are the frequencies of the surname males and females, respectively, P is the proportion narriages between spouses carrying the same sur-ne among all marriages, and the summation is over surnames. We calculated average inbreeding mea-es for each of the 25 villages on the basis of isonymy, ch provides an upper bound (ROGULJIC *et al.* 1997) ble 1).

tatistical analysis and modeling: Comparisons of BP ong *villages* were based on systolic and diastolic BP asurements adjusted for age, body mass index (weight/ ght^2), and smoking status. A step-down multiple re-ssion analysis was performed using MINITAB 12.21 ware to investigate the correlation between *individ-* BP measurements and inbreeding coefficients. The del explored the relationship between systolic and tolic blood pressure (as dependent variables) and a nber of explanatory variables: individual inbreeding fficient (F), island and village of residence, smoking us, and the major known risk factors for hyperten-—age, sex, (log-transformed) height, and (log-trans-ned) weight. Variables that made the least contribu-to the explained variation were dropped one at a e until all the remaining variables were statistically ificant (defined as $P < 0.05$ for main effects and 0.01 for higher-order effects; Table 2). A model developed from quantitative genetic theory to derive wer bound, n_L, for the number of genetic loci of ivalent effect contributing to the dominance vari-e in BP, as

$$n_L = \frac{D_T^2}{V_G} = \frac{D_T^2}{H^2 \cdot V_P}, \quad (1)$$

where D_T is the overall slope of the regression on F, V_G is the total genetic variance, V_P is the total phenotypic variance, and H^2 is the broad-sense heritability (see the APPENDIX). This extends to multiallelic loci the result given by CHARLESWORTH and HUGHES (1999) for the biallelic case and is valid except in the unlikely case of strong overdominance at all, or most, loci. To correct for possible unobserved background inbreeding preceding the earliest generation of which we had knowledge, we inflated F values by a factor equal to the ratio of the mean of village inbreeding levels based on isonymy methods to the mean based on pedigree methods.

Population-attributable fraction: The population-attributable fraction (PAF) for hypertension (defined as either systolic >160 mm Hg or diastolic >90 mm Hg) was calculated by multiple logistic regression allowing for individual differences in the variables: village, sex, age, height, weight, and smoking. We determined the appropriate regression as a function of all associated variables (including F) and then noted each individual's probability of being hypertensive if their F was set equal to 0. The sum of all such probabilities, P_{sum}, is an estimate of the number affected in the absence of inbreeding, but with other variables remaining unaltered. Then PAF = $1 - P_{sum}/N_{aff}$, where N_{aff} is the actual number affected.

Modeling the effects of individual QTL loci: For biallelic loci, the relation between the true number, n say, of recessive QTL loci affecting a trait and n_L (see above) is $n = n_L(1 + \gamma^2)$, where γ denotes the coefficient of variation of the frequency distribution of locus effects. Following ZENG (1992), we modeled this as gamma with parameter $L \leq 1$; *i.e.*, $f(x) = x^{L-1}e^{-x}/\Gamma(L)$. This family of distributions has $\gamma^2 = L^{-1}$. Since the contribution of a biallelic locus with nonadditive effect, x (or D_j in the notation of the APPENDIX), to the dominance variance is just x^2, the distribution of such contributions is also gamma, but with parameter $L + 2$. Hence, for given L, we can compute the minimum proportion of loci contributing any specified proportion of the overall variance. Finally for given n_L, we obtain an estimate of the actual minimum number of loci by multiplying this proportion by $n_L(1 + L^{-1})$. As shown by ZENG (1992) this number is relatively insensitive to L in the range $1/16 \leq L \leq 1$.

RESULTS

Measurements recorded during a survey in 1979–1981 in an *untreated* population permitted analysis of BP as a quantitative trait. Body mass index, diet, education level, occupation, smoking status, and inbreeding values among study participants are shown in Table 1 by village of residence. The prevalence of hypertension among

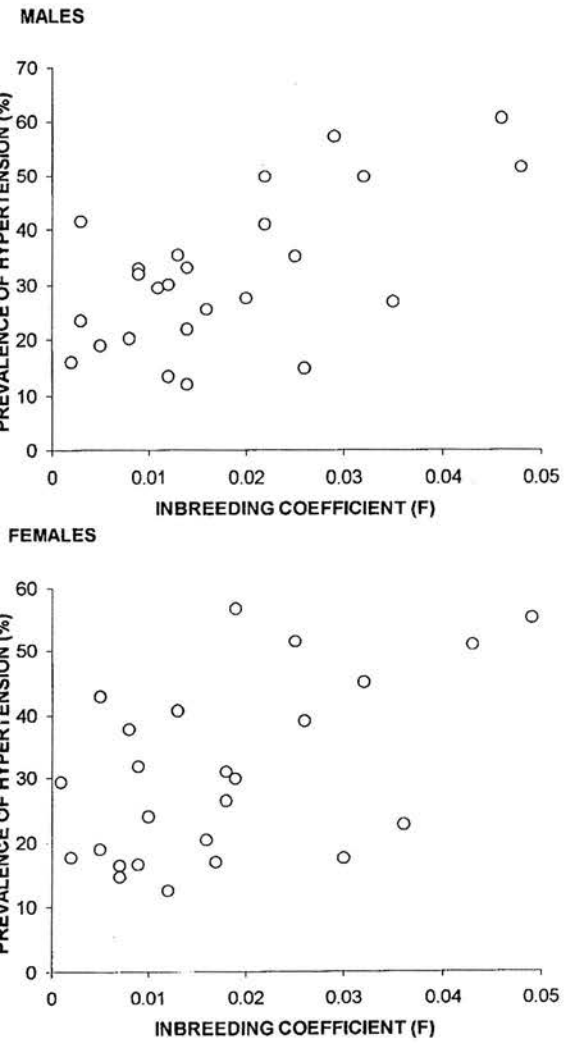
TABLE 1

Ranking of study villages by mean value of *F* computed from genealogical information [*F* (gen)] and isonymy [*F* (iso)]

Village[a]	Sex	*N*	Age [× (SD)]	*F* (gen)	*F* (iso)	Edu (%)	Ocu (%)	Die (%)	Smo (%)	BMI [× (SD)]
Gdinj (H)	M	54	53.6 (10.9)	0.048	0.107	3.7	92.6	94.4	9.3	24.4 (3.3)
	F	76	53.1 (11.7)	0.049		1.3	72.4	93.4	1.3	25.6 (4.7)
Pupnat (K)	M	46	42.8 (12.9)	0.046	0.034	2.2	89.1	95.7	13.0	27.3 (3.6)
	F	51	44.4 (11.8)	0.043		2.0	90.2	88.2	5.9	26.1 (3.7)
Cara (K)	M	63	44.3 (11.7)	0.035	0.040	3.2	90.5	95.2	36.5	26.1 (3.7)
	F	74	45.6 (12.5)	0.030		0.0	87.8	94.6	10.8	26.6 (3.8)
Poljica (H)	M	28	42.3 (9.1)	0.032	0.037	0.0	89.3	92.9	10.7	25.5 (2.7)
	F	20	43.0 (7.6)	0.032		0.0	85.0	85.0	0.0	25.4 (4.2)
Dracevica (B)	M	20	46.8 (16.4)	0.026	0.031	5.0	90.0	90.0	35.0	27.0 (4.2)
	F	22	50.9 (14.2)	0.036		0.0	81.8	90.9	13.6	30.0 (4.3)
Racisce (K)	M	40	43.6 (11.2)	0.029	0.034	2.5	82.5	92.5	25.0	27.8 (4.0)
	F	64	43.2 (11.8)	0.026		1.6	90.7	78.1	21.9	27.0 (4.1)
Bogomolje (H)	M	48	57.6 (15.9)	0.025	0.030	2.1	87.5	95.8	8.3	22.4 (2.6)
	F	33	57.7 (13.4)	0.025		3.0	78.8	90.9	0.0	23.9 (4.1)
Zrnovo (K)	M	98	45.0 (13.0)	0.022	0.029	5.1	79.6	92.3	42.9	26.3 (2.9)
	F	104	45.0 (13.9)	0.019		2.9	78.9	85.6	19.2	26.8 (4.3)
Lozisca (B)	M	17	56.9 (12.2)	0.022	0.029	5.9	82.4	88.2	35.3	26.3 (4.3)
	F	55	49.9 (14.1)	0.018		1.8	72.7	90.9	9.1	28.7 (4.4)
Vrbanj (H)	M	65	38.2 (10.4)	0.020	0.010	4.6	90.8	95.4	10.8	26.5 (2.7)
	F	53	42.3 (9.6)	0.018		1.9	86.8	88.7	1.9	26.8 (3.9)
Selca (B)	M	117	54.4 (14.4)	0.016	0.026	4.3	78.6	89.7	41.9	26.8 (3.3)
	F	185	52.0 (14.9)	0.016		2.2	76.8	90.8	17.8	27.7 (4.4)
Povlja (B)	M	25	56.3 (14.7)	0.014	0.024	0.0	84.0	92.6	40.0	27.6 (3.2)
	F	47	54.1 (14.3)	0.017		2.1	85.1	87.2	10.6	27.5 (4.8)
Vrisnik (H)	M	52	38.0 (10.4)	0.012	0.023	3.8	90.4	94.2	19.2	24.6 (2.6)
	F	44	42.4 (9.6)	0.019		0.0	72.7	88.6	6.8	26.5 (3.6)
Gornji Humac (B)	M	32	51.0 (12.5)	0.014	0.016	3.1	87.5	90.6	59.4	25.7 (3.5)
	F	48	47.4 (13.6)	0.012		2.1	77.1	85.4	25.0	28.3 (4.5)
Zastrazisce (H)	M	70	49.6 (14.8)	0.013	0.013	4.3	85.7	94.3	12.9	24.7 (2.8)
	F	54	50.7 (12.1)	0.013		3.7	87.0	92.6	0.0	25.6 (3.9)
Lumbarda (K)	M	54	43.4 (11.3)	0.014	0.025	7.4	85.1	90.7	37.0	26.8 (2.7)
	F	54	43.0 (12.4)	0.010		1.9	85.2	88.9	5.6	25.6 (3.2)
Novo Selo (B)	M	27	46.6 (16.6)	0.011	0.014	0.0	85.2	88.8	33.3	26.6 (3.9)
	F	30	54.6 (13.3)	0.009		0.0	80.0	86.7	16.7	28.5 (4.3)
Praznice (B)	M	53	48.7 (13.2)	0.012	0.016	3.8	88.7	94.3	45.3	27.2 (3.8)
	F	67	53.0 (14.6)	0.007		1.5	82.1	89.5	10.4	28.9 (4.5)
Sutivan (B)	M	61	50.2 (15.3)	0.009	0.012	6.6	72.1	90.2	27.9	27.9 (3.7)
	F	88	50.5 (15.1)	0.009		3.4	68.1	89.8	23.9	28.0 (4.6)
Smokvica (K)	M	52	43.2 (10.5)	0.009	0.019	3.8	84.6	90.4	42.3	25.8 (3.0)
	F	45	45.7 (12.5)	0.008		2.2	77.8	71.1	11.1	26.2 (3.6)
Svirce (H)	M	74	40.2 (11.0)	0.008	0.008	2.7	91.2	93.2	18.9	25.6 (2.5)
	F	75	38.5 (10.9)	0.007		1.3	69.3	73.3	6.7	25.8 (3.5)
Dol Hvarski (H)	M	53	41.1 (10.7)	0.005	0.004	5.7	79.2	94.3	13.2	26.5 (2.5)
	F	42	42.4 (9.9)	0.005		2.4	71.4	76.2	2.4	25.6 (3.9)
Skrip (B)	M	47	51.4 (10.8)	0.003	0.008	2.1	85.1	93.6	31.9	27.7 (3.2)
	F	56	49.8 (14.5)	0.005		1.8	83.9	87.5	16.1	29.6 (4.7)
Dol Bracki (B)	M	41	44.7 (15.3)	0.003	0.005	2.4	85.4	92.7	29.3	28.0 (3.2)
	F	24	41.8 (15.5)	0.001		0.0	79.2	75.0	25.0	27.8 (4.3)
Nerezisca (B)	M	50	44.7 (12.9)	0.002	0.004	4.0	66.0	90.0	54.0	27.3 (3.4)
	F	62	44.9 (11.9)	0.002		0.0	56.5	71.0	19.4	28.5 (5.0)

Column heads show the number of male and female examinees in each village (*N*), means and standard deviations for age (Age), average inbreeding coefficient computed from genealogical data [*F* (gen)] and isonymy [*F* (iso)] as described in SUBJECTS AND METHODS, the proportion with some college education (Edu), proportion working in agriculture and fishery (Ocu), proportion regularly consuming traditional mediterranean diet (Die), proportion of smokers (Smo), and the means and standard deviations for body mass index (BMI).

[a] H, Hvar; K, Korcula; B, Brac.



IGURE 2.—Relationship between average inbreeding coefnts (F) computed from genealogical information and valence of hypertension in 25 study villages.

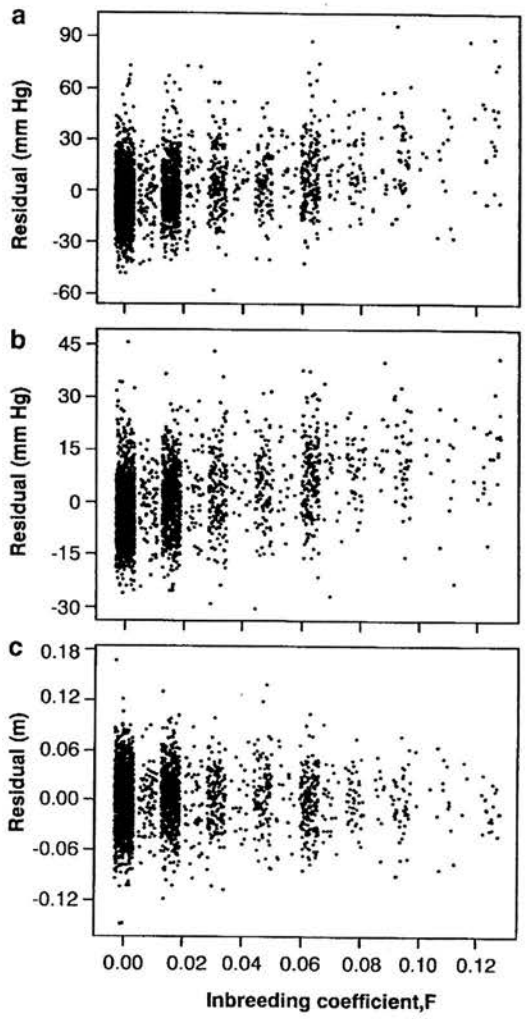


FIGURE 3.—Multiple regression analysis: plot of residuals against inbreeding for (a) systolic BP and (b) diastolic BP after adjusting for effects of village, sex, age, height, weight, and smoking and (c) height after adjusting for village, sex, and age.

viduals with no known inbreeding in their recent estry in the study population was ~20%, and the mean s of those males and females were, respectively, 45.9 13.9) and 47.0 (SD 13.9) years. Average inbreeding asures for each of the 25 villages based on Wright's n method and isonymy gave a consistent pattern of king of villages by level of inbreeding. This supports use of F values as a means of ranking individuals villages by inbreeding coefficient (Table 1).

Ve found a highly significant linear correlation been mean inbreeding coefficient of study individuals ach village and the prevalence of hypertension (Fig-2). To explore this further, we performed multiple ression analysis of systolic and diastolic BP on individnbreeding coefficients (F), controlling for the main ognized determinants of BP (age, sex, height, and ght), village of residence, and smoking status. We nd a strong linear correlation between F and aded systolic and diastolic BP in both males and females (Figure 3). Both systolic and diastolic BP levels correlated positively with age, weight, and individual inbreeding coefficients and negatively with height and smoking status in both males and females. The regression model explained 35–50% of the phenotypic variance in BP. The strongest effect was clearly individual inbreeding coefficients, which alone explained ~15% of the variance in males and 10% in females in both systolic and diastolic levels (Table 2). An increase in F of 0.01 corresponded to an increase of ~3 mm Hg in systolic and 2 mm Hg in diastolic BP in both sexes.

The effect of inbreeding (F) on BP depends on the number and dominance properties of QTL alleles, their frequencies, and average effects on the trait (MUKAI *et al.* 1974; FALCONER and MACKAY 1996). Using result (1) and taking the total phenotypic variance as an upper limit for the value of V_D, we found the genetic component of blood pressure variability in this population to

TABLE 2

Summary of multiple regression analysis for systolic and diastolic BP in males and females: percentage decrease in residual mean square (rms) under different models

	Males				Females			
Model	d.f.	rms[a]	% decr.	% decr. (adj.)[b]	d.f.	rms[a]	% decr.	% decr. (adj.)[b]
Systolic BP								
Model A	1232	210.8	41.1	38.5	1392	249.7	52.6	49.9
Model B	1256	221.6	38.1	36.6	1417	264.9	49.7	47.8
Model C	1257	268.4	25.0	23.3	1418	310.0	41.1	38.9
Model D	1286	357.8	—	—	1472	526.6	—	—
Diastolic BP								
Model A	1232	69.7	46.5	44.2	1416	74.8	48.7	46.6
Model B	1256	72.3	44.6	43.2	1441	78.4	46.2	45.0
Model C	1257	95.5	26.7	25.1	1442	92.5	36.5	35.2
Model D	1286	130.4	—	—	1472	145.7	—	—

Model A, full model, allowing F, F^2, and all two-way interactions between F and other factors; model B, allowing F, but excluding F^2 and all two-way interactions between F and other factors; model C, excluding all terms in F; model D, null model. Models A, B, and C also allow all other factors and two-way interactions, apart from F.

[a] In units of (mm Hg)^2.

[b] % decr. (adj.) is the percentage decrease in rms compared with the rms from model D, but adjusting for the reduction in degrees of freedom.

fluenced by not less than several hundred recessive with 405 and 306 loci for systolic BP and 615 and oci for diastolic BP in males and females, respec-

e distribution of recessive QTL effects can be ap- mated as gamma-type with mode at zero and pa- ter $L < 1$ (ZENG 1992). From our data, if the ated minimum QTL number is ~400, and L is een $1/16$ and 1, then the minimum numbers contrib- the upper 25th and 50th percentiles of the distri- n are, respectively, 8–16 and 30–55 (Figure 4).

ight was analyzed in a similar fashion since in many lations it shows additive variance but no major nance component (KRIEGER 1968; TAMBS *et al.*). The results showed that the slope of the regres- of F on height did not differ significantly from zero, edicted (Figure 3), supporting our interpretation e data.

DISCUSSION

s widely recognized that essential hypertension is r considerable genetic influence. However, apart isolated successes in mapping rare monogenic loci, n account for <5% of hypertension, no major prog- ıas been made in defining the genetic basis of tial hypertension (LIFTON *et al.* 2001). A common, implicit, assumption in mapping studies of such lex traits is that relatively few genetic loci of moder- large effect account for a large component of nderlying genetic variance despite the paucity of rical data to support this. We have demonstrated that the effects of recessive QTL on BP are widespread, accounting for 10–15% of the total variation in BP in this population. These effects are attributable to a very large number of loci (at least 300–600), which will almost certainly show a range of effects on BP.

The model makes several assumptions that may influence these estimates. First, the inbreeding coefficient is based on measures of recent inbreeding (over four to five generations). We therefore calculated isonymy estimates for each village (Table 1) and found that their mean value exceeded the median F value by a factor of 1.35. Since isonymy is widely recognized to overestimate inbreeding (TAY and YIP 1984), this represents an upper bound to the inbreeding estimate. Inflating the F values by 135% in the model reduces the above estimates of minimum QTL numbers by a factor of 1.35^2 (=1.83). On the other hand, if, as seems likely, V_D/V_P is nearer 33% than the 100% assumed here, the effect would be to triple the estimates. CAVALLI-SFORZA and BODMER (1971), for example, estimated V_D/V_P to be 0.38 and 0.33 for systolic and diastolic blood pressure, respectively, using the data of MIALL and OLDHAM (1963). Second, as in many genetic models, all loci were assumed to have equal effects, whereas both theory (BRINK 1967) and empirical data in animals (MACKAY 2001) show that the QTL effects vary widely and may even be of opposite sign. Ignoring this again results in underestimating the true number of QTL. Third, if a substantial proportion of the phenotypic variance is due to epistatic effects, additive and dominance variances may be upwardly biased and lead to an underestimation of the number of QTL. However, many studies suggest

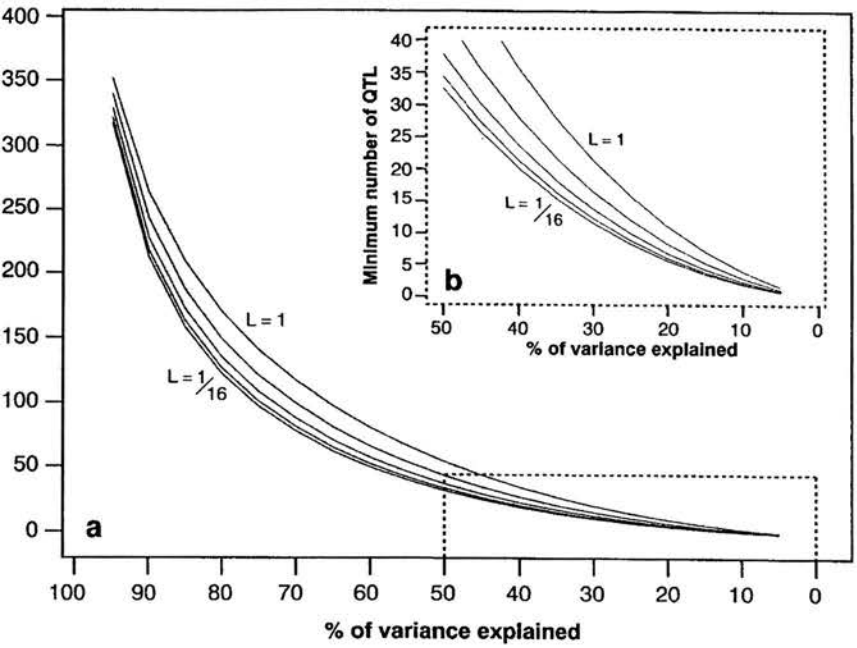


FIGURE 4.—The effect of L on the minimum numbers of loci contributing specified proportions of the overall dominance variance, assuming $n_L = 400$.

% BP variance	10%	25%	50%	75%	90%
No. QTL	2-4	8-16	33-55	98-142	215-266

epistatic QTL effects are uncommon (MACKAY 1). Finally, we assume that recombination between acent loci is sufficiently frequent that the identity-descent (IBD) status of any locus can be considered ependent of its neighbors. In effect, the method ts tightly linked loci as a single "superlocus" (FLINT MOTT 2001), leading to further underestimation he true number of loci.

he magnitude of the inbreeding effect on BP is large uivalent to a rise in systolic BP of ~20 mm Hg and tolic of ~12 mm Hg in offspring of first-cousin mar-es; $F = 0.0625$) but very similar to the only other two lished estimates we could identify in other isolate ulations. KRIEGER (1968) found a 35 mm Hg in-ase in diastolic BP associated with a 0.1 increase in a study of 3465 children in Brazil and MARTIN *et* 1973) reported a 7–28 mm Hg increase in systolic in adult Hutterites associated with an increase in F .0625. This may be because inbreeding has a greater uence on late-onset traits than on traits that are ect to early selection (CHARLESWORTH and HUGHES 6). It is also possible that low environmental varia-, or underestimation of F due to individuals being ted through multiple lines of descent, contributes to size of inbreeding effect in these isolate populations IEGER 1968; MARTIN *et al.* 1973; HALBERSTEIN 1999; EY *et al.* 2001). Thus the observed effect size may ess in more environmentally diverse or outbred pop-ions. The unidirectionality of the effect is also strik-and consistent with a linear unidirectional effect seen in an S-Leut isolate population (MARTIN *et al.* 1973), but the mechanism is unclear. A change in BP with inbreeding is predicted as a consequence of recessive or partially recessive variants with the direction of change toward the value of the more recessive alleles. Physiological homeostasis may also act to support a directional change in BP, for example, through selection against variants tending to reduce BP to maintain circulatory viability. Directional dominance may also occur in late-onset traits when environmental factors are directional (*e.g.*, increase in adult blood pressure due to dietary salt) and when selective constraints are weak compared with blood pressure maintenance in early life.

The estimate of several hundred recessive QTL relevant to human hypertension is realistic and indeed may be conservatively low. It is consistent with a complex and genetically highly variable (HALUSHKA *et al.* 1999) system of blood pressure control mediated by cardiac output, blood vessel architecture, renal function, and central nervous system integration and requiring the interaction of homeostatic systems, including baroreceptors, natriuretic peptides, renin-angiotensin-aldosterone, kinin-kallikrein, adrenergic receptors, and local vasodilator mechanisms (LIFTON *et al.* 2001). Furthermore, published work from animal models of hypertension supports a polygenic rather than oligogenic basis for hypertension (LIFTON *et al.* 2001) and yet these models probably underestimate the genetic complexity, since they are typically bred to achieve fixation of a small subset of the diversity found in wild populations

T and MOTT 2001). The greater genetic complex-
a diverse and outbred human population would
to be self-evident, despite the fact that humans
less haplotype and polymorphic diversity than sev-
ther species, including other primates (REICH *et*
01).

r minimum estimates of the number of recessive
influencing blood pressure control do not in them-
reveal the relative magnitudes of locus effects.
e is, however, good evidence for an L-shaped (lep-
totic) distribution of allelic-effect sizes (SHRIMP-
nd ROBERTSON 1988; TANKSLEY 1993; BOST *et al.*
HAYES and GODDARD 2001; MACKAY 2001; BAR-
nd KEIGHTLEY 2002). In addition, as shown by
(1992), their distribution can be approximated
mma-type with mode at zero (*i.e.*, with parameter
), implying that most loci contribute little to the
l genetic variation and that the number contribut-
large proportion is both small and relatively insen-
to *L*. The model developed from our data predicts
inimum QTL numbers contributing the upper
nd 50th percentiles of the distribution are, respec-
8–16 and 30–55 (Figure 4). Thus, the QTL with
rgest effect account individually for a small propor-
f the total dominance variation and 50–75% of
ariation is mediated by many QTL of very small
, which are probably undetectable using current
ods (TERWILLIGER and GORING 2000).

s study demonstrates an important effect of in-
ing on the genetically complex late-onset disor-
ypertension, which appears to be mediated by a
number of recessive QTL alleles as a result of
sed homozygosity. Several factors support the va-
of the data and reinforce the conclusions: first,
ndard measurement procedures that were adopted
ie exclusion of known confounding factors; sec-
he consistency of findings in diverse populations
GER 1968; MARTIN *et al.* 1973; HALBERSTEIN 1999);
the linear increase in BP with increasing F (preva-
of hypertension rises by 10% for every increment
f 0.01 up to $F = 0.06$); fourth, the overall strength
effect; fifth, the existence of biologically plausible
anisms, all of which point to a causal relationship
en inbreeding and hypertension. Moreover, the
tency of the observation in a random sample of
duals across 25 villages is not explicable by a kin-
ffect. In terms of health impact, the results show
6% of hypertension incidence in this population
e attributed to inbreeding (population-attribut-
action). The population prevalence of hyperten-
mong individuals with no known inbreeding in
ecent ancestry is ~20%, similar to most outbred
ations, but it increases steeply among 50-year-olds
inbreeding coefficient rises (Figure 5).

eeding is generally decreasing among nonimmi-
Western societies but it is highly prevalent globally.
nguineous marriages, defined as a union between
duals related as second cousins or closer (equiva-

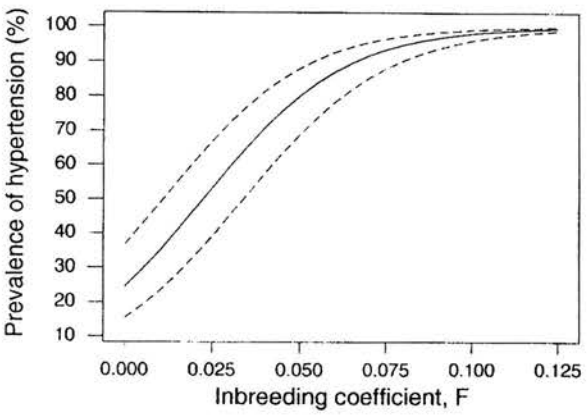


FIGURE 5.—Prevalence of hypertension as a function of inbreeding. The data are shown for nonsmokers of age 50 years, height 1.7 m, and weight 75 kg and with mean village effect, after adjusting for the effects of these variables by binary logistic regression. The dashed lines are 95% confidence limits.

lent to $F \geq 0.0156$ in their progeny), has been conservatively estimated to occur at 1–10% prevalence among 2.811 billion and at 20–50% prevalence among 911 million people globally (BITTLES 1988; BITTLES *et al.* 2001). In addition, the extent of homozygosity by descent in outbred populations may have been underestimated (BROMAN and WEBER 1999). The global impact of inbreeding on hypertension (and stroke) could therefore be significant in health economic terms. In addition, the results provide new insights into the genetic architecture of a common disorder, which should inform and improve the design of QTL-mapping studies and explain some of the observed differences in trait distributions among different populations.

The authors thank Professor Bill Hill, Professor Brian Charlesworth, and Dr. Peter Visscher for helpful discussions, comments, and suggestions. This work was supported by the Wellcome Trust (IRDA) grant to H.C. and I.R., the Croatian Ministry of Science and Technology (CMST) grant 01960101 to P.R., 0196005 to P.R., 0196001 to N.S.N., and 0108330 to I.R., and the joint British Council and CMST grant ALIS 054 to H.C. and I.R. I.R. was supported by funds from the UK Medical Research Council, the University of Edinburgh, and the Overseas Research Scheme.

LITERATURE CITED

ABNEY, M., M. S. MCPEEK and C. OBER, 2001 Broad and narrow heritabilities of quantitative traits in a founder population. Am. J. Hum. Genet. **68:** 1302–1307.

BARTON, N. H., and P. D. KEIGHTLEY, 2002 Understanding quantitative genetic variation. Nat. Rev. Genet. **3:** 11–21.

BITTLES, A. H., 1988 Empirical estimates of the global prevalence of consanguineous marriage in contemporary societies. Working Report 74, Morrison Institute for Population and Resource Studies, Stanford University, Stanford, CA.

BITTLES, A. H., and J. V. NEEL, 1994 The costs of human inbreeding and their implications for variations at the DNA level. Nat. Genet. **8:** 117–121.

BITTLES, A. H., W. M. MASON, J. GREENE and N. A. RAO, 1991 Reproductive behaviour and health in consanguineous marriages. Science **252:** 789–794.

BITTLES, A. H., H. S. SAVITHRI, H. S. V. MURTHY, G. BASKARAN, W.

WANG *et al.*, 2001 Consanguinity: a familiar story full of surprises, pp. 68–78 in *Health and Ethnicity*, edited by H. MACBETH and P. SHETTY. Taylor and Francis, London.
IACEK, N., 1964 Tristan da Cunha and Susak. Lijec. Vjesn. **86:** 1412–1416.
T, B., D. DE VIENNE, F. HOSPITAL, L. MOREAU and C. DILLMANN, 2001 Genetic and nongenetic bases for the L-shaped distribution of quantitative trait loci effects. Genetics **157:** 1773–1787.
NK, A. (Editor), 1967 *Heritage From Mendel.* University of Wisconsin, Madison, WI.
'MAN, K. W., and J. L. WEBER, 1999 Long homozygous chromosomal segments in reference families from the centre d'Etude du polymorphisme humain. Am. J. Hum. Genet. **65:** 1493–1500.
'ALLI-SFORZA, L. L., and W. F. BODMER, 1971 *The Genetics of Human Populations.* W. H. Freeman, San Francisco.
.RLESWORTH, B., and D. CHARLESWORTH, 1999 The genetic basis of inbreeding depression. Genet. Res. **74:** 329–340.
.RLESWORTH, B., and K. A. HUGHES, 1996 Age-specific inbreeding depression and components of genetic variance in relation to the evolution of senescence. Proc. Natl. Acad. Sci. USA **93:** 6140–6145.
.RLESWORTH, B., and K. A. HUGHES, 1999 The maintenance of genetic variation in life-history traits, pp. 369–391 in *Evolutionary Genetics: From Molecules to Morphology*, Vol. 1, edited by R. S. SINGH and C. B. KRIMBAS. Cambridge University Press, Cambridge, UK.
'W, J. F., 1980 The estimation of inbreeding from isonymy. Hum. Biol. **52:** 1–14.
:ONER, D. S., 1964 *Introduction to Quantitative Genetics.* Oliver & Boyd, Edinburgh.
:ONER, D. S., and T. F. C. MACKAY, 1996 *Introduction to Quantitative Genetics*, Ed. 4. Longman, Harlow, UK.
'HER, J., B. BOUADJAR, R. HEILIG, M. HUBER, C. LEFEVRE *et al.*, 2001 Mutations in the gene encoding SLURP-1 in Mal de Meleda. Hum. Mol. Genet. **10:** 875–880.
'T, J., and R. MOTT, 2001 Finding the molecular basis of quantitative traits: successes and pitfalls. Nat. Rev. Genet. **2:** 437–445.
BERSTEIN, R. A., 1999 Blood pressure in the Caribbean. Hum. Biol. **71:** 659–684.
USHKA, M. K., J. B. FAN, K. BENTLEY, L. HSIE, N. SHEN *et al.*, 1999 Patterns of single-nucleotide polymorphisms in candidate genes for blood-pressure homeostasis. Nat. Genet. **22:** 239–247.
DY, G. H., 1960 *A Course of Pure Mathematics*, Ed. 10. Cambridge University Press, Cambridge, UK.
ES, B., and M. E. GODDARD, 2001 The distribution of the effects of genes affecting quantitative traits in livestock. Genet. Sel. Evol. **33:** 209–230.
.WICH, B. J., B. ROSNER, N. NUBANI, E. H. KASS and F. I. LEWITTER, 1982 Familial aggregation of blood pressure in a highly inbred community, Abu Ghosh, Israel. Am. J. Epidemiol. **115:** 646–656.
CIJEVIC, B., M. BAKRAN, S. S. PAPIHA, A. CHAVENTRE and D. F. ROBERTS, 1994 Serogenetic analysis in the study of the population structure of the eastern Adriatic (Croatia). Hum. Biol. **66:** 991–1003.
RIC, I. M., L. BARAC, D. BUKOVIC, I. FURAC, G. GEBER *et al.*, 2001a Short tandem repeat (STR) variation in eight village populations of the island of Korcula (Croatia). Ann. Hum. Biol. **28:** 281–294.
RIC, I. M., L. JIN, R. CHAKRABORTY, R. DEKA, L. BARAC *et al.*, 2001b Inter- and intra-Island genetic diversity in Adriatic populations of Croatia: implications for studying complex diseases in isolated populations. Am. J. Hum. Genet. **69** (Suppl.): 394.
AJTIC, B., M. DUJMOVIC, Z. KOLACIO and V. KOGOJ-BAKIC, 1995 Enclaves of hereditary dwarfism on the island of Krk, Croatia. Coll. Anthropol. **19:** 365–370.
.GER, H., 1968 Inbreeding effects on metrical traits in Northeastern Brazil. Am. J. Hum. Genet. **21:** 537–546.
ON, R. P., A. G. GHARAVI and D. S. GELLER, 2001 Molecular mechanisms of human hypertension. Cell **104:** 545–546.
KAY, T. F., 2001 The genetic architecture of quantitative traits. Annu. Rev. Genet. **35:** 303–339.
TIN, A. O., T. W. KURCZYNSKI and A. G. STEINBERG, 1973 Familial studies of medical and anthropometric variables in a human isolate. Am. J. Hum. Genet. **25:** 581–593.
TINOVIC, I., S. MASTANA, B. JANICIJEVIC, V. JOVANOVIC, S. S. PAPIHA *et al.*, 1998 VNTR DNA variation in the population of the island of Hvar, Croatia. Ann. Hum. Biol. **25:** 489–499.
TINOVIC, I., L. BARAC, I. FURAC, B. JANICIJEVIC, M. KUBAT *et al.*, 1999 STR polymorphisms in the population of the island of Hvar. Hum. Biol. **71:** 341–352.
MIALL, W. E., and P. D. OLDHAM, 1963 The hereditary factor in arterial blood pressure. Brit. Med. J. **19:** 75–80.
MUKAI, T., R. A. CARDELLINO, T. K. WATANABE and J. F. CROW, 1974 The genetic variance for viability and its components in a local population of *Drosophila melanogaster.* Genetics **78:** 1195–1208.
REICH, D. E., M. CARGILL, S. BOLK, J. IRELAND, P. C. SABETI *et al.*, 2001 Linkage disequilibrium in the human genome. Nature **411:** 199–204.
ROGULJIC, D., I. RUDAN and P. RUDAN, 1997 Estimation of inbreeding, kinship, genetic distances and population structure from surnames: example from the island of Hvar, Croatia. Am. J. Hum. Biol. **9:** 595–608.
ROBERTS, D. F., Z. M. NOOR, S. S. PAPIHA and P. RUDAN, 1992 Genetic variation in Brac, Croatia. Ann. Hum. Biol. **19:** 539–557.
RUDAN, I., H. CAMPBELL and P. RUDAN, 1999 Genetic epidemiological studies of eastern Adriatic island isolates, Croatia: objectives and strategies. Coll. Anthropol. **23:** 531–546.
RUDAN, P., D. SIMIC, N. SMOLEJ-NARANCIC, L. A. BENNETT, B. JANICIJEVIC *et al.*, 1987 Isolation by distance in Middle Dalmatia, Yugoslavia. Am. J. Phys. Anthropol. **74:** 417–426.
RUDAN, P., A. SUJOLDZIC, D. SIMIC, L. A. BENNETT and D. F. ROBERTS, 1992 Population structure in the eastern Adriatic: the influence of historical processes, migration patterns, isolation and ecological pressures, and their interaction, pp. 204–218 in *Isolation, Migration and Health*, edited by D. F. ROBERTS, N. FUJIKI and K. TORIZUKA. Cambridge University Press, Cambridge, UK.
SHRIMPTON, A. E., and A. ROBERTSON, 1988 The isolation of polygenic factors controlling bristle score in Drosophila. II. Distribution of third chromsome bristle effects with chromosome sections. Genetics **118:** 445–459.
STOLL, M., A. E. KWITEK-BLACK, A. W. COWLEY, JR., E. L. HARRIS, S. B. HARRAP *et al.*, 2000 New target regions for human hypertension via comparative genomics. Genome Res. **10:** 473–482.
TAMBS, K., T. MOUM, L. J. EAVES, M. C. NEALE, K. MIDTHJELL *et al.*, 1992 Genetic and environmental contributions to the variance of body height in a sample of first and second degree relatives. Am. J. Phys. Anthropol. **88:** 285–294.
TANKSLEY, S. K., 1993 Mapping polygenes. Annu. Rev. Genet. **27:** 205–233.
TAY, J. S., and W. C. YIP, 1984 The estimation of inbreeding from isonymy: relationship to the average inbreeding coefficient. Ann. Hum. Genet. **48:** 185–194.
TERWILLIGER, J., and H. H. H. GORING, 2000 Gene mapping in the 20th and 21st centuries: statistical methods, data analysis, and experimental design. Hum. Biol. **72:** 63–132.
THOMAS, J. D., M. M. DOUCETTE, D. C. THOMAS and J. D. STOECKLE, 1987 Disease, lifestyle and consanguinity in 58 American gypsies. Lancet **2:** 377–379.
TOLK, H. V., L. BARAC, M. PERICIC, I. M. KLARIC, B. JANICIJEVIC *et al.*, 2001 The evidence of mtDNA haplogroup F in a European population and its ethnohistoric implications. Eur. J. Hum. Genet. **9:** 717–723.
WADDLE, D. M., R. SOKAL and P. RUDAN, 1998 Factors affecting population variation in Eastern Adriatic isolates, Croatia. Hum. Biol. **70:** 845–864.
WAHID SAEED, A. A., F. J. AL SHAMMARY, T. A. KHOJA, T. J. HASHIM, C. C. ANOKUTE *et al.*, 1996 Prevalence of hypertension and sociodemographic characteristics of adult hypertensives in Riyadh City, Saudi Arabia. J. Hum. Hypertens. **10:** 583–587.
WEINER, J. S., and J. A. LOURIE, 1969 *Human Biology—A Guide to Field Methods.* Blackwell, Oxford.
WRIGHT, S., 1922 Coefficients of inbreeding and relationship. Am. Nat. **56:** 330–338.
ZENG, Z., 1992 Correcting the bias of Wright's estimates of the number of genes affecting a quantitative character: a further improved method. Genetics **131:** 986–1001.

Communicating editor: D. CHARLESWORTH

APPENDIX: THE EFFECT OF INBREEDING ON A MULTILOCUS PHENOTYPE

Model, notation, and assumptions: The phenotype of the ith individual is modeled by

TABLE A1

Computation of the components of genetic variance at locus *j*

Genotype	$A_{jk}A_{jk}$	$A_{jk}A_{jk'}$ $(k \neq k')$
Frequency (HWE)	p_{jk}^2	$2p_{jk}p_{jk'}$
Assigned value	$2a_{jk}$	$a_{jk} + a_{jk'} + d_{jkk'}$
Genetic value	$2a_{jk} - \mu_j$	$a_{jk} + a_{jk'} + d_{jkk'} - \mu_j$
Breeding value (additive)	$2(a_{jk} + T_{jk} + S_j - \mu_j)$	$a_{jk} + a_{jk'} + T_{jk} + T_{jk'} + 2S_j - 2\mu_j$
Dominance deviation	$D_j - 2T_{jk}$	$D_j - T_{jk} - T_{jk'} + d_{jkk'}$

$$y_i = \mu + x_{i1} + x_{i2} + \ldots + x_{in} + \varepsilon_i, \qquad \text{(A1)}$$

x_{ij} denotes the contribution to the phenotype of enotype at the jth locus ($j = 1, \ldots, n$), and ε_i is dom "environmental" contribution, uncorrelated en individuals and with mean 0 and variance σ^2. locus is assumed to have K_j alleles, A_{jk}, with frequen- $_{jk}$ ($k = 1, \ldots, K_j$), and loci are assumed to act vely and independently. Individuals with geno- $A_{jk}A_{jk}$ and $A_{jk}A_{jk'}$ ($k \neq k'$) are distributed around phenotypic values of $2a_{jk}$ and $a_{jk} + a_{jk'} + d_{jkk'}$, ctively. Thus, $d_{jkk'} = 0$ represents additivity, and $a_{jk} - a_{jk'}$ complete dominance of A_{jk} over $A_{jk'}$.

e effect of inbreeding: Assuming Hardy-Weinberg brium (HWE),

$$E(x_{ij}|\text{HWE}) = 2S_j + D_j, \qquad \text{(A2)}$$

$$S_j = \sum_k a_{jk}p_{jk}, \qquad \text{(A3)}$$

$$D_j = \sum_k \sum_{k'} p_{jk}p_{jk'}d_{jkk'}, \qquad \text{(A4)}$$

Σ_k denotes summation from $k = 1, \ldots, K_j$, and se for $\Sigma_{k'}$. (Note that for mathematical conformity, sume $d_{jkk} = 0, \forall\ k$.) If the two alleles at locus j are hen

$$E(x_{ij}|\text{IBD}) = 2S_j. \qquad \text{(A5)}$$

e, if the level of inbreeding, $P(\text{IBD})$, of the ith dual is F_i, we have

$$E(y_i) = \mu + 2\sum_j S_j + \sum_j D_j - \left\{\sum_j D_j\right\}F_i, \qquad \text{(A6)}$$

Σ_j denotes summation from $j = 1, \ldots, n$. Thus, of y_i against F_i is linear with slope $-\Sigma_j D_j$.

e components of genetic variance: Table A1 shows teps needed to compute the additive (V_{Aj}) and nance (V_{Dj}) components of total genetic variance at locus j, defining

$$\mu_j = 2S_j + D_j \qquad \text{(A7)}$$

$$T_{jk} = \sum_{k'} p_{jk'}d_{jkk'}. \qquad \text{(A8)}$$

s a generalization of Falconer's treatment for the biallelic case (FALCONER 1964). By adding the squared deviations weighted by their frequencies and simplifying, we find

$$V_{Aj} = 2\left\{\sum_k p_{jk}(a_{jk} + T_{jk})^2 - (S_j + D_j)^2\right\} \qquad \text{(A9)}$$

and

$$V_{Dj} = D_j^2 - 2\sum_k p_{jk}T_{jk}^2 + E_j, \qquad \text{(A10)}$$

where

$$E_j = \sum_k \sum_{k'} p_{jk}p_{jk'}(d_{jkk'})^2. \qquad \text{(A11)}$$

Since loci are assumed to be independent the overall components of variance are derived by summing over all j. Hence,

$$\text{Var}(x_{ij}) = V_{Gj} + (W_{Gj} - V_{Gj} + D_j^2)F_i - D_j^2F_i^2, \qquad \text{(A12)}$$

where

$$W_{Gj} = \text{Var}(x_{ij}|\text{IBD}) = 4\left\{\sum_k a_{jk}^2 p_{jk} - S_j^2\right\}. \qquad \text{(A13)}$$

Lower limit for the number of loci, *n*: We make use of the mathematical result that, for any two sets of real numbers $\{z_i,\ i = 1, \ldots, n\}$ and $\{w_i,\ i = 1, \ldots, n\}$, if $z_i^2 \leq w_i$ $(i = 1, \ldots, n)$ then

$$\frac{(z_1 + \ldots + z_n)^2}{w_1 + \ldots + w_n} \leq n. \qquad \text{(A14)}$$

This is an application of Cauchy's inequality (see, *e.g.*, HARDY 1960). In the present context, if we set $z_j = D_j$ and $w_j = V_j$ (where V_j can denote any component of variance, V_{Aj}, V_{Dj}, or V_{Gj}, as required), and provided that we can show that $D_j^2 \leq V_j$ for all j, then we have a sufficient, but not necessary, condition that the quantity

$$n_L = \frac{D_T^2}{V_T} \qquad \text{(A15)}$$

is a lower bound for n. Here, $D_T = -\Sigma_j D_j$ and $V_T = \Sigma_j V_j$ denote the overall slope and variance, respectively, the additivity of both relationships being a consequence of assuming that different loci act independently.

Special cases: The condition $D_j^2 \leq V_j$ does not hold in all circumstances. However, consideration of special cases suggests that the circumstances under which it

eaks down are rather exceptional. In the biallelic case odel BA), the condition always holds since D_j^2 is iden-al to V_{Dj}. For the multiallelic case we consider two odels (MA1 and MA2), in both of which all alleles at ery locus have equal frequency and successive homo-gotes are evenly spaced—that is, $p_{jk} = 1/K_j$ and $a_{jk} = k - (K_j + 1)/2]$ $(k = 1, \ldots, K_j; j = 1, \ldots, n)$. In odel MA1, the dominance effects are assumed to be ual in absolute magnitude, *i.e.*, $d_{jkk'} = a_j d_j$ $(\forall\ j$ and $k \neq k')$, whereas in model MA2 they are assumed to proportional to the interhomozygote distances; *i.e.*, $= a_j\rho_j|k - k'|$ $(\forall\ j, k, k')$. With this definition, $|\rho_j| =$ corresponds to full dominance of one allele in each ssible pair, and $|\rho_j| = 0$ to complete absence of domi-nce. By analogy, it seems logical in model MA1 to le $d_{jkk'}$ by half the mean interhomozygote distance, $a_j(K_j + 1)/3$.

By straightforward though tedious algebra it can be own that the condition $D_j^2 \leq V_{Gj}$ is satisfied under all odels unless the level of dominance exceeds $\sqrt{(3/2)}$ odel MA1) or $\sqrt{3}$ (model MA2). These asymptotic its are bounded from above as $K_j \to \infty$. Since such els imply a considerable degree of overdominance the same direction and between all pairs of alleles, s unlikely to apply in most situations. For example, sickle-cell anemia, in perhaps the most widely quoted d extreme case of overdominance in human genetics, LCONER (1964) quotes relative fitness values of 0.80, 0, and 0.25 for the "normal" homozygote, the hetero-ote, and the sickling-trait homozygote, respectively. e implied level of overdominance (ρ) is then 1.73.

Extreme overdominance: In the most extreme form overdominance, all homozygotes have one assigned ue (0, say), and all heterozygotes have another (d,). Then at a single locus, and dropping the suffix j, have

$$D = -d(1 - R_2) \tag{A16}$$

$$V_A = 2d^2(R_3 - R_2^2) \tag{A17}$$

$$V_D = d^2(R_2 - 2R_3 + R_2^2) \tag{A18}$$

d

$$V_G = V_A + V_D = d^2R_2(1 - R_2), \tag{A19}$$

ere

$$R_i = \textstyle\sum_k p_k^i. \tag{A20}$$

Hence,

$$\frac{D^2}{V_G} = \frac{1 - R_2}{R_2} \leq K - 1 \tag{A21}$$

since $R_2 \geq K^{-1}$. On summing over all loci and applying Cauchy's inequality, we obtain

$$\frac{D_T^2}{V_{GT}} \leq \sum_j (K_j - 1) = \text{total number of alleles} - \text{total number of loci}. \tag{A22}$$

Note that this depends in turn on the easily proved result that, for $A_j > 0$,

$$-A_j \leq B_j \leq A_j \quad \forall\ j \Rightarrow \left(\sum_j B_j\right)^2 \leq \left(\sum_j A_j\right)^2. \tag{A23}$$

Conclusions: The models explored here suggest that a sufficient condition for n_L (with $V_T = V_{GT} = \Sigma_j V_{Gj}$) to be a lower bound for n is likely to be satisfied in most practical circumstances and will fail only in situations of extreme overdominance. In the most extreme such situation, when all homozygotes have the same genetic value and all heterozygotes have a different one, Cauchy's inequality leads to the result that

$$\text{Total number of alleles}\left(\sum_j K_j\right) - \text{total number of loci } (n) \geq n_L. \tag{A24}$$

and hence, if $\bar{K}$ denotes the mean number of alleles per locus, that

$$n \geq \frac{n_L}{K - 1}. \tag{A25}$$

On the other hand, because the condition is *sufficient* but not *necessary* it will in practice be more widely applicable than the above models suggest. For example, in the multiallelic models, the requirement that the absolute dominance is less than a certain limit may allow dominance to be much greater for some pairs of alleles than for others and even of opposite sign, so long as the *average* dominance remains within the required limit.

Finally, it should be borne in mind that the present method reveals nothing about the relative *magnitude* of the dominance effects at different loci or of course about the presence of additive effects.

LETTER TO JMG

nbreeding and risk of late onset complex disease

Rudan, D Rudan, H Campbell, A Carothers, A Wright, N Smolej-Narancic, B Janicijevic, L Jin, Chakraborty, R Deka, P Rudan

J Med Genet 2003;40:925–932

nbreeding has been shown in almost all species to be associated with impairment of function because of homozygosity of recessive alleles. This occurs across a wide range traits and suggests a large number of deleterious alleles in e human genome. This has been predicted from the duced early survival of offspring in first cousin marriages nd from similar results in other organisms.[1–3] As most entified genetic variants causing complex disease in umans are partially recessive[1] we predict that inbreeding humans might influence a wide range of complex diseases. Numerous reports on the health effects of inbreeding have cused mainly on its impact on reproduction, childhood ortality, and rare Mendelian disorders.[2 3] For example, a 4–% increase in childhood mortality has been found in the ffspring of first cousin marriages, and similar results have een reported in other species.[2 4 5] However, the effects of breeding on late onset disorders are largely unknown, espite the fact that deleterious effects of inbreeding in other pecies are known to increase with age, as predicted by lection theory.[6 7] The reported finding of greater inbreeding ffects for traits such as blood pressure and serum cholesterol middle age compared with early adult life is consistent ith this.[8]

In order to investigate the hypothesis that the heritable mponent of late onset diseases includes a major class of eleterious recessive alleles,[9] we recently studied the effects inbreeding on blood pressure among 2760 adult individuals from 25 villages in a Dalmatian island isolate. The udy showed a large effect of inbreeding on blood pressure uivalent to a rise in systolic blood pressure of ~20 mm Hg nd diastolic blood pressure of ~12 mm Hg in offspring of rst cousin marriages. The effect appeared to be mediated by veral hundred recessive alleles as a result of increased mozygosity.[10] In support of this finding, several studies of nall inbred communities worldwide have reported an creased prevalence of hypertension.[8 11–15]

In the present study, we extend this observation by vestigating the relation between inbreeding and the evalence of 10 late onset complex diseases of public health nportance. The study was carried out in 14 of the original 25 olate villages on three neighbouring islands in middle almatia, Croatia. These island populations present a wide nge of levels of inbreeding and endogamy, reduced genetic riation at both individual and (sub)population levels, and lative uniformity of environment,[10 16–18] and thus provide a od setting for investigating inbreeding effects.

Key points

- From arguments derived from population genetics, we propose that the genetic basis of common late onset diseases includes a major class of deleterious recessive alleles. Inbreeding is therefore predicted to increase the incidence of such diseases.
- Among 10 late onset conditions studied in a genetic isolate population, inbreeding was found to be a significant (positive) predictor for coronary heart disease, stroke, cancer, uni/bipolar depression, asthma, gout, and peptic ulcer, but not for type 2 diabetes
- It appears that inbreeding causes an increase in homozygosity at many genetic loci with small deleterious effects on homeostatic pathways, resulting in increased disease risk.
- The results indicate that between 23% and 48% of the incidence of these disorders in this population sample (other than type 2 diabetes) could be attributed to recent inbreeding. The global impact of inbreeding could thus be substantial, as an estimated one billion people globally show rates of consanguineous marriages greater than 20%.

ETHODS

udy cohort

e village populations of three neighbouring islands in the stern Adriatic, part of Middle Dalmatia, Croatia (Brac, var, and Korcula, see fig 1), represent well characterised netic isolates. The tendency towards inbreeding in each llage has been influenced by geographic isolation, political *Pastrovic"*) privileges given to residents of certain communities, and sociocultural factors.[16–18] A survey of 1339 adult individuals selected randomly from voting lists to form approximately 20–30% of the total population of these 14 villages was undertaken in late 1970s and early 1980s by the Institute for Anthropological Research in Zagreb in collaboration with the Smithsonian Institute, Washington, USA. The mean adult ages in individual villages varied from 41 to 65 years (detailed age/sex profiles for each village are given in appendix 1). For each individual, information was collected on the highest level of education, occupation, diet, smoking habits, and body mass index (table 1).

Computation of individual inbreeding coefficients

A single researcher (IR) computed individual inbreeding coefficients for each study participant, based on pedigree information on four to five ancestral generations, recorded during the initial field work and supplemented by a study of parish registries. The individual inbreeding coefficients (F) were then computed according to Wright's path method[19]:

$$F = \sum_{(1\to c)} \left(\frac{1}{2}\right)^{(n_i+m_i+1)}$$

where m_i and n_i refer to the number of paths from a common ancestor, and c refers to the number of common ancestors. The genealogical inbreeding coefficient for each village was then computed as the average of all individual F values. To further support these estimates, F was calculated from isonymy as suggested by Tay and Yip,[20] and mean values

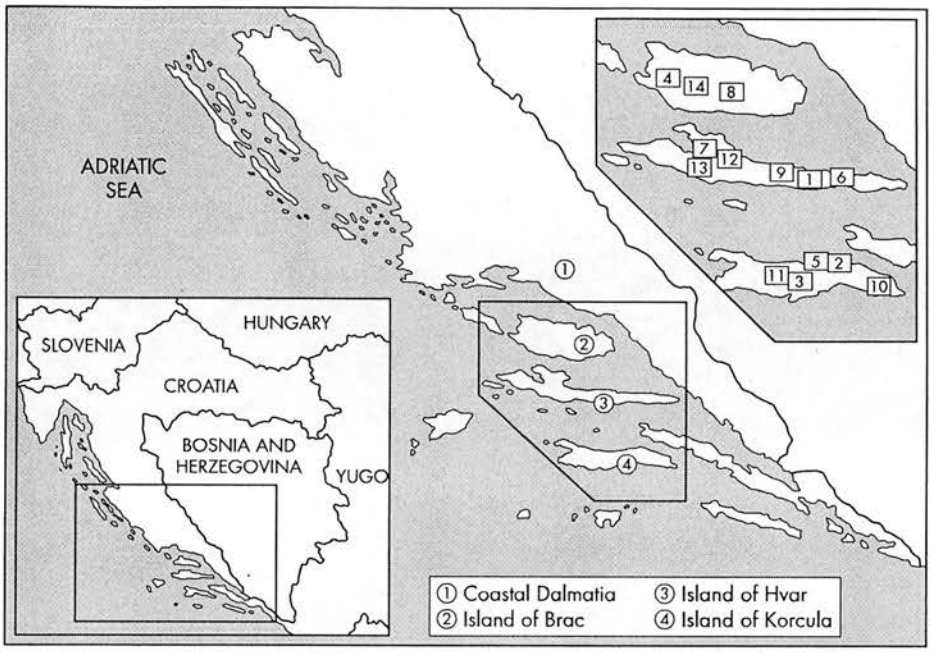


Figure 1 Map of the Dalmatian island genetic isolate showing study islands and villages (numbered 1 to 14).

were derived for each of the 14 villages. Estimates based on isonymy are generally thought to be positively biased, and so to provide an upper bound for F (table 1).

Follow up data on disease status of cohort individuals

Population census data in 1981, 1991, and 2001 from study villages show significant depopulation with minimal immigration over the last two decades. Thus only 480 individuals who were still resident in the study villages were available for follow up in the year 2000. Specific diagnostic criteria were established for each of 10 commonly occurring disorders in this population (coronary heart disease, stroke, cancer, schizophrenia, epilepsy, uni/bipolar depression, asthma, adult type diabetes, gout, and peptic ulcer) following those presented in *Merck's Manual*. In collaboration with local general practitioners, two medical doctors, who were blind to the inbreeding status of each individual, inspected the medical records between March and October 2000 and recorded whenever appropriate diagnostic criteria were met. Diagnoses were supported wherever possible by medical records from the local general hospital in Split.

Statistical analysis and modelling

Disease prevalence was first investigated by comparisons between villages grouped by the level of inbreeding as high, moderate, or low (table 2). Disease prevalence rates were

Table 1 Genetic and environmental characteristics of 14 village populations ranked according to average inbreeding coefficient computed from genealogical data

Village (island)	N(sam)	N(pop)	F(gen)	F(iso)	Edu (%)	Occu (%)	Diet (%)	Smo (%)	BMI (avg)
Gdinj (H)	135	153	0·049	0·107	2·3	80·8	93·8	4·6	25·1
Pupnat (K)	96	326	0·044	0·034	2·1	89·7	91·8	9·3	26·7
Cara (K)	139	464	0·032	0·040	1·5	89·0	94·9	22·6	26·4
Dracevica (B)	20	66	0·031	0·031	2·4	85·7	90·5	23·8	28·6
Racisce (K)	103	290	0·027	0·034	1·9	87·5	83·6	23·1	27·3
Bogomolje (H)	90	102	0·025	0·030	2·5	84·0	93·8	4·9	23·0
Vrisnik (H)	105	216	0·015	0·023	2·1	82·3	91·6	13·5	25·5
G Humac (B)	43	214	0·013	0·016	2·5	81·3	87·5	38·8	27·3
Zastrazisce (H)	133	210	0·013	0·013	4·0	86·3	93·6	5·9	25·1
Lumbarda (K)	59	354	0·012	0·025	4·7	85·2	89·8	21·3	26·2
Smokvica (K)	97	866	0·008	0·019	3·1	81·4	81·4	27·8	26·0
Svirce (H)	152	375	0·008	0·008	2·0	80·1	83·2	12·8	25·7
Dol (H)	102	154	0·005	0·004	4·2	75·8	86·3	8·4	26·1
Nerezisca (B)	65	106	0·002	0·004	1·8	60·7	79·5	34·8	28·0

Islands: H, Hvar; B, Brac; K, Korcula.
BMI (avg), average body mass index among examinees; Diet, proportion of examinees regularly consuming traditional Mediterranean diet; Edu, proportion of examinees with some college education; F(gen), average inbreeding coefficients computed from genealogical data; F(iso), average inbreeding coefficients computed from isonymy; N(pop), number of individuals in the village population in year 2000 (year in which disease prevalence was determined); N(sam), number of individuals in the sample (represents 20–30% random samples of total village population obtained between 1979 and 1981); Occu, proportion of examinees working in agriculture and fishery; Smo: proportion of examinees smoking (smoking recorded using Brinkman index*).

Brinkman GL, Coates EO. The effect of bronchitis, smoking and occupation on ventilation. *Am Rev Respir Dis* 1963;**87**:384–9.

Table 2 Age and sex standardised prevalences of 10 complex diseases in groups of villages with relatively "high," "moderate," and "low" average inbreeding coefficient values

	"High inbreeding" (N=1401)	"Moderate inbreeding" (N=998)	"Low inbreeding" (N=1501)
	Average F_{gen}=0.036 Average F_{iso}=0.052	Average F_{gen}=0.013 Average F_{iso}=0.019	Average F_{gen}=0.006 Average F_{iso}=0.009
Coronary heart disease	13.28%	11.95%	11.23%
Stroke	2.43%	2.79%	1.73%
Cancer	4.54%***	3.44%*	1.93%
Schizophrenia	1.23%***	0.96%*	0.14%
Uni/bipolar depression	10.26%***	7.63%**	4.51%
Asthma	3.63%	2.64%	2.60%
Type II diabetes	6.02%	7.35%	6.77%
Gout	9.25%***	7.19%***	3.96%
Peptic ulcer	6.92%***	4.29%**	2.18%
Epilepsy	1.47%***	0.78%	0.31%

Statistical significance (p values) in highly and moderately inbred groups is calculated against the low inbreeding group: *p<0.05; **p<0.01; ***p<0.001.
F_{gen}, weighed average inbreeding coefficient computed from genealogical data; F_{iso}, weighed average inbreeding coefficient computed from isonymy data.

andardised by sex and age to the total population of all 14 llages included in the study, using 10 year age intervals and rect standardisation.

In an attempt to overcome the possible confounding effects unmeasured environmental exposures or population ratification, the relation between *individual* inbreeding efficients and disease outcomes was investigated among e 480 individuals. Data on age, sex, education, occupation, et, smoking status, village of residence, height, weight, and dividual inbreeding coefficient (F) were analysed in a gistic regression with disease status as the outcome. Age d sex were forced into the prediction model irrespective of nether they were formally significant. Other main effects nbreeding, smoking, height, weight and village) were tered with p = 0.05, and all higher order effects and teractions with p = 0.01.

opulation attributable risk

opulation attributable risk (PAR) estimates for inbreeding ere calculated by logistic regression, allowing for individual fferences in the variables village, sex, age, height, weight, d smoking. The appropriate regression was determined as a nction of all associated variables (including F), then the obability for each individual of having the disease outcome as calculated assuming an F value set at 0. The sum of all ch probabilities, P_{sum}, is an estimate of the number fected in the absence of inbreeding, but with other riables remaining unaltered. Then $PAR = 1 - P_{sum}/N_{aff}$, here N_{aff} is the actual number of affected individuals. In riving the PAR values, the effects estimated from the bset of 14 villages were applied to the full dataset from all llages.

The original survey was carried out with the informed nsent of the participants and ethical approval for the recent eld work and analyses was granted by the appropriate search ethics committees in Scotland and Croatia.

ESULTS

ble 1 presents selected characteristics of the study villages: mber and name of village, island, average inbreeding efficients computed from genealogical data and from onymy, education level, occupation, diet, smoking, and dy mass index. The table presents village data in three oups defined by the average level of inbreeding. The ocation of villages to these three groups according to the timates of F based on genealogy (F_{gen}) is broadly consistent with those based on isonymy (F_{iso}). On a log–log scale, the correlation between the two measures of inbreeding across villages was 0.92, with F_{iso} exceeding F_{gen} on average by a factor of 1.32.

Table 2 presents age and sex standardised disease prevalence data for each of the 10 diseases under investigation. A stepwise increase in disease prevalence across villages stratified by (increasing) average inbreeding coefficient was found for gout, depression, peptic ulcer, schizophrenia, cancer, epilepsy, coronary heart disease, and asthma (the last two not statistically significant).

Table 3 includes data from 480 individuals from 14 villages, with age and sex forced into the multiple logistic regression model. Other main effects (inbreeding, smoking, log_weight, log_height, and village) were entered with p = 0.05, and all higher order effects and interactions with p = 0.01. Schizophrenia and epilepsy were excluded from this analysis because of the small number of cases (four each) and thus low study power to investigate predictors for these conditions. Inbreeding remained a significant (positive) predictor for every condition except type 2 diabetes. The forced inclusion of age and sex in the model acted to reduce slightly the significance of the effect of inbreeding, because of a small positive correlation between inbreeding and age. Village of residence was found to be a significant predictor only for coronary heart disease. Weight was a significant positive predictor for type 2 diabetes and gout and a significant negative predictor for cancer.

In terms of health impact, the results on population attributable risk (table 3) show that 23–48% of the incidence of these disorders (other than type 2 diabetes) in this population can be attributed to recent inbreeding.

DISCUSSION

The impact of inbreeding on reproduction, childhood mortality, and Mendelian disorders is well documented.[2 3] In contrast, very little has been published on the effects of inbreeding on late onset diseases. This is despite the fact that inbreeding may have a greater influence on late onset traits than on traits that are subject to early selection.[6 7] This study shows an important effect of inbreeding on several genetically complex late onset diseases which are of major public health importance. This is consistent with our proposal that an important genetic influence on these disorders is mediated by numerous deleterious recessive alleles, suggesting that

Table 3 Summary of results of multiple logistic regression

Disease	Predictor	Coefficient	SE	p Value	PAR (SE) (%)†
CHD	Constant	−6.20	1.27	0.000	34.3 (6.6)*
	SEX	0.47	0.26	0.070	
	AGE·1	0.69	0.14	0.000	
	VILL	–	–	0.022	
	F·1	26.63	4.88	0.000	
Stroke	Constant	−3.28	0.71	0.000	26.3 (12.3)
	SEX	−0.30	0.40	0.464	
	AGE·1	0.34	0.20	0.095	
	F·1	13.37	5.85	0.022	
Cancer	Constant	10.17	6.27	0.105	23.2 (11.7)
	SEX	0.83	0.54	0.128	
	AGE·1	0.02	0.20	0.909	
	LN_WT	−3.49	1.40	0.013	
	F·1	13.0	6.17	0.035	
Depression	Constant	−3.86	0.58	0.000	42.2 (7.6)
	SEX	0.69	0.31	0.028	
	AGE·1	0.02	0.14	0.896	
	F·1	24.54	4.43	0.000	
Asthma	Constant	−3.74	0.80	0.000	47.6 (12.6)
	SEX	−0.40	0.45	0.374	
	AGE·1	0.36	0.23	0.110	
	F·1	23.10	6.04	0.000	
Diabetes (type II)	Constant	−28.16	4.58	0.000	0
	SEX	0.97	0.31	0.002	
	AGE·1	0.31	0.14	0.027	
	LN_WT	5.60	0.99	0.000	
Gout	Constant	−14.90	4.08	0.000	31.0 (7.6)
	SEX	−0.17	0.29	0.552	
	AGE·1	0.38	0.14	0.006	
	LN_WT	2.85	0.89	0.001	
	F·1	17.69	4.16	0.000	
Peptic ulcer	Constant	−3.23	0.63	0.000	30.6 (10.5)
	SEX	−0.18	0.35	0.615	
	AGE·1	0.36	0.18	0.043	
	F·1	15.68	5.06	0.002	

*Excludes village effects, even though they were significant in the dataset from 14 villages, as they could not be estimated for the remaining 11 villages in the full set.
†Population attributable risk estimated from individual inbreeding levels (see text).
SEs calculated by "delta" method.
CHD, coronary heart disease; F·1, inbreeding; LN_WT, weight; PAR, population attributable risk; VILL, village.

ıbreeding increases disease risk as a result of increased omozygosity.[9]

alidity of findings

is important to consider whether these results can be xplained by chance, bias, or confounding.

hance

he fact that this was our major a priori hypothesis, taken ogether with the levels of statistical significance reported, rgues strongly against chance as an explanation for these ndings.

ias

Vith respect to selection bias, the 480 individuals on whom e were able to obtain disease outcome data were a subset of ıe original cohort from 1979–81. However, census data evealed that the major reason for loss to follow up, other ıan deaths of cohort members, was emigration from the llages over the 20 year period, which should not result in ıbstantial bias.

With respect to measurement bias, we do not believe that isease outcomes were ascertained or recorded differently mong individuals who differed by inbreeding status. tandard and explicit clinical criteria were adopted by the vo study doctors, who were blind to the inbreeding status of ıdividuals. Furthermore, the results cannot be explained by ifferent diagnostic practices in different villages, as the llage term was not found to be statistically significant in the multiple logistic regression analysis (except for coronary heart disease).

Confounding

Various potential confounding factors (age, sex, smoking status, education level, general diet, occupational group, height, weight) were measured and their effects adjusted for in the multivariate analysis. Although a degree of imprecision is inevitable in measuring some of these factors, we do not believe that confounding could have accounted for the large and consistent effects demonstrated.

Factors supporting the validity of the data

Several factors support the validity of the data. First, the findings support our prior hypothesis and are consistent with similar findings on hypertension in the same population[10] and with other reports of inbreeding effects on blood pressure.[8 11–15] Second, the overall strength of the effect argues against bias or confounding. Third, we have presented detailed arguments that biologically plausible mechanisms underpinning this effect are consistent with population genetic theory and observations in a wide range of organisms.[9] Finally, the data are consistent with the few other published reports of inbreeding effects on late onset diseases in which inbreeding was measured rather than self reported (table 4).

Size of inbreeding effect in late onset diseases

The magnitude of the inbreeding effect on disease prevalence was large. However, the effect on prevalence of stroke and

ronary heart disease, for example, is consistent with our evious report of a rise in diastolic blood pressure by nm Hg for an increase in F of 0.01,[10] and both cohort udies and randomised trials show that an increase of nm Hg diastolic blood pressure is associated with a 33% crease in stroke risk and a 20% increase in risk of ischaemic art disease.[21] The effect size is supported by the only two eviously published estimates of inbreeding on blood pressure at we could identify in other isolate populations.[8 11]

The large effect may reflect the greater influence of breeding on late onset traits than on traits that are subject early selection.[6 7] It is also possible that low environmental riation, or underestimation of F because of individuals ing related through multiple lines of descent, contribute to e size of inbreeding effect in these isolate popula-ns.[8 11 15 22] Thus the magnitude of the inbreeding effect ative to the overall variation may be smaller in more vironmentally diverse populations.

The ecological analysis at the village level (table 2) suggests an inbreeding effect on the prevalence of epilepsy and schizophrenia, but there were insufficient outcome events to permit an analysis at the individual level. The effect of inbreeding was shown in seven of the other eight diseases studied. The lack of observed effect of inbreeding on type 2 diabetes may reflect the lower heritability and stronger environmental influences involved in the aetiology of this condition[23] or heritable components that are mainly additive or dominant rather than recessive.

Mechanism of inbreeding effects in late onset diseases

We have argued that the genetic component of late onset diseases may be due principally to large numbers of rare variants in numerous genes—the common disease/rare variant (CD/RV) hypothesis.[9] The possibility that a significant fraction of the genetic variation in complex traits is caused by rare alleles maintained by mutation–selection balance is

Table 4 Review of the studies investigating the effect of inbreeding on complex late onset diseases

Disease	Author/reference	Study design	Reported effect of inbreeding
Coronary heart disease	Shami *et al. Lancet* 1991;**338**:954	Case–control study, hospital based	Hospital cases had significantly greater inbreeding coefficients than controls (individuals in population from which they were recruited)
	Puzyrev *et al. Arctic Med Res* 1992;**51**:136–42	Ecological study	Increased risk of myocardial ischaemia among endogamous males
	Jaber *et al.** *Am J Med Genet* 1997;**70**:346–8	Case–control study (self reported exposure and outcome status)	No difference in self reported prevalence among students with or without consanguineous parents
	Rudan *et al* (*present study*), 2003	Cohort study (with parallel ecological study)	Increased risk associated with greater inbreeding coefficient (relative risk of 1.2 per 3% inbreeding)
Cancer	Simpson *et al. Am J Obst Gynecol* 1981;**141**:629–36	Case–control study, population based	Increased risk of breast, endometrial, and ovarian cancer associated with greater inbreeding coefficient in women under 45 years
	Lebel & Gallagher. *Am J Med Genet* 1989;**33**:1–6	Case–control study, population based	Increased risk associated with greater inbreeding coefficient, (especially for multiple and early onset cancers)
	Shami *et al. Lancet* 1991;**338**:954	Case–control study, hospital based	Hospital cases had significantly greater inbreeding coefficients than controls (individuals in population from which they were recruited)
	Rudan. *Hum Biol* 1999;**71**:173–87	Ecological study	Stepwise increase of 20 year cancer incidence associated with greater inbreeding in five island communities
	Denic & Bener.**Br J Cancer* 2001;**85**:1675–9	Case–control study, population based (self reported exposure status)	Decreased risk of breast cancer among women who self reported being offspring from consanguineous unions, no effect on cervical cancer
	Rudan *et al* (*present study*), 2003	Cohort study (with parallel ecological study)	Increased risk associated with greater inbreeding coefficient (relative risk of 2 per 3% inbreeding)
Psychiatric disorders	Abaskuliev & Skoblo. *Genetika* 1975;**11**:145–8	Case–control study, population based	Increased frequency of consanguinity among parents of schizophrenia cases
	Sangstad & Odegard. *Clin Genet* 1986;**30**:261–75	Case–control study, population based	No increase in first cousin matings among parents of psychiatric patients (major changes in diagnostic criteria over time reported by authors as important confounder)
	Gindilis *et al. Genetika* 1989;**25**:734–43	Ecological study	Severe schizophrenia 2–3 times more prevalent in inbred communities
	Rudan *et al.* (*present study*) 2003	Cohort study (with parallel ecological study)	Increased risk associated with greater inbreeding coefficient (relative risk of 2.5 [depression] and 5 [schizophrenia] per 3% inbreeding)
Alzheimer's disease	Vezina *et al. Genet Epidemiol* 1999;**16**:412–25	Case–control study, hospital based	205 necropsy confirmed cases of late onset Alzheimer's disease significantly more inbred than controls
Multiple sclerosis	Roberts *et al. J Epidemiol Community Health* 1979;**33**:229–35	Case–control population based (with parallel ecological study)	Average inbreeding coefficient greater among cases than controls; increased prevalence in genetic isolate population)
	Roberts *et al. J Epidemiol Community Health* 1983;**37**:281–5	Case–control population based (with parallel ecological study)	Average inbreeding coefficient greater among cases than controls; increased prevalence in genetic isolate population)
Type 2 diabetes	Jaber *et al.** *Am J Med Genet* 1997;**70**:346–8	Case–control study (self reported exposure and outcome status)	No difference in self reported prevalence among students with or without consanguineous parents
	Rudan *et al.* (*present study*) 2003	Cohort study (with parallel ecological study)	No consistent increase in risk or prevalence associated with greater inbreeding coefficient
Gout	Ombra *et al. Am J Hum Genet* 2001;**68**:1119–29	Ecological study	Increased prevalence of hyperuricaemia and uric stones in a highly inbred community
	Rudan *et al.* (*present study*) 2003	Cohort study (with parallel ecological study)	Increased risk associated with greater inbreeding coefficient (relative risk of 2.5 per 3% inbreeding)
Asthma	Jaber *et al.** *Am J Med Genet* 1997;**70**:346–8	Case–control study (self reported exposure and outcome status)	No difference in self reported prevalence among students with or without consanguineous parents
	Rudan *et al.* (*present study*) 2003	Cohort study (with parallel ecological study)	Increased risk associated with greater inbreeding coefficient (relative risk of 1.5 per 3% inbreeding)
Peptic ulcer	Jaber *et al.** *Am J Med Genet* 1997;**70**:346–8	Case–control study (self reported exposure and outcome status)	No difference in self reported prevalence among students with or without consanguineous parents
	Rudan *et al.* (*present study*) 2003	Cohort study (with parallel ecological study)	Increased risk associated with greater inbreeding coefficient (relative risk of 3 per 3% inbreeding)

*These studies are based on self reported consanguinity which in our experience is unreliable as most "inbreeding loops" are found in the third and fourth parental generations and are thus unknown to most people.

nsistent with extensive research into the genetics of uantitative traits in simpler organisms.[7] Recent estimates[24] ggest that each person carries, on average, 500–1200 ghtly deleterious mutations, most of which are rare nd present in heterozygous form. Many of these variants ill become homozygous in inbred individuals, who are pected to show correspondingly large effects. We have eviously reported an estimate of several hundred recessive nes underlying human hypertension,[10] consistent with a mplex and genetically highly variable system of blood essure control and with published work from ontaneous and engineered animal models of hyperten-on.[25 26] The late onset disorders under investigation in this udy are likely to be similarly complex at a physiological vel so that significant effects of inbreeding are again pected.

The observed effect of inbreeding on the prevalence of veral different late onset diseases is consistent with the esence of many deleterious recessive alleles located roughout the genome. It is also consistent with a more neral effect of inbreeding, with increased homozygosity at ese loci leading to an accumulation of small deleterious fects on homeostatic pathways, which cumulatively crease disease risk. This suggests a greater sensitivity of meostatic mechanisms to inbreeding in later life, as edicted by findings in animals.[5 6] Decay of homeostatic pacity would also be expected to lead to reduced capacity to spond appropriately to diverse stimuli. This is supported by e recently reported observation that the reduced survival und in inbred animals is greater in the natural habitat than a controlled laboratory environment.[27]

The inbreeding data do not allow an easy distinction to be ade between the relative contributions of common versus re variants but do inform two somewhat neglected aspects the genetic architecture underlying complex diseases.[9] rst, the results provide indirect evidence in support of a ajor polygenic component to disease susceptibility. The breeding coefficient is shown to be a significant predictor of ronary heart disease, stroke, cancer, depression, asthma, out, and peptic ulcer, with population attributable risks arying between 23% ands 48% (table 3). Second, the cessive or partially recessive nature of complex disease sceptibility has received little emphasis. Both factors have plications for the identification of individual disease sceptibility alleles. If disease susceptibility is indeed highly lygenic then it implies the need to reduce the phenotypic mplexity of "disease" by means of genetically simpler but ntributory quantitative traits (QT) or disease subgroups. hose with extreme values of QT distributions or early sease age of onset will be those most likely to harbour sceptibility alleles of large effect and hence to provide a alistic possibility of gene identification. A significant mponent of genetic susceptibility appears to result from ariants that are recessive or partially recessive. This implies at the study of inbred populations would be advantageous, the increased gene dosage of such variants in inbred dividuals will tend to amplify their phenotypic effects mpared with outbred populations, where most alleles are resent in heterozygotes.

Public health implications

The population attributable risk estimates from this study suggest that 23–48% of the incidence of the disorders showing an inbreeding effect in this population can be attributed to inbreeding. We have previously reported that 36% of hypertension incidence in this population can be attributed to inbreeding.[10] These estimates are specific to this population and may be absent or considerably smaller in other populations. Nevertheless, inbreeding is highly prevalent globally and inbreeding effects may explain some of the observed differences in disease prevalence among different populations. Consanguineous marriages, defined as a union between individuals related as second cousins or closer (equivalent to $F \geq 0.0156$ in their progeny), have been conservatively estimated to occur at 1–10% prevalence among 2811 million people globally and at 20–50% prevalence among 911 million.[28 29] In addition, the extent of inbreeding even in outbred populations may have been underestimated.[30] Further details, including updated tables of global consanguinity studies, can be found at www.consang.net. The global impact of inbreeding on late onset disorders of public health importance could therefore be significant in health economic terms. This effect may be mediated by the observed inbreeding effect on important physiological traits such as blood pressure[10] and cholesterol,[8] recently shown to be two of the most important determinants of the global burden of disease.[31] As inbreeding declines owing to increased population movement and admixture, the prevalence of late onset disorders may also decline, and as common late onset diseases are correlated with longevity,[32] this may influence life expectancy.

ACKNOWLEDGEMENTS

The initial field work in the 14 villages was funded by grants from the Smithsonian Institution, Washington DC, USA. The reconstruction of genealogies was supported partly by the Ministry of Science and Technology of the Republic of Croatia. The classification of disease status was sponsored partly by the University of Cincinnati, USA (from grant ES06096 from the National Institutes of Health, USA) and the Ministry of Science and Technology of the Republic of Croatia. Subsequent analysis was supported by the Wellcome Trust (IRDA) grant to HC and IR, the Croatian Ministry of Science and Technology (CMST) grants to PR, NS-N, and IR. IR was supported by funds from the UK Medical Research Council, the University of Edinburgh, and the Overseas Research Scheme.

Authors' affiliations

I Rudan, School of Public Health "Andrija Stampar", University Medical School, Zagreb, Croatia
H Campbell, Public Health Sciences, University of Edinburgh, Edinburgh, UK
D Rudan, N Smolej-Narancic, B Janicijevic, P Rudan, Institute for Anthropological Research, Zagreb, Croatia
A Carothers, A Wright, Medical Research Council, Human Genetics Unit, Western General Hospital, Edinburgh, UK
L Jin, R Chakraborty, R Deka, Department of Environmental Health, University of Cincinnati, Cincinnati, Ohio, USA

Correspondence to: Dr Harry Campbell, Public Health Sciences, University of Edinburgh, Edinburgh EH8, 9AG, UK; Harry.Campbell@ed.ac.uk

APPENDIX 1
Age/sex profiles of the 14 study villages

Village (No)	Population	Male	Female	M/F ratio	Age (mean)	Age (SD)
Dracevica (4)	66	37	29	1.2759	52.800	22.890
Nerezisca (14)	106	56	50	1.1200	51.640	22.820
G. Humac (8)	214	97	117	0.8291	53.450	21.550
Dol (13)	154	75	79	0.9494	55.930	21.440
Svirce (12)	375	179	196	0.9133	47.720	22.420
Vrisnik (7)	216	112	104	1.0769	51.220	20.150
Zastrazisce (9)	210	96	114	0.8421	53.620	21.560
Gdinj (1)	153	76	77	0.9870	65.940	17.060
Bogomolje (6)	102	35	67	0.5224	60.880	22.370
Smokvica (11)	866	411	455	0.9033	42.777	24.061
Cara (3)	464	240	224	1.0714	43.380	22.830
Pupnat (2)	326	173	153	1.1307	45.840	23.830
Lumbarda (10)	354	169	185	0.9135	41.280	23.710
Racisce (5)	290	134	156	0.8590	43.310	27.030

No = the village number given in fig 1 and is equivalent to the rank according to average inbreeding coefficient given in table 1.

APPENDIX 2
Criteria adopted in this study for establishing diagnoses of 10 selected late onset diseases

1. **Coronary heart disease** (includes angina pectoris and myocardial infarction)

Angina pectoris

May be diagnosed by GP.

Clinical syndrome characterised by repeating episodes of precordial discomfort or pressure, typically precipitated by exertion or relieved by rest or sublingual glyceryl trinitrate with or without reversible ischaemic ECG changes.

Myocardial infarction

Must be diagnosed by a consultant in local general hospital.

Based on presenting symptoms (deep substernal radiating pain) and supported by combination of ECG findings (deep Q waves, elevated or depressed ST segments, deeply inverted T waves), and raised myocardial component of creatine kinase and lactic dehydrogenase, with or without myocardial imaging.

2. **Cerebral stroke**

Must be diagnosed by a consultant in local general hospital.

Based on presenting symptoms including variable neurological defects that increase over 24–48 hours.

Diagnosis may be supported by CT/MRI scan or arteriography.

3. **Cancer**

Must be diagnosed by a consultant in local general hospital.

Diagnosis requires abnormal cellular growth of any site to be histologically confirmed as malignant.

4. **Diabetes type II**

May be diagnosed by GP.

Symptomatic hyperglycaemia (polyuria, polydipsia, polyphagia, weight loss) or diabetic ketoacidosis or non-ketotic hyperglycaemic-hyperosmolar coma; or plasma (or serum) glucose level greater than 140 mg/dl after an overnight fast on two occasions; or development of any of the late complications (retinopathy, nephropathy, atherosclerotic changes on peripheral or coronary arteries, neuropathy).

5. **Schizophrenia**

- Must be diagnosed by a consultant in local general hospital.
- Chronic mental disorder (continuous signs of illness for at least six months) characterised by psychotic symptoms involving disturbances of thought, perception, feeling and behaviour
- Psychotic symptoms such as delusions, hallucinations, and formal thought disorder; deterioration from previous level of functioning; a tendency toward onset before the age of 45.
- Diagnosis should exclude mood (affective) disorder, organic mental disorder or mental retardation.

6. **Manic depression**

- Must be diagnosed by a consultant in local general hospital.
- Combination of symptomatic picture of depression, chronic course, family history, and response to treatment
- Diagnosis may be supported by TRH stimulation test or dexamethasone suppression test.

7. **Epilepsy**

- Must be diagnosed by a consultant in local general hospital.
- Recurrent paroxysmal disorder characterised by sudden brief attacks of altered consciousness, motor activity, sensory phenomena, or inappropriate behaviour caused by abnormal excessive discharge of cerebral neurones.
- Diagnosis may be supported by abnormalities in EEG, CT, or MRI.

8. **Asthma**

- May be diagnosed by GP.
- Airways obstruction that is usually reversible, presenting with attacks of tachypnoea, tachycardia, and audible

wheezes, airway inflammation, and increased airways responsiveness to a variety of stimuli.

Diagnosis may be supported by a family history of allergy or asthma.

9. **Gout**

May be diagnosed by GP.

Acute gouty arthritis (recurrent acute mono/polyarticular pain in peripheral joints, often nocturnal, progressively more severe, with swelling, warmth, redness, and tenderness).

Diagnosis may be supported by any of the following: raised serum urate (greater than 7 mg/dl), demonstration of urate crystals in tissue or synovial fluid, or dramatic response (within 24 hours) to colchicine.

10. **Peptic ulcer**

Must be diagnosed by a consultant in local general hospital.

Circumscribed ulceration of the gastric or duodenal mucous membrane causing a chronic and recurrent burning, gnawing, aching, soreness, or empty feeling in the epigastrium.

Diagnosis must be supported by endoscopic findings and/or *x* ray studies with barium.

PPENDIX 3
st of variables entered into multiple logistic egression

Name	Definition	p Value to enter
Main		
SEX	1 = male 2 = female	Forced
AGE.1	0.1*(AGE-50)	Forced
VILL	Village	0.05
F.1	F/128	0.05
IX_SMO.1	Smoking index (1–5)	0.05
LN_HT	Ln(HEIGHT/m)	0.05
LN_WT	Ln(WEIGHT/kg)	0.05
EDU	Education (see text)	0.05
OCU	Occupation (see text)	0.05
NUT	Nutrition (see text)	0.05
Higher order		
AGE.2	$(AGE.1)^2$	0.01
AGE.3	$(AGE.1)^3$	0.01
IX_SMO.2	$(IX\_SMO.1)^2$	0.01
F.2	$(F.1)^2$	0.01
Interactions		
All two way interactions between main effects		0.01

EFERENCES

Bittles AH, Neel JV. The costs of human inbreeding and their implications for variations at the DNA level. *Nat Genet* 1994;**8**:117–21.

2 **Charlesworth B**, Hughes KA. The maintenance of genetic variation in life-history traits. In: Singh RS, Krimbas CB, eds. *Evolutionary genetics: from molecules to morphology*, vol 1. Cambridge: Cambridge University Press, 1999.
3 **Charlesworth B**, Charlesworth D. The genetic basis of inbreeding depression. *Genet Res* 1999;**74**:329–40.
4 **Wright AF**, Hastie ND. Complex genetic diseases: controversy over the Croesus code. *Genome Biol* 2001;**2**:1–8.
5 **Bittles AH**, Mason WM, Greene J, Rao NA. Reproductive behaviour and health in consanguineous marriages. *Science* 1991;**252**:789–94.
6 **Charlesworth B**, Hughes KA. Age-specific inbreeding depression and components of genetic variance in relation to the evolution of senescence. *Proc Natl Acad Sci USA* 1996;**93**:6140–5.
7 **Hughes KA**, Alipaz JA, Drnevich JM, Reynolds RM. A test of evolutionary theories of ageing. *Proc Natl Acad Sci USA* 2000;**99**:14286–91.
8 **Martin AO**, Kurczynski TW, Steinberg AG. Familial studies of medical and anthropometric variables in a human isolate. *Am J Hum Genet* 1973;**25**:581–93.
9 **Wright AF**, Charlesworth B, Rudan I, Carothers A, Campbell H. Polygenic nature of late onset disease. *Trends Genet* 2003;**19**:97–106.
10 **Rudan I**, Smolej-Narancic L, Campbell H, Carothers A, Wright AF, Rudan P. Inbreeding and the genetic complexity of human hypertension. *Genetics* 2003;**163**:1011–21.
11 **Krieger H**. Inbreeding effects on metrical traits in Northeastern Brasil. *Am J Hum Genet* 1968;**21**:537–46.
12 **Hurwich BJB**, Rosner N, Nubani E, Kass H, Lewitter HI. Familial aggregation of blood pressure in a highly inbred community, Abu Ghosh, Israel. *Am J Epidemiol* 1982;**115**:646–56.
13 **Thomas JD**, Douchette MM, Thomas DC, Stoeckle JD. Disease, lifestyle and consanguinity in 58 American gypsies. *Lancet* 1987;**2**:377–9.
14 **Wahid Saeed AA**, Al Shammary FJ, Khoja TA, Hashim TJ, Anokute CC, Khan SB. Prevalence of hypertension and socio-demographic characteristics of adult hypertensives in Riyadh City, Saudi Arabia. *J Hum Hypertens* 1996;**10**:583–7.
15 **Halberstein RA**. Blood pressure in the Caribbean. *Hum Biol* 1999;**71**:659–84.
16 **Rudan P**, Simic D, Smolej-Narancic N, *et al*. Isolation by distance in Middle Dalmatia-Yugoslavia. *Am J Phys Anthropol* 1987;**74**:417–26.
17 **Rudan P**, Sujoldzic PA, Simic D, Bennett LA, Roberts DF. Population structure in the eastern Adriatic: the influence of historical processes, migration patterns, isolation and ecological pressures, and their interaction. In: Roberts DF, Fujiki N, Torizuka K, eds. *Isolation, migration and health*. Cambridge: Cambridge University Press, 1992.
18 **Rudan I**, Campbell H, Rudan P. Genetic epidemiological studies of Eastern Adriatic Island Isolates, Croatia: objectives and strategies. *Coll Anthropol* 1999;**23**:531–46.
19 **Wright S**. Coefficients of inbreeding and relationship. *Am Naturalist* 1922;**56**:330–8.
20 **Tay JS**, Yip WC. The estimation of inbreeding from isonymy: relationship to the average inbreeding coefficient. *Ann Hum Genet* 1984;**48**:185–94.
21 **Law MR**, Wald NJ. Risk factor thresholds: their existence under scrutiny. *BMJ* 2002;**324**:1570–6.
22 **Abney M**, McPeek MS, Ober C. Broad and narrow heritabilities of quantitative traits in a founder population. *Am J Hum Genet* 2001;**68**:1302–7.
23 **Ravussin E**, Valencia ME, Esparza J, Bennett PH, Schulz LO. Effects of a traditional lifestyle on obesity in Pima Indians. *Diabetes Care* 1994;**17**:1067–74.
24 **Fay JC**, Wyckoffand GJ, Wu CI. Positive and negative selection on the human genome. *Genetics* 2001;**158**:1227–34.
25 **Lifton RP**, Gharavi AG, Geller DS. Molecular mechanisms of human hypertension. *Cell* 2001;**104**:545–6.
26 **Stoll M**, Kwitek-Black AE, Cowley AW, *et al*. New target regions for human hypertension via comparative genomics. *Genome Res* 2001;**10**:473–82.
27 **Jimenez JA**, Hughes KA, Alaks G, Graham L, Lacy RC. An experimental study of inbreeding depression in a natural habitat. *Science* 1994;**266**:271–3.
28 **Bittles AH**. Empirical estimates of the global prevalence of consanguineous marriage in contemporary societies. In: *Morrison Institute for Population and Resource Studies Working Report 74*. Stanford: Stanford University, 1988.
29 **Bittles AH**, Savithri HS, Murthy G, Baskaran G, Wang W. Consanguinity: a familiar story full of surprises. In: Macbeth H, Shetty P, eds. *Health and ethnicity*. London: Taylor and Francis, 2001.
30 **Broman KW**, Weber JL. Long homozygous chromosomal segments in reference families from the centre d'Etude du polymorphisme humain. *Am J Hum Genet* 1999;**65**:1493–500.
31 **Ezzati M Lopez AD**, Rodgers A, Vander Hoom S, Murray CJ. Selected major risk factors and global and regional burden of disease. *Lancet* 2002;**360**:1347–60.
32 **Vaupel JW**. Trajectories of mortality at advanced ages. In: Wachter KW, Finch CE, eds. *Between Zeus and the salmon; the biodemography of longevity*. Washington DC: National Academy Press, 1997:17–37.

Coll. Antropol. **28** (2004) 2: 585–601
UDC 575.222.5:616.71-007.234
Original scientific paper

Inbreeding and Susceptibility to Osteoporosis in Croatian Island Isolates

Igor Rudan*[,1,2], Tatjana Škarić-Jurić*[,3], Nina Smolej-Narančić[3], Branka Janićijević[3], Diana Rudan[4], Irena Martinović Klarić[3], Lovorka Barać[3], Marijana Peričić[3], Radoslav Galić[5], Margaret Lethbridge-Čejku[6] and Pavao Rudan[3]

[1] School of Public Health »Andrija Štampar«, University Medical School, Zagreb, Croatia
[2] Department of Public Health Sciences, University Medical School, Edinburgh, UK
[3] Institute for Anthropological Research, Zagreb, Croatia
[4] General Hospital »Sveti Duh«, Zagreb, Croatia
[5] Department of Statistics, School of Medicine, University »J. J. Strossmayer«, Osijek, Croatia
[6] U. S. Department of Health and Human Services, Center for Disease Control and Prevention, National Center for Health Statistics, Hyattsville, MD, USA

ABSTRACT

The aim of this study was to investigate a recessive genetic component in susceptibility to osteoporosis (OP) by comparing its prevalence in isolated villages of three Croatian islands: Brač, Hvar and Korčula with different levels of inbreeding. A random sample of 20–30% adults from 14 villages was obtained, including a total of 1,389 examinees. The average inbreeding coefficient (F) of examinees from each village population was estimated using Wright's path method (based on genealogical information). The morphometry of the metacarpal bones was performed on hand-wrist radiographs of both hands in all examinees. OP was defined as values of cortical index smaller than 2 standard deviations based on distribution of values in examinees of the same sex under 45 years of age. Mean values of cortical index (CI) and prevalence of OP (both standardized by age and weighted for the sample size) in each village were correlated to the mean inbreeding coefficient (F). The coefficient of correlation (r) between F values and CI was –0.28 in males ($p=0.08$) and –0.42 in females ($p=0.005$), and between F and OP prevalence 0.32 in males ($p<0.001$) and 0.43 in females ($p<0.001$). These results indicate a trend of increased susceptibility to osteoporosis with increasing level of inbreeding in isolated communities of Croatian islands.

***Key words**: inbreeding, cortical index, osteoporosis, isolate populations, Croatia*

Received for publication October 12, 2004
* These authors equally contributed to the paper

Introduction

Osteoporosis is a common public health problem of post-reproductive age, characterized by reduced bone mass, changes in micro-architecture of the bone tissue and increased risk of fractures subsequent to those changes[1,2]. Similarly to many other common complex diseases of late onset, most cases in population probably result from the action of many different genes and their interaction with the environment[2,3]. However, osteoporosis seems to have much stronger genetic basis than most of the other late-onset diseases[4–7]. Twin and family studies indicate that a large majority of variance in quantitative traits such as bone mineral density can be explained by hereditary factors, and there is also a high correlation between siblings in skeletal geometry and bone turnover[1,4]. Some cases in population arise as a consequence of single-gene disorders, e.g. »osteoporosis-pseudoglioma syndrome«. Genome-wide linkage studies identified multiple candidate loci on chromosomes 1, 2, 5 and 11, but those associations seem to be modified by dietary calcium and vitamin D intake, which makes them difficult to repeat[4]. Significant non-hereditary risk factors include low calcium intake, vitamin D deficiency, physical inactivity, cigarette smoking, excessive consumption of protein, caffeine and alcohol, low body mass index and use of bone-resorbing medications[8–13].

There is great current interest in understanding genetic architecture of common complex diseases such as osteoporosis, as this is expected to lead to the development of genetic markers of increased disease risk and new therapeutical targets[3]. In this paper, we present an approach to study of osteoporosis that could provide a support for its predominantly genetic determination, with susceptibility mediated through a number of recessive genetic variants, most of them having a small individual effect on disease risk. The reasoning is simple: if a modest increase in number of genes identical by descent (e.g. an increase of inbreeding coefficient from 0% to 3%, predicted to affect about 800 genes) leads to significant changes in prevalence of osteoporosis, this is only consistent with large number of genomic loci influencing the disease. This conclusion is more valid if the study is conducted in an isolate population, in which the variation in environmental pressures is minimal and consanguinity is prevalent. Therefore, the studied population included 14 isolate villages from the eastern Adriatic islands of Hvar, Brač and Korčula, Croatia, an isolate resource well characterized through long-term multidisciplinary researches[14–17].

In this unique metapopulation of distinct human isolates, there is a long history of anthropological research into the determinants of skeleton-related biological traits. Initially, comparisons of within-population and between-population variation in traits such as metacarpal bone dimensions were used along with a larger number of other traits to assess population structure. The studies performed by the staff of the Institute for Anthropological Research in Zagreb, Croatia, consistently showed excellent compliance with models of population structure such as »isolation by distance« in several island populations, i.e. better »fit« to the model than observed for most of the other studied biological, bio-cultural and socio-cultural traits[18–27]. This group also characterized in detail the effects of gender and aging on bone loss[24–30] and reported on specific populations in which the expected effects of age and gender could not be shown[31]. The effects of occupation on bone loss and osteoarthritis were investigated[32,33]. An attempt to analyse genetic basis of the metacarpal bone dimensions was initially made through the analyses of their latent structure[34]. Recently, this

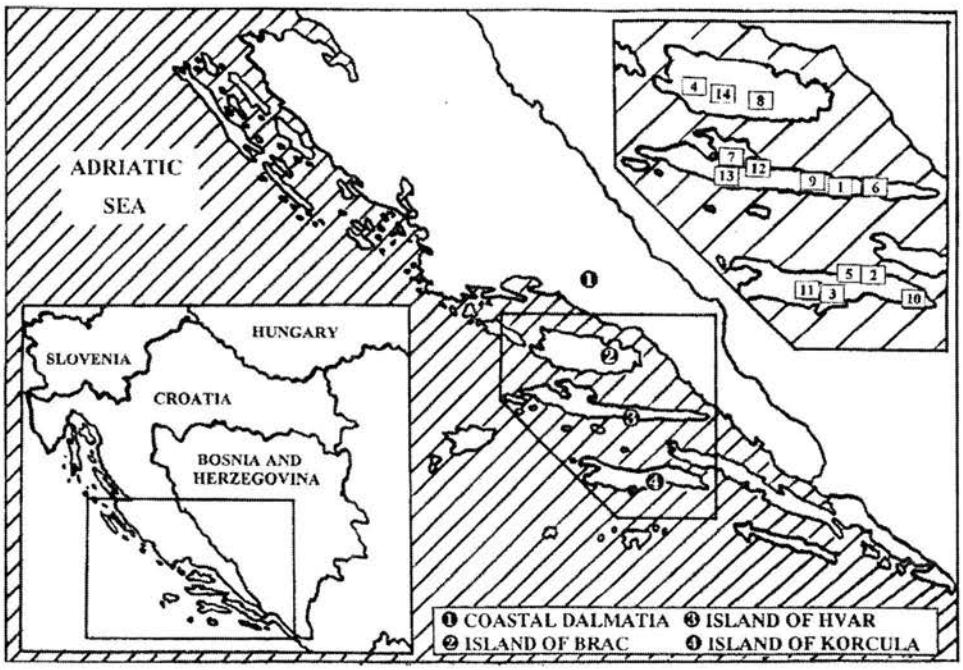


Fig. 1. Map of Dalmatian island genetic isolates showing study islands and villages (1–14).

group has also made progress in studying genetic determinants of bone-related measurements using pedigree data analyses[5,35].

Materials and Methods

A. Population choice

Croatia has 15 Adriatic Sea islands with population greater than 1,000. The villages on the islands have unique population histories and relative isolation from their neighboring villages and from the mainland through centuries. The 14 village populations of the three neighboring islands of the eastern Adriatic, Middle Dalmatia, Croatia (Brač, Hvar and Korčula, Figure 1) have been investigated by the Institute for Anthropological Research in Zagreb since early 1970's. More than 200 studies on these populations have been published to date in collaboration with many international scientists. The long-term researches included characterization of ethnohistory, migration patterns, genealogical reconstruction, analyses of many quantitative and qualitative biological traits, records of medical problems and study of environmental, socio-cultural and genetic characteristics. The most informative overviews of the results can be found in the papers of Rudan et al.[14–16,27], Bennett et al.[17] and Waddle et al.[36].

In Table 1, we list main sources that provide detailed information on population structure, inbreeding effects, monogenic (Mendelian) diseases and rare genetic variants found in particular islands and reported in the literature[37–69].

The 14 villages chosen for this study on the islands of Brač, Hvar and Korčula were founded during one of the three pe-

TABLE 1
BRIEF REVIEW OF POPULATION GENETIC RESEARCH UNDERTAKEN IN CROATIAN ISLAND ISOLATES IN THE PAST 33 YEARS

Type of research	Island	Ref.
A. Studies of population genetic variation		
STR polymorphisms	Krk, Brač, Hvar, Korčula	37–40
VNTR polymorphisms	Hvar	41
Y-chromosome haplogroups	Krk, Brač, Hvar, Korčula	42
mtDNA haplogroups	Krk, Brač, Hvar, Korčula	43,44
HLA markers or immunoglobulin allotypes	Krk, Hvar, Silba, Olib, Pag	45–48
Serogenetic polymorphisms	Brač, Hvar, Korčula, Silba, Olib,	49–52
B. Reports on autochthonous Mendelian diseases		
Dwarfism	Krk	53–55
Albinism	Krk	55
Progressive spastic quadriplegia	Krk	55
Familial cognitive dysfunction	Susak	56
Familial congenital hip dislocation	Lastovo	57
Familial ovarian cancer	Lastovo	58,59
Keratoderma palmoplantaris transgrediens	Mljet	60
C. Reports of high population frequencies of extremely rare genetic variants		
Deleted/triplicated alpha-globin gene	Silba	61
PGM1*W3 phosphoglucomutase-1 variant	Olib	62
mtDNA haplogroup F	Hvar	63
Y-chromosome haplogroup P*	Hvar	42
D. Studies of inbreeding effects		
Incidence of cancer	Brač, Hvar, Korčula, Vis, Lastovo	59
Prevalence of hypertension	Brač, Hvar, Korčula	64,65
Prevalence of 10 complex chronic diseases	Brač, Hvar, Korčula	66
Prevalence of learning disability	Brač, Hvar, Korčula, Susak	67
Prevalence of nephrolithiasis	Brač, Hvar, Korčula	68
Prevalence of malocclusion	Hvar	69

riods: BC era (by admixture of Illyrians, Greeks and succeeding Romans), 7th century AD (by Croats who immigrated from Asia) and 16–18th century AD (by Croats who fled from Balkans peninsula fearing Ottoman expansion). The subsequent tendency towards inbreeding in each village was influenced by a combination of geographic reasons (isolation), political reasons (»*The Paštrović Privileges*«) and socio-cultural reasons which were all extensively discussed elsewhere[14–17,27,70–72].

B. Field work and sample collection

During the field research between 1978 and 1987 undertaken by the Institute for Anthropological Research in Zagreb, the information was collected and various measurements performed on 1,389 adult individuals (682 males and 707 females) selected randomly from voting lists to form approximately 20–30% of the total population of these 14 villages. The collected data characterizing environmental variation included proportion of inhabitants with some college education (EDU), occupation in agriculture and fishery (OCC), regular consumption of traditional Mediterranean diet (NUT), smoking habits (SMO) and measurement of body mass index (BMI) (Table 2). The details on nutritional status assessment based on BMI are reported by Smolej Narančić[73]. Table 2 supports the hypothesis of decreased variation in most of the studied characteristics related to environment. This is important, as socio-economic status, occupation, diet, obesity and climate are usually highlighted as potential environmental risk factors for many common late onset diseases.

C. Estimation of mean inbreeding coefficients in each village

Genetic characterization of the 14 villages (Figure 1) included the computation of average inbreeding coefficient of each village based on three different methods: (i) reconstruction of genealogies for each examinee, (ii) analysis of parental isonymy from surname distribution and (iii) analysis of genotype distributions of MN, Ss and Kk serogenetic polymorphisms from blood samples obtained from all the examinees. Individual inbreeding coeffi-

TABLE 2
PREVALENCE OF FACTORS RELATED TO SOCIO-ECONOMIC STATUS, LIFESTYLE AND ENVIRONMENT IN 14 VILLAGES UNDER STUDY

Village[1]	EDU (%)	OCC (%)	NUT (%)	SMO (%)	BMI ($\bar{x}$)
Gdinj (H)	2.3	80.8	93.8	4.6	25.1
Pupnat (K)	2.1	89.7	91.8	9.3	26.7
Čara (K)	1.5	89.0	94.9	22.6	26.4
Dračevica (B)	2.4	85.7	90.5	23.8	28.6
Račišće (K)	1.9	87.5	83.6	23.1	27.3
Bogomolje (H)	2.5	84.0	93.8	4.9	23.0
Vrisnik (H)	2.1	82.3	91.6	13.5	25.5
G. Humac (B)	2.5	81.3	87.5	38.8	27.3
Zastražišće (H)	4.0	86.3	93.6	5.9	25.1
Lumbarda (K)	4.7	85.2	89.8	21.3	26.2
Smokvica (K)	3.1	81.4	81.4	27.8	26.0
Svirče (H)	2.0	80.1	83.2	12.8	25.7
Dol (H)	4.2	75.8	86.3	8.4	26.1
Nerežišća (B)	1.8	60.7	79.5	34.8	28.0

EDU = percentage of the population with higher education degree; OCC = percentage of the population working in agriculture/fishery; NUT = percentage of the population consuming »Mediterranean diet«; SMO = percentage of the population smoking; BMI = mean value of body mass index. The capital letters next to village names represent the islands (B=Brač, H=Hvar, K=Korčula).

cients were computed independently for each of 1,389 individuals. The pedigree information on 2–3 ancestral generations that was recorded for each examinee during the initial field work (1978–1987) was expanded during 1997–2000 through insight into the parish registries stored in local churches to allow the completion of the information on 4 ancestral generations in each examinee. The individual inbreeding coefficients (F) were then computed according to Wright's path method[74]:

$$F = \sum_{(1\to c)} (1/2)^{(n_i+m_i+1)}$$

where m and n refer to the number of paths from a common ancestor, and c refers to the number of common ancestors. The genealogical inbreeding coefficient for each village was then computed as the average of all individual F values. To further support these estimates, F was calculated from isonymy as[75,76]:

$$F = (P - \Sigma p_k q_k) / 4(1 - \Sigma p_k q_k) = (\Sigma p_k q_k) / 4 - (P - \Sigma p_k q_k)(\Sigma p_k q_k) / 16(1 - \Sigma p_k q_k)$$

where p_k is the frequency of the surname k in males, q_k in females, P is the proportion of marriages between spouses carrying the same surname among all marriages, and the summation is over all surnames. Apparently, in this approach the units of investigation are marital pairs of examinees' parents and not the examinees themselves. Finally, the inbreeding in each village was also assessed from the departure in the frequency of heterozygotes from the expectations based on Hardy-Weinberg equilibrium[77], where F is calculated as:

F = 1 – (observed proportion of heterozygotes / expected proportion of heterozygotes)

and where the expected proportion of heterozygotes calculated from allelic frequencies p and q of 2 alleles present in the population (MN, Ss and Kk) equals $2pq$. F values calculated from each of the 3 polymorphisms were added and divided by 3 to obtain the average F value for each village. The process of blood sample collection, storage, transport and analysis at the University of Newcastle upon Tyne was described in detail by Roberts et al.[49]. Table 3 reviews the average coefficients of inbreeding obtained by the applied three methods. Tables 2 and 3 indicate that the selected village populations represent an excellent model for studying effects of inbreeding, as a wide range of inbreeding coefficients is present while the concern over considerable confoundings related to environmental variance is reduced.

D. Cortical index and osteoporosis prevalence estimation

The osteometric dimensions of metacarpal bones are an efficient and practical method for investigation and monitoring bone mass. At the time when data used in the present analysis have been gathered, the radiogrammetry of the metacarpal bones has been widely used as the best available screening method of bone status in population studies. Procedure of metacarpal bones osteometry, as thoroughly described by Barnett and Nordin[78], was followed in field studies performed on all three investigated islands. Hand-wrist radiographs were taken using a single portable X-ray. Total diaphysis width (D) and medullary canal width (d) of the second left metacarpal bone was determined by a single, experienced observer[78]. Measurements were performed by one investigator using a millimeter ruler and a magnifying glass (×10) with a scale permitting 0.05-mm accuracy. Measurements were rounded to 0.1 mm. For each individual and for each bone, the cortical index (CI) was computed as:

$$CI = (D - d) \times 100 / D.$$

Tables 4 and 5 show the sample sizes and descriptive statistics of age and corti-

TABLE 3
GENETIC STRUCTURE OF 14 VILLAGES UNDER STUDY. THE VILLAGES ARE RANKED ACCORDING TO THE ESTIMATED AVERAGE INBREEDING COEFFICIENT CALCULATED FROM THE GENEALOGICAL DATA *(Fgen)*

Village	F(gen)	F(iso)	F(sgp)
Gdinj (H)	0.049	0.107	0.041
Pupnat (K)	0.044	0.034	0.108
Čara (K)	0.032	0.040	0.016
Dračevica (B)	0.031	0.031	0.049
Račišće (K)	0.027	0.034	0.004
Bogomolje (H)	0.025	0.030	0.018
Vrisnik (H)	0.015	0.023	0.013
G. Humac (B)	0.013	0.016	–0.016
Zastražišće (H)	0.013	0.013	–0.021
Lumbarda (K)	0.012	0.025	–0.010
Smokvica (K)	0.008	0.019	–0.107
Svirče (H)	0.008	0.008	–0.018
Dol (H)	0.005	0.004	–0.041
Nerežišća (B)	0.002	0.004	–0.076

The corresponding estimated based on isonymy *(Fiso)* and codominant serogenetic polymorphisms MN, Ss and Kk *(Fsgp)* are presented to support the results obtained through Wright's »path«-method *(Fgen)*. The capital letters next to village names represent the islands (B=Brač, H=Hvar, K=Korčula).

cal index by village in male and female examinees. As expected, the larger effect of age on cortical index and on osteoporosis prevalence was observed in females than in males (Figures 2 and 3). It is apparent that the effects of age are minimal in younger subjects (plateau) and around menopausal ages a decline in cortical index could be observed. This trend is more pronounced in females, which is in accordance with the findings in other populations. Complex segregation analysis of cortical index of the metacarpal bones has been performed by Ginsburg et al.[35] using pedigree data from the same populations. CSA model implemented included sex-specific parameterizations of inflection points of cortical index data and results – being 45 years for males and females in two most parsimonious models – which allows us to use 45 years as the reliable referent point for current analysis.

Prevalence of osteoporosis was established according to the statistical criteria. It was based on the distribution of cortical index values in people under the age of 45 in each gender separately. A »cut-off« value of cortical index was defined as the mean minus two standard deviations of the distribution in each sex. As this criterion has been widely used, and the measurements were performed by a single device and technician and analyzed by a single experienced assessor, we believe that the likelihood of substantial procedural errors is minimal.

TABLE 4
DESCRIPTIVE STATISTICS OF AGE AND CORTICAL INDEX BY VILLAGE IN MALE EXAMINEES.

Village	N	Age (yrs.)				Cortical index (%)			
		$\bar{x}$	SD	Min.	Max.	$\bar{x}$	SD	Min.	Max.
Gdinj (H)	51	54.53	11.72	24	81	55.97	7.89	37.78	72.22
Pupnat (K)	46	42.76	12.91	23	71	55.17	6.84	44.85	68.35
Čara (K)	63	42.59	11.87	20	71	53.11	8.42	32.37	72.07
Dračevica (B)	20	46.80	16.37	23	74	66.28	7.15	47.62	80.00
Račišće (K)	40	43.30	11.22	20	62	55.14	7.30	36.23	73.11
Bogomolje (H)	46	57.85	15.65	24	81	48.90	6.17	32.14	64.44
Vrisnik (H)	44	37.68	10.52	23	54	57.22	6.77	42.11	69.89
G. Humac (B)	27	51.04	16.81	22	85	62.87	6.92	49.49	76.14
Zastražišće (H)	69	50.09	14.78	20	77	53.41	7.45	32.43	76.09
Lumbarda (K)	53	43.21	11.36	22	77	59.70	7.60	41.09	77.83
Smokvica (K)	52	40.87	10.53	24	59	53.46	7.36	37.75	69.41
Svirče (H)	70	39.94	11.10	21	55	59.36	9.78	39.81	82.76
Dol (H)	50	41.44	10.37	20	56	56.73	6.15	40.00	70.51
Nerežišća (B)	51	45.59	13.86	22	81	62.31	9.07	44.12	82.02
Total	682	45.25	13.73	20	85	56.50	8.59	32.14	82.76

TABLE 5
DESCRIPTIVE STATISTICS OF AGE AND CORTICAL INDEX BY VILLAGE IN FEMALE EXAMINEES.

Village	N	Age (yrs.)				Cortical index (%)			
		$\bar{x}$	SD	Min.	Max.	$\bar{x}$	SD	Min.	Max.
Gdinj (H)	71	53.48	12.28	27	82	55.64	9.99	37.21	84.15
Pupnat (K)	50	44.76	11.74	21	61	56.84	8.67	34.53	75.00
Čara (K)	75	43.63	12.38	20	63	56.29	8.74	37.63	83.09
Dračevica (B)	23	52.48	15.82	21	87	65.11	12.46	42.22	92.41
Račišće (K)	63	43.29	11.83	21	71	58.22	8.58	40.18	77.14
Bogomolje (H)	32	58.50	15.62	19	78	50.44	9.89	29.35	75.00
Vrisnik (H)	28	41.43	8.69	23	54	61.20	9.28	43.37	76.40
G. Humac (B)	44	49.09	13.81	25	82	64.93	11.24	39.76	91.21
Zastražišće (H)	54	52.37	12.93	24	83	54.46	9.41	37.76	83.33
Lumbarda (K)	55	43.09	12.27	22	74	59.58	9.94	37.70	80.00
Smokvica (K)	42	44.48	12.41	23	61	58.28	11.17	32.08	82.45
Svirče (H)	65	39.25	10.88	20	55	63.90	9.81	40.26	87.18
Dol (H)	43	41.86	10.08	22	56	63.30	9.29	46.07	84.00
Nerežišća (B)	62	44.76	12.15	19	77	65.02	10.82	37.21	91.55
Total	707	46.15	13.21	19	87	59.31	10.60	29.35	92.41

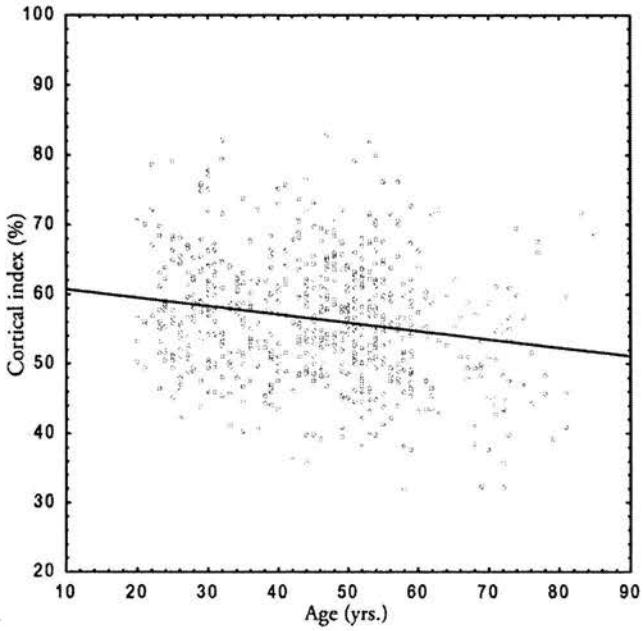


*Fig. 2. Effect of age on cortical index in males (r = –0.19, p<0.001, y = 61.932 – 0.120*x).*

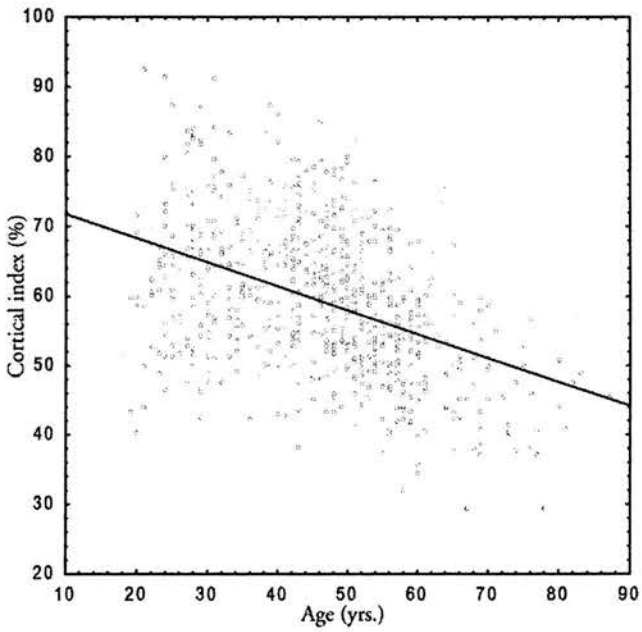


*Fig. 3. Effect of age on cortical index in females (r = –0.43, p<0.001, y = 72.251 – 0.345*x).*

E. Statistical analyses

To investigate the relationship between inbreeding and the cortical index and prevalence of osteoporosis, the ecologic epidemiological design was used[79]. The cortical index values and the prevalence of osteoporosis were compared among 14 villages with various levels of mean inbreeding. The cortical index values were adjusted for age effects in each sex by means of multiple regression (age, age^2) and mean values of standardized residuals were calculated for each village. As the estimated OP prevalence could be considerably influenced by variations in sex and age distribution of the sample sizes of different village they were adjusted according to the age and sex distribution in total sample and the results of the weighted data were presented separately for each sex. All statistical analyses were performed using »Statistica 6« software.

Results

In present study, the possible influence of inbreeding on bone mass was tested using ecologic-epidemiological design. Cortical index of metacarpal bones has been used as an indicator of total bone mass and good screening tool for susceptibility for osteoporosis appropriate for the population studies.

The coefficient of correlation (Figures 4 and 5) between F-values of each investigated village and mean values of the standardized age-adjusted residuals of metacarpal cortical index was –0.28 in males (p=0.08) and –0.42 in females (p=0.005). Correlations between F-values and estimated prevalence of osteoporosis weighted for sample sizes (Figures 6 and 7) were 0.32 in males (p<0.05) and 0.43 in females (p<0.001). Although the correlations in both traits were more significant in females than in males, there is a general concordance between values obtained in both sexes from the same village, which supports the hypothesis that the findings are indeed due to village-specific effects.

Obtained results clearly show a tendency of increased susceptibility to osteoporosis with increasing level of inbreeding

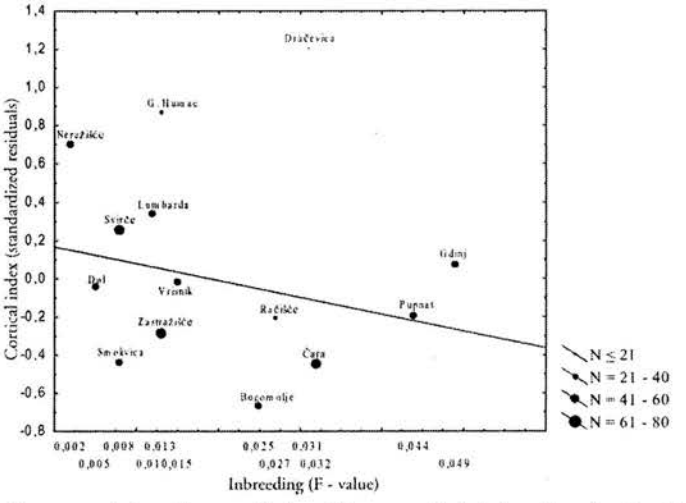


*Fig. 4. Effect of average inbreeding coefficient (F) on cortical index (standardized age-adjusted residuals weighted according to sample size) in male examinees in 14 villages (r = –0.28, p = 0.077, y = 0.168 – 8.823*x).*

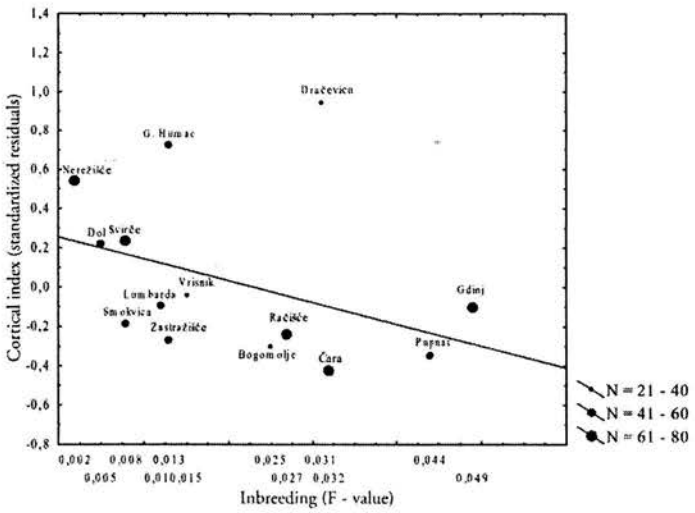


*Fig. 5. Effect of average inbreeding coefficient (F) on cortical index (standardized age-adjusted residuals weighted according to sample size) in female examinees in 14 villages (r = –0.42, p = 0.005, y = 0.255 – 11.057*x).*

in isolated communities of three Croatian islands. However, few villages are displaying as outliers in this general trend: Dračevica (Brač) with exceptionally high cortical index values and Čara (Hvar) with exceptionally high prevalence of osteopo-

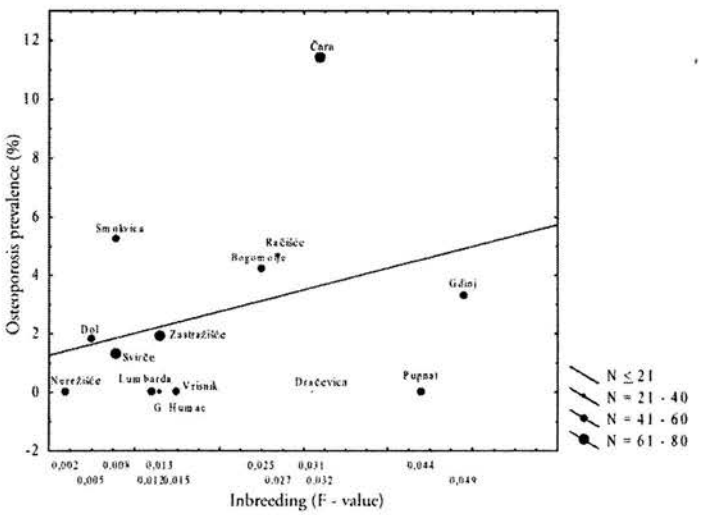


*Fig. 6. Effect of average inbreeding coefficient (F) on prevalence of osteoporosis (standardized by age, weighted according to sample size) in male examinees in 14 villages (r = 0.32, p<0.001, y = 1.269 + 74.598*x).*

rosis. We suspect that those villages are the sites of some possibly very intriguing genetic effects (drift) and therefore could be the promising sites for incoming studies of susceptibility/protection genes osteoporosis.

Discussion

There are several possible reasons why the effect of inbreeding on late-onset complex chronic diseases has not been widely evaluated to date. Firstly, in countries where inbreeding is prevalent in the population, life expectancy is generally considerably shorter than in western communities, and late-onset diseases do not represent the main public health problem. Therefore, the effects of inbreeding were mostly investigated in small isolated communities (geographic, cultural, linguistic, ethnic or religious isolates) in developed countries, which could be more easily reached. In most of the developed countries, however, the isolated human populations characterized by inbreeding often do not have the same level of access to health care as general (especially urban) population where the public health sector is well developed and the majority of epidemiological studies are being undertaken. Therefore, the health status of human isolates is not easily evaluated from medical records or communicated with their local physicians. In addition, there are not many isolate popu- lations world-wide with well-preserved parish registries from which reliable estimates of inbreeding coefficients can be determined based on the familial relationships over at least several ancestral generations. Furthermore, there are usually multiple concerns over confounding factors, as it is often quite difficult to find a non-inbred control population which would match the studied inbred isolate in environmental exposures and differ significantly only in genetic structure. Contemporary isolates usually share specific climate and environment, as well as a multitude of common socio-cultural factors such as diet, lifestyle, religion practices and socio-eco-

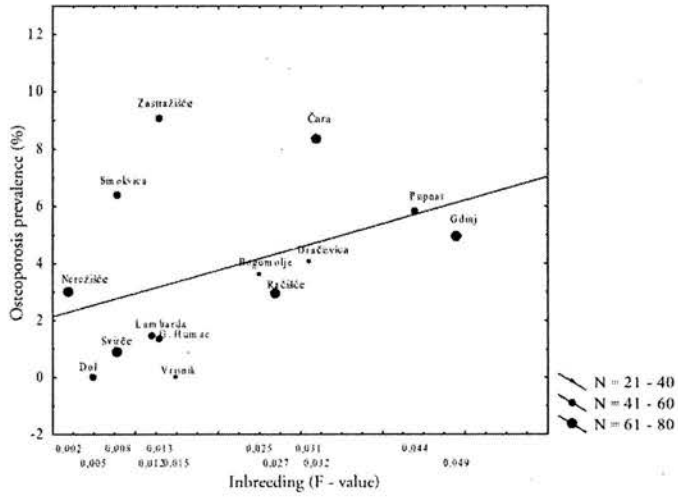


*Fig. 7. Effect of average inbreeding coefficient (F) on prevalence of osteoporosis (standardized by age, weighted according to sample size) in female examinees in 14 villages (r=0.43, p<0.001, y = 2.136 + 81.568*x).*

nomic status for which it is very difficult to control. In small and isolated communities the phenomena such as founder effect, genetic drift and inbreeding can significantly affect allele and genotype frequencies, which invalidates comparisons to control populations where genotype frequencies follow Hardy- Weinberg equilibrium[68]. Khoury concluded that, despite rare attempts, there is still hardly any study that would satisfactorily deal with all the usual concerns including small sample sizes, unreliable inbreeding estimates, doubtful disease dia- gnoses and inappropriate control populations[80]. We recently attempted to conduct a number of studies that would satisfactorily deal with those issues[64–69].

In this study, several precautions were taken to control for potential confounding effects. Primarily, a unique population was chosen, well characterized genetically and epidemiologically through almost 3 decades of continuing research by both local and international anthropologists. The subdivision of the islanders of the eastern Adriatic into small villages and a diversity of their attitudes towards inbreeding influenced by geographic isolation, political privileges in the past, as well as socio-cultural reasons resulted in a range of inbreeding coefficients present at both individual and population level. At the same time, the environmental variation within and between these populations is reduced (shared climate, religion, lifestyle, nutrition, predominant occupation and the level of education), as shown in Table 2. Therefore, we believe that these populations represent exceptionally good setting for undertaking the study of the effects of the inbreeding on complex chronic disease such as osteoporosis[59,64–69].

Further advantage of Croatian island isolates is that each of selected 14 villages in this study has its own church with well preserved records of births, marriages and deaths dating back to 1750, which helped reconstruct the genealogies and determine inbreeding coefficients. In addition, each village also has its own health clinic with the full-time local general practitioner and a nurse. The bone X-rays were performed in these clinics by the same device and same assessor from the Institute of Anthropological Research in Zagreb, as a very objective measurement of phenotype. As the results were later analyzed by a single and experienced observer, we believe that the like- lihood of substantial measurement errors is minimal (both observer-related and village-specific). An attempt was also made to avoid significant confounding effects of genetic drift and founder effect in specific villages. This has been achieved by including several villages of similar inbreeding levels from different islands into the groups with »high«, »moderate« and »low« inbreeding, as it is unlikely that the two population genetic phenomena would affect gene frequencies in the same direction in all villages[68].

The results showed a significant decreasing effect of inbreeding on cortical index (initially standardized by age) across 14 villages (Figures 4 and 5). Similarly, the prevalence of osteoporosis found in these villages appeared to increase from about 1.3% in villages with low inbreeding prevalence to nearly 4.0% in villages with average inbreeding coefficient close to F=0.05, i.e. a 3-fold increase. This result reaffirms the conclusion of our recent study where non-specific risk of inbreeding has been reported for a number of complex chronic diseases[66]. Analysis by village (Figures 6 and 7) shows that the standardized prevalence of osteoporosis follows the increase in average inbreeding coefficients, although this correlation is not really linear as there are villages with extremely increased prevalence in both sexes (e.g. Čara, with OP prevalence of about 10%). In addition, not all villages with higher rates of inbreeding reveal

high prevalence of osteoporosis (e.g. Pupnat in males). Those observations might be due to specific genetic structure of those villages which are characterized by increased or decreased frequencies of rare alleles with large effects mediating susceptibility to osteoporosis, possibly due to combined effect of genetic drift and founder effect.

If we accept the conclusion that the increase in inbreeding of about 2–3% (from F=0.005 to 0.03) could be, to some extent, responsible for the observed increase in prevalence of osteoporosis, the central question becomes what does it tell us about the genetic basis of the disease. Apparently, the susceptibility seems to be controlled, at least partly, by the recessive genetic variants that are more likely to be rare than common, as the inbreeding effects on rare variants are more apparent. In addition, if we accept that the total number of genes in the human genome, according to the most recent estimates, is about 25,000[81], then the increase in inbreeding of 3% would correspond to having about 800 random genes across the genome identical by descent. If this unrecombined autozygosity in only 3% of genes could lead to a notable effect in the prevalence of osteoporosis, there are two main mechanisms that could explain it. Firstly, inbreeding may mainly act through the rare variants of major effect, that may be enriched in frequency in each village by combined effects of founder effect and subsequent genetic drift. Second, the genes controlling this trait are of small effect and very numerous, scattered across the entire genome. The design of this study provides some arguments against the first explanation. Major effect genes arise after mutations that are considered to be very rare, as the probability of random mutation causing small effect is much greater. Therefore, even if such mutations were present in some of the studied villages, it is extremely unlikely that similar effects of inbreeding would be observed across several villages, as our results indicated. In addition, under such assumption the differences between inbred and outbred individuals would normally be much larger than it was the case in our study. It is more plausible that the genetic control of cortical index, bone mass loss and susceptibility to osteoporosis in humans is at least partly controlled by a larger number of genes of small individual effects[68].

This study attempted to discuss the genetic basis of susceptibility to osteoporosis through inbreeding study in relatively isolated communities. The results provided further evidence on polygenic basis of susceptibility for osteoporosis with detectable effects of recessive genes.

Acknowledgements:

This paper was partly supported by the grants from the Ministry of Science, Education and Sport of the Republic of Croatia to P.R. (0196005), N.S.N. (0196001) and I.R. (0108330), The Royal Society to Professor Harry Campbell (H.C.) and I. R., and The Wellcome Trust (IRDA) to H. C. and I.R.

REFERENCES

1. RALSTON, S. H., Quart. J. Med., 90 (1997) 247. — 2. STEWART, T. L., S. H. RALSTON, J. Endocrinol., 166 (2000) 235. — 3. ALBAGHA, O. M., S. H. RALSTON, Endocrinol. Metab. Clin. North Am., 32 (2003) 65. — 4. RALSTON, S. H., J. Clin. Endocrinol. Metab., 87 (2002) 2460. — 5. ŠKARIĆ-JURIĆ, T., P. RUDAN, Coll. Antropol., 21 (1997) 447. — 6. KARASIK, D., E. GINSBURG, G. LIVSHITS, O. PAVLOVSKY, E. KOBYLIANSKY, Genet. Epidemiol., 19 (2000) 410. — 7. KOBYLIANSKY, E., D. KARASIK, V. BELKIN, G. LIVSHITS, Ann. Hum. Biol., 27 (2000) 433. — 8. DEAL, C. L., Am. J. Med., 102 (1997) 35S. — 9. PRENTICE, A., Public Health Nutr., 7 (2004) 227. — 10. NGUYEN, T. V., J. R. CENTER, J. A.

EISMAN, Med. J. Aust., 180 Suppl 5 (2004) S18. — 11. MATKOVIĆ, V., K. KOSTIAL, I. ŠIMONOVIĆ, R. BUZINA, A. BRODAREC, B. E. C. NORDIN, Am. J. Clin. Nutr., 32 (1979) 540. — 12. MATKOVIĆ, V.: Influence of age, sex and nutrition on bone loss. (Ph.D. Thesis, University of Zagreb, Zagreb, 1976) (In Croat.). — 13. STINI, W. A., Coll. Antropol., 27 (2003) 23. — 14. RUDAN, P., D. ŠIMIĆ, N. SMOLEJ NARANČIĆ, L. A. BENNETT, B. JANIĆIJEVIĆ, V. JOVANOVIĆ, M. F. LETHBRIDGE, J. MILIČIĆ, D. F. ROBERTS, A. SUJOLDŽIĆ, L. SZIROVICZA, Am. J. Phys. Anthropol., 74 (1987) 417. — 15. RUDAN, P., A. SUJOLDŽIĆ, D. ŠIMIĆ, L. A. BENNETT, D. F. ROBERTS, In: ROBERTS, D. F., N. FUJIKI, K. TORIZUKA (Eds): Isolation, Migration and Health. (Cambridge University Press, Cambridge, 1992). — 16. RU- DAN, I., H. CAMPBELL, P. RUDAN, Coll. Antropol., 23 (1999) 531. — 17. BENNETT, L. A., J. L. ANGEL, D. F. ROBERTS, P. RUDAN, Coll. Antropol., 7 (1983) 195. — 18. RUDAN, P., D. ŠIMIĆ, L.A. BENNETT, Am. J. Phys. Anthropol., 77 (1988) 97. — 19. ŠIMIĆ, D., P. RUDAN, Hum. Biol., 62 (1990) 113. — 20. RUDAN, I., P. RUDAN, L. SZIROVICZA, D. ŠIMIĆ, L. A. BENNETT, Homo, 47 (1996) 257. — 21. RUDAN, I., P. RUDAN, A. CHAVENTRE, B. JANIĆIJEVIĆ, J. MILIČIĆ, N. SMOLEJ NARANČIĆ, A. SUJOLDŽIĆ, Homo, 49 (1998) 201. — 22. RUDAN, I., P. RUDAN, In: BODZSAR, B. E., C. SUSSANNE (Eds.): Studies in Human Biology. (Eotvos University Press, Budapest, 1996, pp. 351–367). — 23. ŠIMIĆ, D., P. RUDAN, L. A. BENNETT, In: ROBERTS, D. F., A. CHAVENTRE (Eds.): Pluridisciplinary Approach of Human Isolates. (INED, Paris, 1990, pp. 149–162). — 24. RUDAN, P., J. L. ANGEL, L. A. BENNETT, B. FINKA, B. JANIĆIJEVIĆ, V. JOVANOVIĆ, M. F. LETHBRIDGE, J. MILIČIĆ, M. MIŠIGOJ, N. SMOLEJ NARANČIĆ, A. SUJOLDŽIĆ, L. SZIROVICZA, D. ŠIMIĆ, P. ŠIMUNOVIĆ: Anthropological Investiga- tions of the Eastern Adriatic, Book 1: Biological and Cultural Microdifferentiation among the village populations of the island of Korčula and the Pelješac peninsula. (HAD, Zagreb, 1987) (In Croat.). — 25. RUDAN, P., B. FINKA, B. JANIĆIJEVIĆ, V. JOVANOVIĆ, V. KUŠEC, J. MILIČIĆ, M. MIŠIGOJ- DURAKOVIĆ, D. F. ROBERTS, Lj. SCHMUTZER, N. SMOLEJ NARANČIĆ, A. SUJOLDŽIĆ, L. SZIROVICZA, D. ŠIMIĆ, P. ŠIMUNOVIĆ, S. M. SPOLJAR- VRZINA: Anthropological Investigations of the Eastern Adriatic, Book 2: Biological and Cultural Microdifferentiation among the village populations of the island of Hvar. (HAD, Zagreb, 1990) (In Croat.). — 26. RUDAN, P., L. A. BENNETT, B. FINKA, B. JANIĆIJEVIĆ, V. JOVANOVIĆ, V. KUŠEC, M. LETHBRIDGE-CEJKU, J. MILIČIĆ, Lj. SCHMUTZER, N. SMOLEJ NARANČIĆ, A. SUJOLDŽIĆ, D. ŠIMIĆ, P. ŠIMUNOVIĆ, S. M. ŠPOLJAR-VRŽINA: Anthropological Investigations of the Eastern Adriatic, Book 3: Biological and Cultural Microdifferentiation among the village populations of the island of Brač. (HAD, Zagreb, 1990) (In Croat). — 27. RUDAN, P., B. JANIĆIJEVIĆ, V. JOVANOVIĆ, J. MILIČIĆ, N. SMOLEJ-NARANČIĆ, A. SUJOLDŽIĆ, L. SZIROVICZA, T. ŠKARIĆ-JURIĆ, L. BARAĆ LAUC, T. LAUC, I. MARTINOVIĆ KLARIĆ, M. PERIČIĆ, D. RUDAN, I. RUDAN, Coll. Antropol. 28 Suppl. 2 (2004) 321. — 28. KUŠEC, V., D. ŠIMIĆ, A. CHAVENTRE, J. D. TOBIN, C. C. PLATO, P. RUDAN, Coll. Antropol., 12 (1988) 309. — 29. LOVASIĆ, I., T. ŠKARIĆ-JURIĆ, B. BUDISELIĆ, L. SZIROVICZA, Coll. Antropol., 22 (1998) 307. — 30. MARTINOVIĆ KLARIĆ, I., F. LOVASIĆ, B. BUDISELIĆ, T. ŠKARIĆ-JURIĆ, L. SZIROVICZA, A. CHAVENTRÉ, Coll. Antropol., 23 (1999) 91. — 31. BEHLULI, I., M. LETHBRIDGE-CEJKU, C. C. PLATO, P. RUDAN, I. RUDAN, W. A. STINI, J. D. TOBIN, Med. Jad., 21 (1991) 55. — 32. LETHBRIDGE-ČEJKU, M.: Osteoarthritis of the hands in a rural population (Anthropological research on the island of Brac, Croatia). (Ph.D. Thesis, University of Zagreb, Zagreb, 1995) (In Croat.). — 33. LETHBRIDGE-ČEJKU, M., C. C. PLATO, P. RUDAN, Am. J. Hum. Biol., 9 (1997) 136. — 34. ŠIMIĆ, D., A. CHAVENTRE, C. C. PLATO, J. D. TOBIN, P. RUDAN, Ann. Physiol. Anthrop., 11 (1992) 3. — 35. GINSBURG, E., T. ŠKARIĆ-JURIĆ, E. KOBILYANSKY, D. KARASIK, I. MALKIN, P. RUDAN, Am. J. Hum. Biol., 13 (2001) 398. — 36. WADDLE, D. M., R. R. SOKAL, P. RUDAN, Hum. Biol., 70 (1998) 845. — 37. MARTINOVIĆ KLARIĆ, I., Am. J. Hum. Biol., 12 (2000) 509. — 38. MARTINOVIĆ KLARIĆ, I., L. BARAĆ, D. BUKOVIĆ, I. FURAČ, G. GEBER, B. JANIĆIJEVIĆ, M. KUBAT, M. PERIČIĆ, B. VIDOVIĆ, P. RUDAN, Homo, 51 (2000) 141. — 39. MARTINOVIĆ, I., L. BARAĆ, I. FURAČ, B. JANIĆIJEVIĆ, M. KUBAT, M. PERIČIĆ, B. VIDOVIĆ, P. RUDAN, Hum. Biol., 71 (1999) 341. — 40. MARTINOVIĆ KLARIĆ, I., L. BARAĆ, D. BUKOVIĆ, I. FURAČ, G. GEBER, B. JANIĆIJEVIĆ, M. KUBAT, M. PERIČIĆ, B. VIDOVIĆ PUPIĆ, P. RUDAN, Ann. Hum. Biol., 28 (2001) 281. — 41. MARTINOVIĆ, I., S. MASTANA, B. JANIĆIJEVIĆ, V. JOVANOVIĆ, S. S. PAPIHA, D. F. ROBERTS, P. RUDAN, Ann. Hum. Biol., 25 (1998) 489. — 42. BARAĆ, L., M. PERIČIĆ, I. MARTINOVIĆ KLARIĆ, S. ROOTSI, B. JANIĆIJEVIĆ, T. KIVISILD, J. PARIK, I. RUDAN, R. VILLEMS, P. RUDAN, Eur. J. Hum. Genet., 11 (2003) 535. — 43. TOLK, H.-V., M. PERIČIĆ, L. BARAĆ, I. MARTINOVIĆ KLARIĆ, B. JANIĆIJEVIĆ, I. RUDAN, J. PARIK, R. VILLEMS, P. RUDAN, Coll. Antropol., 24 (2000) 267. — 44. TORRONI, A., H.-J. BANDELT, V. MACAULAY, M. RICHARD, M. CRUCIANI, C. RENGO, V. MARTINEZ-CABRERA, R. VILLEMS, T. KIVISILD, E. METSPALU, J. PARIK, H.-V. TOLK, K. TAMBETS, P. FORSTER, B. KARGER, P. FRANCALACCI, P. RUDAN, B. JANICIJEVIC, O. RICKARDS, M.-L. SAVONTAUS, K. HUOPONEN, V. LAITINEN, S. KOIVUMAKI, B. SYKES, E. HICKEY, A. NOVELLETTO, P. MORAL, D. SELLITTO, Am. J. Hum. Genet., 69 (2001) 844. — 45. MARTINOVIĆ, I., M. BAKRAN, A. CHAVENTRE, B. JANIĆIJEVIĆ, V. JOVANOVIĆ, N. SMOLEJ NARANČIĆ, A. KAŠTELAN, Z. GRUBIĆ, Z. ŽUNEC, D. F. ROBERTS, P. RUDAN, Hum. Biol., 69 (1997) 819. — 46. GRUBIĆ, Z., R. ŽUNEC, E. ČEČUK-JELIČIĆ, V. KERHIN- BRKLJAČIĆ, D. KAŠTELAN, L. BARAĆ, B. JANIĆIJEVIĆ, I. MARTINOVIĆ, M. PERIČIĆ, L. A. BENNETT, P. RUDAN, A. KAŠTELAN, Coll. An-

tropol., 22 (1998) 157. — 47. CAMBON-THOMSEN, A., E. SOMMER, A. SEVIN, A. CHAVENTRE, N. BOROT, E. OHAYON, P. RUDAN, Coll. Antropol., 13 (1989) 311. — 48. BOROT, N., J. M. DUGOUJON, B. JANIĆIJEVIĆ, P. RUDAN, A. CHAVENTRE, Coll. Antropol., 15 (1991) 247. — 49. ROBERTS, D. F., Z. M. NOOR, S. S. PAPIHA, P. RUDAN, Ann. Hum. Biol., 19 (1992) 539. — 50. JANIĆIJEVIĆ, B., M. BAKRAN, S. S. PAPIHA, A. CHAVENTRE, D. F. ROBERTS, Hum. Biol., 66 (1994) 991. — 51. JANIĆIJEVIĆ, B., Coll. Antropol., 12 (1988) 369. — 52. ARNAUD, J., N. BOROT, A. CHAVENTRE, B. JANIĆIJEVIĆ, A. E. SAMMARTINO, A. SUJOLDŽIĆ, P. RUDAN, R. JAMBOU, Coll. Antropol., 13 (1989) 281. — 53. KOPAJTIĆ, B., M. DUJMOVIĆ, Z. KOLACIO, V. KOGOJ-BAKIĆ, Coll. Antropol., 19 (1995) 365. — 54. KRŽIŠNIK, C., Z. KOLACIO, T. BATTELINO, M. BROWN, J. S. PARKS, Z. LARON, J. Endocr. Genet., 1 (1999) 9. — 55. ZERGOLLERN, L., Birth Defects Orig. Artic. Ser., 7 (1971) 28. — 56. BOHAČEK, N., Liječ. Vjesn., 86 (1964) 1412. — 57. MARIĆEVIĆ, A., Liječ. Vjesn., 117 (1995) 126. — 58. RUDAN, I., Hum. Biol., 73 (2001) 871. — 59. RUDAN, I., Hum. Biol., 71 (1999) 173. — 60. BAKIJA-KONSUO, A., A. BASTA-JUZBAŠIĆ, I. RUDAN, M. SITUM, M. NARDELLI-KOVAČIĆ, S. LEVANAT, J. FISCHER, D. HOHL, D. LONČARIĆ, S. SEIWERT, H. CAMPBELL, Dermatology, 205 (2002) 32. — 61. TURČINOV, D., R. KRISHNAMOORTHY, B. JANIĆIJEVIĆ, I. MARKOVIĆ, C. LAPOUMEROULIE, A. CHAVENTRE, P. RUDAN, Coll. Antropol., 24 (2000) 295. — 62. BOROT, N., J. ARNAUD, P. RUDAN, A. CHAVENTRE, J. SEVIN, Hum. Hered., 41 (1991) 309. — 63. TOLK, H.-V., L. BARAĆ, M. PERIČIĆ, I. MARTINOVIĆ KLARIĆ, B. JANIĆIJEVIĆ, H. CAMPBELL, I. RUDAN, T. KIVISILD, R. VILLEMS, P. RUDAN, Eur. J. Hum. Genet., 9 (2001) 717. — 64. SMOLEJ NARANČIĆ, N., I. RUDAN, J. Physiol. Anthropol. Appl. Human Sci., 20 (2001) 85. — 65. RUDAN, I., N. SMOLEJ NARANČIĆ, H. CAMPBELL, A. CAROTHERS, A. WRIGHT, B. JANIĆIJEVIĆ, P. RUDAN, Genetics, 163 (2003) 1011. — 66. RUDAN, I., D. RUDAN, H. CAMPBELL, A. CAROTHERS, A. WRIGHT, N. SMOLEJ-NARANČIĆ, B. JANIĆIJEVIĆ, L. JIN, R. CHAKRABORTY, R. DEKA, P. RUDAN, J. Med. Genet., 40 (2003) 925. — 67. RUDAN, I., D. RUDAN, H. CAMPBELL, Z. BILOGLAV, L. SIBBETT, B. JANIĆIJEVIĆ, N. SMOLEJ NARANČIĆ, P. RUDAN, Coll. Antropol., 26 (2002) 421. — 68. RUDAN, I., M. PADOVAN, D. RUDAN, H. CAMPBELL, Z. BILOGLAV, B. JANIĆIJEVIĆ, N. SMOLEJ NARANČIĆ, P. RUDAN, Coll. Antropol., 26 (2002) 11. — 69. LAUC, T., P. RUDAN, I. RUDAN, H. CAMPBELL, J. Orthod., 30 (2003) 301. — 70. FORENBAHER, S., Coll. Antropol., 26 (2002) 361. — 71. MALNAR, A., Coll. Antropol., 26 (2002) 411. — 72. ŠKREBLIN, L., L. ŠIMIČIĆ, A. SUJOLDŽIĆ, Coll. Antropol., 26 (2002) 333. — 73. SMOLEJ NARANČIĆ, N., Coll. Antropol., 23 (1999) 59. — 74. WRIGHT, S., Am. Naturalist, 56 (1922) 330. — 75. TAY, J. S., W. C. YIP, Ann. Hum. Genet., 48 (1984) 185. — 76. ROGULJIĆ, D., I. RUDAN, P. RUDAN, Am. J. Hum. Biol., 9 (1997) 595. — 77. FALCONER, D. S., T. F. C. MACKAY: Introduction to quantitative genetics, 4th Ed. (Longman, Harlow, UK, 1996). — 78. BARNETT, L., B. E. C. NORDIN, Clin. Radiol., 11 (1960) 166. — 79. ROTHMAN, K. J., S. GREENLAND: Modern epidemiology, 2nd edition. (Lippincott, Williams & Wilkins Publishers, 1998). — 80. KHOURY, M. J., B. H. COHEN, T. H. BEATY: Fundamentals of genetic epidemiology, 1st edition. (Oxford University Press, Oxford, 1993). — 81. INTERNATIONAL HUMAN GENOME SEQUENCING CONSORTIUM, Nature, 431 (2004) 931.

P. Rudan

Institute for Anthropological Research, Amruševa 8, 10000 Zagreb, Croatia
E-mail: pavao.rudan@inantro.hr

SRODIVANJE I SKLONOST OSTEOPOROZI U IZOLIRANIM OTOČNIM POPULACIJAMA HRVATSKE

SAŽETAK

Cilj ovog istraživanja bio je analizirati recesivnu genetsku komponentu vrijednosti kortikalnog indeksa druge metakarpalne kosti te procijenjenu prevalenciju osteoporoze (OP) u 14 sela hrvatskih otoka Brača, Hvara i Korčule s različitom razinom srodivanja. Slučajni uzorak uključio je 20–30% odraslog stanovništva tih sela i obuhvatio ukupno 1389 ispitanika. Prosječan koeficijent urođenosti (F) ispitanika svakog sela procijenjen

je Wrightovom »path« metodom (temelji se na informacijama pohranjenima u rodoslovlju). Morfometrija metakarpalnih kostiju učinjena je na rendgenskim snimkama šake i ručnog zgloba na obje ruke u svih ispitanika. Osteoporozu smo definirali kao vrijednosti kortikalnog indeksa manju od 2 standardne devijacije temeljenu na raspodjeli vrijednosti u ispitanika istog spola i dobi ispod 45 godina. Prosječne vrijednosti kortikalnog indeksa i prevalencije osteoporoze u svakom selu (standardizirane s obzirom na dob ispitanika te ponderirane s obzirom na veličinu uzorka) korelirane su s prosječnim koeficijentom urođenosti sela. Koeficijent korelacije (r) između F vrijednosti i CI bio je –0,28 u muškaraca (p=0,08) i –0,42 u žena (p=0,005), a između F vrijednosti i standardiziranih prevalencija OP iznosio je 0,32 u muškaraca (p<0,001) i 0,43 u žena (p<0,001). Ovi rezultati ukazuju na značajnu »depresiju srođivanja« kortikalnog indeksa kao pokazatelja smanjenja koštane mase, što upućuje na povećanu sklonost osteoporozi u starijoj životnoj dobi (OP) u srođenim zajednicama hrvatskih otoka.

Coll. Antropol. **26** (2002) 2: 421–428
UDC 572.026:616.89-008.454
Original scientific paper

Inbreeding and Learning Disability in Croatian Island Isolates

I. Rudan[1,2], D. Rudan[3], H. Campbell[2], Z. Biloglav[1], R. Urek[3], M. Padovan[4], L. Sibbett[2], B. Janićijević[5], N. Smolej Narančić[5] and P. Rudan[5]

[1] School of Public Health »A. Štampar«, School of Medicine, University of Zagreb, Zagreb, Croatia

[2] Department of Public Health Sciences, University Medical School, Edinburgh, UK

[3] General Hospital »Sveti Duh«, Zagreb, Croatia

[4] General Hospital »Dubrovnik«, Dubrovnik, Croatia

[5] Institute for Anthropological Research, Zagreb, Croatia

ABSTRACT

The aim of this study was to investigate the prevalence of learning disability (LD) in isolate populations with different inbreeding coefficients (F). Prevalence of LD and F were determined in 10 villages from five Croatian islands: Brač, Hvar, Korčula, Lastovo and Susak. For the purpose of this study, LD was defined as the inability to attend the public school system. As the elementary schools (grade 1–8) in the place of the study are both public and compulsory, the assessment of child's inability to attend the school is performed at the age of six. This is required by all children in the country based on standard set of tests of cognitive performance defined by the Ministry of Education and Culture of the Republic of Croatia. The average inbreeding coefficients in each village population (F) were estimated in a random sample of 20–30% adults in each of the 10 villages based on 4 ancestral generations and using Wright's path method. Prevalence of LD ranged from 0.43% to 2.47%, and the inbreeding coefficients ranged from 0.8% to 4.9%. The Pearson's correlation coefficient between F and LD prevalence was 0.80 ($p<0.01$). Although the relative risk per 5% inbreeding appeared very high (about 10), the absolute risk only increased from 0.18% to 1.77%. The genetic effect of inbreeding (GEI) was approximately 0.69% and the population-attributable fraction 76.6%. A review of the literature and the results of this study lead to a conclusion that a very large number of predominantly recessive genetic factors might mediate the genetic susceptibility to various forms of LD in these populations.

Received for publication September 12, 2002

Introduction

Despite the common belief that consanguineous unions are associated with increased risk of some form of learning disability (LD) in their children, a review of scientific evidence in support of this association[1–22] failed to identify rigorous evidence in its favor[23,24]. The origins of this belief certainly lie in a distant past, as this association has been widely mentioned in various historical works of literature in many cultures[25]. We were only able to identify 22 well-designed studies investigating the effect of inbreeding on cognitive performance. Only 2 of these were published after 1980, suggesting that this has been a topic avoided by the research community, possibly due to its sensitive nature. This is unfortunate since advances in the understanding of human disorders made possible in this post-genomic era provide realistic hope that the mechanisms of disease may be better understood and lead to new prevention and treatment strategies. We believed that an investigation into whether the observed increase in LD in inbred communities was due to numerous recessive genetic variants of small effect, or a small number of rare variants of large effect, or simply cultural or socio-economic bias would be a useful contribution to improving understanding of the disease mechanisms which underlie LD.

In this paper, we present one approach to the study of LD that aimed to study the relationship between inbreeding patterns and LD whilst attempting to correct for cultural and socio-economical bias. We further aimed to determine the relative and absolute risk of LD that might be attributable to inbreeding. The studied population included 10 isolate villages from the eastern Adriatic islands of Hvar, Brač, Korčula, Lastovo and Susak, in Croatia, a resource well characterized through a long-term multidisciplinary anthropologic and biomedical research[26–30].

Materials and Methods

Study design

The prevalence of learning disability was determined in 10 isolate villages on 5 different Croatian islands (Figure 1). These villages are characterized by reduced environmental variation and their inhabitants share very similar environmental factors (climate, nutrition, socio economic status, occupation, education, housing), as it has been demonstrated in previous studies[28,31]. In theory this should create a favorable setting for study since it should help limit socio-economic and cultural bias in the interpretation of the results.

Another favorable characteristic of these populations for our study is the diversity of the attitudes towards inbreeding[26,28]. This was influenced by geographic isolation, political privileges in the past and socio-cultural reasons and resulted in a range of inbreeding coefficients present at both individual and population level[26,27].

Previously conducted studies compared the prevalence of LD in an inbred cohort with non-inbred controls. This raises issues about the social and cultural comparability of controls and the possible clustering of a Mendelian disease (a single large effect gene) in the inbred cases. In contrast, this study investigated 10 populations with similar environment and culture but with a spectrum of inbreeding coefficients and quite different founding populations.

We hypothesized that if the study found comparable prevalence of LD in all 10 populations, this would not support any inbreeding effect and the LD prevalence will be assumed to be determined mainly by factors related to environment. However, if we found a consistently positive correlation between inbreeding levels and LD prevalence across 10 villages, this would clearly point to an effect of inbreeding. A further advantage of having 10 distinct

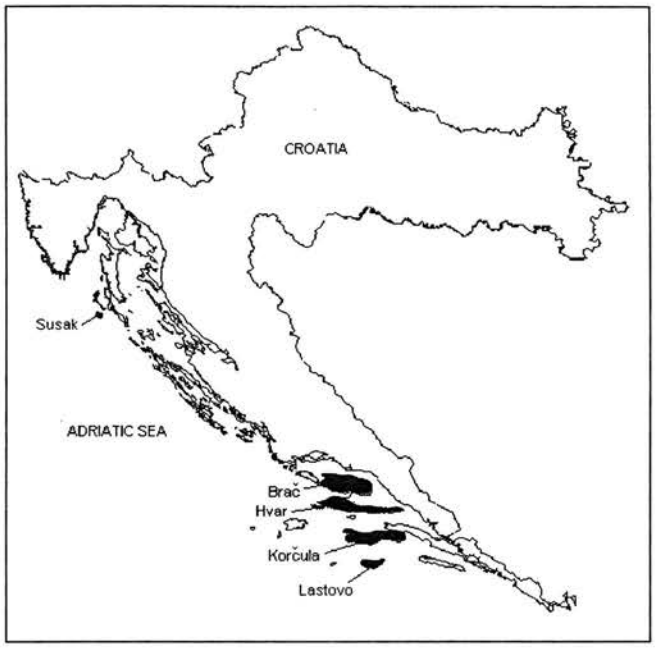


Fig. 1. Geographic location of the investigated islands of Brač, Hvar, Korčula, Lastovo and Susak.

populations under investigation is to rule out the possibility of a rare Mendelian disease clustering, as it is unlikely that the same rare variant would be present in all 10 populations having very different founding populations and ethnohistory.

We further hypothesized that if a modest increase in sharing of genes identical by descent (e.g. an increase in inbreeding coefficient at the level of entire population from 0% to 5%) led to a significant change in prevalence of LD across several isolate populations that share very similar environmental effects, then this would be most consistent with the model including a very large number of genomic loci influencing the disease, as Morton has suggested in his review of the problem[19].

Estimation of the prevalence of learning disability

For the purpose of this study, learning disability was defined as the inability to attend the public school system. As the elementary schools (grade 1–8) in the place of the study are both public and compulsory, the assessment of child's ability to attend the school is performed at the age of six. The assessment is based on standard set of tests, as required by the Ministry of Education and Culture of the Republic of Croatia[32]. These tests include: (a) perception test, test of point linkage, test of knowing facts, drawing test and numerical test; (b) intelligence test based on drawing a human image; (c) »Bender Gestalt« test; (d) Raven's progressive colored matrices[32]. Data on the individuals unable to attend school were retrieved from local general practitioners and were considered to be complete. The prevalence of LD was calculated as the proportion of these individuals in the total population of each village (as of January 2001). Ethical approval for this study was obtained from the Ethics Committee of

the Institute for Anthropological Research, Zagreb, Croatia.

Computation of inbreeding coefficients (F)

Genetic characterization of these villages included the computation of average inbreeding coefficient of each village based on reconstruction of genealogies of a sample of examinees which formed 20–30% of adult village population. The pedigree information on 2–3 ancestral generations was recorded for each examinee during the fieldwork between 1979–1981 performed by the Institute of Anthropological Research in Zagreb. It was later expanded during 1997–2000 through insight into the parish registries stored in local churches to allow the completion of the information on 4 ancestral generations in each examinee. The individual inbreeding coefficients (*F*) were then computed according to Wright's path method[33]:

$$F = \Sigma_{(1\rightarrow c)}(1/2)^{(n_i+m_i+1)}$$

where *m* and *n* refer to the number of paths from a common ancestor, and *c* refers to the number of common ancestors. The genealogical inbreeding coefficient for each village was then computed as the average of all individual F values.

Statistical data analysis

Linear regression analysis of LD prevalence on F was performed using the data from all ten villages. The corresponding Pearson's coefficient of correlation (r) and the regression coefficient (b) were determined using the SYSTAT 7.0 software.

The observed prevalence of LD in each of the studied populations was considered to approximate reasonably well the absolute risk of LD in pre-school age in each population. The relative risk for each unit increase of 5% inbreeding was inferred from the slope of the linear regression curve as the ratio of the expected LD prevalence at the points of F = 5% and F = 0%.

As pointed out by Freire-Maia[34], in certain instances the absolute and relative risk measures can be artificially low. An index – called »the genetic effects of inbreeding« (GEI) – was suggested as an alternative[34], and is calculated as:

$$GEI = (Pi–Po) / (1–Po),$$

where Pi is a probability of the event (in this case LD) in an inbred person (in this case a village with an average F greater than 3%), and Po is the probability of the event in a non-inbred person (in this case a village with an average F less than 1%).

The population-attributable fraction (PAF) was calculated by logistic regression, noting each village's probability LD prevalence value if their F was set equal to zero. The sum of all such probabilities, is an estimate of the LD prevalence in the absence of inbreeding. Then:

$$PAF = 1 - Psum / Npop,$$

where Npop is the total population size[35].

Results

Table 1 presents the data on studied villages coded from A to J and their respective total populations, the average coefficients of inbreeding (computed as above), the number of cases of LD and the prevalence of LD in each village. The inbreeding coefficients in these villages ranged from 0.8% to 4.9%, and the prevalence of LD from 0.43% to 2.47%.

Figure 2 presents the linear regression between F and LD and the corresponding Pearson's correlation coefficient, which was 0.80 ($p<0.01$). Although the relative risk per unit increase of 5% inbreeding appeared to be quite high (about 10), the absolute risk (defined as prevalence at the intercept of the regression line with F=0% and F=5%) only increased from 0.18% to 1.77%. The genetic effect of inbreeding (GEI) was 0.69% which relates to the difference in ex-

TABLE 1
TOTAL POPULATION OF 10 STUDIED VILLAGES (AS OF JANUARY 2001), AVERAGE POPULATION INBREEDING COEFFICIENT (F) DETERMINED FROM GENEALOGIES, NUMBER OF LEARNING DISABILITY (LD) CASES AND THE PREVALENCE OF LEARNING DISABILITY

Village (island)	N (Population)	F (Genealogical)	LD Cases	LD Prevalence
A (Hvar)	153	0.049	2	1.31%
B (Susak)	81	0.047	2	2.47%
C (Korčula)	326	0.044	6	1.84%
D (Korčula)	464	0.032	2	0.43%
E (Korčula)	290	0.027	3	1.03%
F (Brač)	214	0.013	1	0.47%
G (Korčula)	354	0.012	3	0.85%
H (Lastovo)	899	0.011	5	0.56%
I (Korčula)	866	0.008	4	0.46%
J (Hvar)	375	0.008	2	0.53%

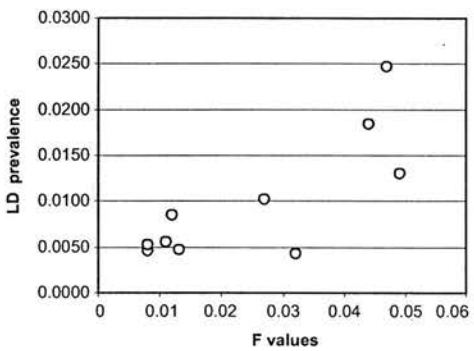


Fig. 2. Plot of the values of prevalence of learning disability (LD) against average coefficients of inbreeding (F) in ten studied villages.

TABLE 2
PEARSON'S COEFFICIENT OF CORRELATION BETWEEN INBREEDING COEFFICIENT AND PREVALENCE OF LEARNING DISABILITY, REGRESSION COEFFICIENT, ESTIMATES OF RELATIVE RISK, GENETIC EFFECT OF INBREEDING (GEI) AND POPULATION ATTRIBUTABLE FRACTION (PAF)

Measure	Value
Pearson's coefficient of correlation (r)	0.80
Regression coefficient (b)	0.33
Relative risk (F=5% vs. F=0%)	10.23
GEI (F3% vs. F1%)	0.69%
PAF	76.6%

pected prevalence of LD when the villages with F greater than 3% and less than 1% are compared. The population attributable fraction of cases due to inbreeding in all ten populations was high and it amounted to 76.6% (Table 2).

Discussion

Early work on the relationship between inbreeding and cognitive performance was reported by Penrose in 1938[1], but the most influential studies were carried out in 1970's by Schull, Neel and their coworkers in Japan[4,7,8,10,11,15], Danielov in Russia[12,13,17] and Costeff in Israel[9,16]. These three groups published about a half of all the available studies in the world literature to date. The results of these studies were in very close agreement: they all found statistically significant effects of the inbreeding on cognitive performance (measured as a quantitative trait) or the prevalence of learning disability (measured as a qualitative trait). Although the associated relative risks were also rather large, it still needs to be

understood that in absolute terms any risk attributable to inbreeding was still small. The average regression coefficient of inbreeding on IQ found in these studies, after weighting by the reciprocal variance, was –44.0 (±12.3)[15]. In addition, the risk of some form of LD in offspring of non-inbred marriages was estimated to about 1.2%, and in the offspring of first-cousin marriages to about 6.2%[19]. Although the risk in the latter is reportedly increased 5 times, this still means that 94 of 100 children born in such unions did not suffer of any form of LD.

In our study, learning disability was measured as a qualitative trait. The relative risk (RR) per unit increase 5% inbreeding seemed to be about 10, although this was based on the intercept of our linear regression line with F = 0 as we only studied inbred populations. However, when this is replaced by the prevalence of LD in general Croatian population, kindly provided by the Croatian Ministry of Education and Culture, the more realistic estimate of absolute risk in non-inbred population of about 4.0 is observed. This is in close agreement of RR = 5 per 6.25% inbreeding, reported in previous studies. The basal prevalence of LD in non-inbred populations in our study was somewhat smaller than in other reports or in general Croatian population (0.43–0.56% in comparison to about 1.2%), but this can be explained by random fluctuations in relatively small populations of the villages in this study. Thus, if we accept the conclusion that an absolute increase in inbreeding of about 4–5% (from F = 0.005 to 0.05) could be responsible for the observed 4-fold increase in prevalence of LD, the central question becomes what does it tell us about the genetic mechanisms underlying this complex syndrome. An effect of inbreeding implies that the susceptibility is most probably controlled, at least partly, by recessive genetic variants. In addition, if we accept that the total number of human genes is between 30,000 and 40,000, then an absolute increase in inbreeding of 5% would correspond to having about 1,750 random genes across the genome identical by descent. If this unrecombined homozygosity in only 5% of genes could lead to a measurable effect in the prevalence of LD, then there are two main mechanisms that could explain it: (a) this brings together some rare major effect genes in a simple Mendelian fashion, or (b) the genes controlling this trait are of small effect but very numerous, scattered across the genome. The design of this study provides evidence against the first explanation. Major effect genes arise after mutations that are considered to be extremely rare, as the probability of random mutation causing a small effect is much greater. Therefore, even if such mutations were present in some of the studied villages, it is extremely unlikely that similar effects of inbreeding would be observed across several villages, as our results indicate. In addition, under this assumption the differences between inbred and non-inbred individuals would normally be much larger than was the case in our study, where the differences were very small but consistent across many distinct populations. We conclude, therefore, that it is more likely that the genetic susceptibility to learning disability, a highly heterogeneous group of syndromes, is at least in part controlled by a large number of recessive genes.

Acknowledgements

This work was supported by the Wellcome Trust (IRDA) grant to H.C. and I.R., the Croatian Ministry of Science and Technology grants to I.R. (0108330), to N.S.-N. (0196001) and to P.R. (0196005) and the joint British Council and CMST grant ALIS 054 to H.C. and I.R. I.R. was supported by funds from the UK Medical Research Council, the University of Edinburgh and the Overseas Research Scheme.

REFERENCES

1. PENROSE, L. S., J. Ment. Sci., 84 (1938) 693. — 2. BOOK, J. A., Ann. Hum. Genet., 21 (1957) 1957. — 3. SLATIS, H. M., R. E. HOENE, Am. J. Hum. Genet., 13 (1961) 28. — 4. SCHULL, W. J., J. V. NEEL: The effects of inbreeding on Japanese children. (Harper and Row, New York, 1965). — 5. ROBERTS, D. F., Br. Med. J., 4 (1967) 336. — 6. SCHREIDER, E., Am. J. Phys. Anthropol., 30 (1969) 215. — 7. NEEL, J. V., W. J. SCHULL, M. YAMAMOTO, S. UCHIDA, T. YANASE, N. FUJIKI, Am. J. Hum. Genet., 22 (1970) 263. — 8. TANAKA, K., Jinrui Idenkagu Zasshi, 16 (1972) 170. — 9. COSTEFF, H., B. E. COHEN, L. WELLER, Acta Paediatr. Scand., 61 (1972) 452. — 10. KUDO, A., K. ITO, K. TANAKA, Jpn. J. Hum. Genet., 17 (1972) 231. — 11. NAKAZAWA, S., T. SHIOKAWA, T. ISHIKAWA, K. TANAKA, A. KATUI, Jinrui Idenkagu Zasshi, 17 (1972) 219. — 12. DANIELOV, M. B., Genetika, 11 (1975) 121. — 13. DANIELOV, M. B., Genetika, 12 (1976) 130. — 14. BASHI, J., Nature, 266 (1977) 440. — 15. TANAKA, K., Jinrui Idenkagu Zasshi, 22 (1977) 55. — 16. COSTEFF, H., B. E. COHEN, L. WELLER, D. RAHMAN, Am. J. Hum. Genet., 29 (1977) 339. — 17. DANIELOV, M. B., Zh. Nevropatol. Psikhiatr. Im. S. S. Korsakova, 78 (1978) 898. — 18. OLIVIER, G., G. DEVIGNE, C. R. Acad. Sci. Hebd. Seances D, 286 (1978) 611. — 19. MORTON, N. E., Proc. Natl. Acad. Sci. U.S.A., 75 (1978) 3906. — 20. KAMIN, L. J., Psychol. Bull., 87 (1980) 469. — 21. JANCAR, J., S. J. JOHNSTON, J. Ment. Defic. Res., 34 (1990) 483. — 22. BADARUDDOZA, B., M. AFZAL, Behav. Genet., 23 (1993) 343. — 23. BITTLES, A. H., J. V. NEEL, Nat. Genet., 8 (1994) 117. — 24. BITTLES, A. H., W. M. MASON, J., GREENE, N. A. RAO, Science, 252 (1991) 789. — 25. FARRALL, M., Nat. Genet., 5 (1993) 107. — 26. RUDAN, P., D. ŠIMIĆ, N. SMOLEJ-NARANČIĆ, L. A. BENNETT, B. JANIĆIJEVIĆ, V. JOVANOVIĆ, M. F. LETHBRIDGE, J. MILIČIĆ, D. F. ROBERTS, A. SUJOLDŽIĆ, L. SZIROVICZA, Am. J. Phys. Anthropol., 74 (1987) 417. — 27. RUDAN, P., A. SUJOLDŽIĆ, D. ŠIMIĆ, L. A. BENNETT, D. F. ROBERTS, In: ROBERTS, D. F., N. FUJIKI, K. TORIZUKA (Eds.): Isolation, migration and health. (Cambridge University Press, Cambridge, 1992). — 28. RUDAN, I., H. CAMPBELL, P. RUDAN, Coll. Antropol., 23 (1999) 531. — 29. BENNETT, L. A., J. L. ANGEL, D. F. ROBERTS, P. RUDAN, Coll. Antropol., 7 (1983) 195. — 30. WADDLE, D. M., R. SOKAL, P. RUDAN, Hum. Biol., 70 (1998) 845. — 31. SMOLEJ-NARANČIĆ, N., Coll. Antropol., 23 (1999) 59. — 32. CROATIAN MINISTRY OF HEALTH AND CROATIAN MINISTRY OF EDUCATION AND CULTURE, Narodne novine, 13 (1991) 425. — 33. WRIGHT, S., Am. Naturalist, 56 (1922) 330. — 34. FREIRE-MAIA, N., Am. J. Med. Genet., 35 (1990) 118. — 35. ROTHMAN, K. J., S. GREENLAND (Eds.), Modern epidemiology. 2nd ed. (Lippincott-Raven, Philadelphia, 1998).

I. Rudan

Department of Public Health Sciences, University of Edinburgh Medical School, Edinburgh, EH8 9AG, Scotland, UK

SRODIVANJE I POTEŠKOĆE U UČENJU MEĐU STANOVNIŠTVOM HRVATSKIH OTOKA

SAŽETAK

Cilj ovoga istraživanja je odrediti prevalenciju poteškoća u učenju (LD) u izoliranim otočnim populacijama s različitim vrijednostima koeficijenta urođenosti (F). Prevalencija poteškoća u učenju i vrijednosti F određene su u deset sela s pet različitih hrvatskih otoka: Brača, Hvara, Korčule, Lastova i Suska. Za potrebe ovoga istraživanja, LD je definirana kao nemogućnost pohađanja sustava osnovnog obrazovanja. Kako su osnovne škole (1.–8. razred) u mjestu istraživanja državne i obvezne, procjena nemogućnosti svakog djeteta da pohađa osnovnu školu vrši se u dobi od 6 godina. Testiranje je obvezno za svu djecu a temelji se na skupu testova kojeg propisuje Ministarstvo prosvjete. Prosječni koeficijent urođenosti u svakom selu izračunat je na temelju genealoških podataka 20–30% odraslog stanovništva upotrebom Wrightove »path«

metode. Prevalencija LD kretala se između 0.43% i 2.47%, a vrijednost koeficijenata urođenosti između 0.8% i 4.9%. Pearsonov koeficijent korelacije između F i prevalencije LD bio je 0.80 ($p<0.01$). Iako je relativni rizik za svakih 5% urođenosti bio velik (oko 10), apsolutni rizik pritom raste s procijenjenih 0.18% na samo 1.77%. Genetički učinak srođivanja procijenjen je na 0.69% a pripisivi populacijski udio na 76.6%. Pregled raspoložive literature i rezultati ovoga istraživanja ukazuju na zaključak da bi vrlo velik broj pretežno recesivnih genetskih čimbenika mogao biti odgovoran za genetsku sklonost različitim tipovima LD među stanovništvom.

polygenic basis for late-onset disease

ın Wright[1], Brian Charlesworth[2], Igor Rudan[3,4], Andrew Carothers[1] and ry Campbell[3]

Human Genetics Unit, Western General Hospital, Crewe Road, Edinburgh EH4 2XU, UK
itute of Cell, Animal and Population Biology, King's Buildings, University of Edinburgh, Edinburgh EH9 3JT, UK
artment of Public Health Sciences, Teviot Place, University of Edinburgh, Edinburgh EH8 9AG, UK
artment of Epidemiology, University Medical School, Rockefellerova 4, Zagreb, Croatia

biological basis of late-onset disease has been ıed by genetic factors subject to varying degrees of utionary constraint. Late-onset traits are not only e sensitive to environmental variation, owing to the kdown of homeostatic mechanisms, but they also v higher levels of genetic variation than traits ctly influencing reproductive fitness. The origin and re of this variation suggests that current strategies poorly suited to identifying genes involved in many plex diseases.

ıjor focus of current interest lies in the genetic vari- ı underlying susceptibility to common, late-onset an diseases such as heart disease, diabetes and er. These diseases result from the cumulative break- ı of many quantitatively varying physiological sys- over the course of decades of life. They are orders of nitude more common than individual mendelian ders and are typically more prevalent in post- ɔductive life, which means that they may be less ect to SELECTIVE CONSTRAINTS (see Glossary). The se mechanisms maintaining genetic variation in such s are poorly understood, but three broad categories dentifiable [1]. First, variants that are deleterious in early and later life, which are therefore efficiently ened by natural selection and held at low population ıencies. Second, variants that are selectively neutral ırly life but show late deleterious effects; this means they are subject only to weak selection and can reach er frequencies. Third, variants that are favourable in life but deleterious later on; these can be maintained election at intermediate frequencies. Strategies for ifying disease susceptibility genes depend both on the ıce of common and rare variants maintained in the lation, and on whether these occur at a limited OGENIC) or a large (POLYGENIC) number of loci. In this le, evolutionary and population genetic arguments ısed to examine these issues and to suggest that ntly favoured strategies could be poorly suited to ifying disease susceptibility genes.

ıe strategy assumes that most disease susceptibility nts are common in the population (frequency >0.01) COMMON DISEASE/COMMON VARIANT (CD/CV) HYPOTH- [2]. This proposes that individuals with disease have an excess of common susceptibility alleles, and that these are potentially detectable in large-scale patient–control association studies. However, if late-onset diseases are due to large numbers of rare variants at many loci – the COMMON DISEASE/RARE VARIANT (CD/RV) HYPOTHESIS – this strategy would fail and the contribution of most individual variants would be too small to further our understanding of disease [3]. To evaluate these issues, we first examine our current knowledge of human genetic diversity.

Hidden genetic diversity

The human population is both evolutionarily young and genetically uniform, with less diversity than most other species, including other primates [4]. The most abundant differences between individuals are single nucleotide POLYMORPHISMS (SNPs), which account for most of the observed variability in typical sequence surveys [5]. The great majority of SNPs occur outside coding regions and their distribution is broadly consistent with SELECTIVE NEUTRALITY [6]. There are ~10 million predicted SNPs with allele frequencies above 0.01 [7]. Under the CD/CV hypothesis, these provide the major genetic substrate for common diseases. However, this picture may give a misleading impression of the genetic variation underlying the emergent diseases of modern civilizations.

The principal reason is that the vast majority of DNA sequence variants, including most of those with functional effects, are expected to be rare [8]. Genetic theory predicts that the distribution of neutral sites is heavily skewed towards low-frequency variants with as many below a frequency of 0.01 as above it [9]. But the proportion of rare variants is even higher for two reasons. First, most mutations with phenotypic effects are deleterious [10], so that their frequency is reduced by selection. Second, the human population has been expanding, generating large numbers of rare alleles by mutation [11] (Fig. 1). The overall pattern is therefore one of relatively few common SNPs and many individually rare single nucleotide variants.

The majority of disease-causing alleles in early-onset mendelian disorders are recent, diverse and rare, resulting in extreme allelic heterogeneity. This is expected for deleterious alleles exposed to early selection, but is also found in later-onset diseases, including familial forms of cancer, coronary artery disease and Alzheimer dementia [12]. For example,

responding author: Alan Wright (alan.wright@hgu.mrc.ac.uk).

js.trends.com 0168-9525/03/$ - see front matter PII: S0168-9525(02)00033-1

sary

onistic pleiotropy: see Trade-off model.
cing selection: selection that maintains more than one allele in the ation at intermediate frequencies.
non disease/common variant (CD/CV) hypothesis: susceptibility to on disease results from a small number of common polymorphic ts at one or more loci.
non disease/fixed variant (CD/FV) hypothesis: susceptibility to common e in a given population results from invariant sites at one or more loci. can differ between populations and contribute to differences in disease otibility.
non disease/rare variant (CD/RV) hypothesis: susceptibility to common e results from numerous rare variants at many loci.
ional dominance: dominance is directional when the value of the zygous effect (h) deviates from the expected intermediate value in zygotes (e.g. $h < 0.5$ for most loci causing inbreeding depression).
nance: nonadditive interactions between alleles at the same locus, which ontinuously from complete recessivity (heterozygous effect, h, is zero) gh additivity ($h = 0.5$) to complete dominance ($h = 1$) (Box 1).
ve population size (N_e): the number of individuals in a population outing genes to succeeding generations, which predicts the rate of c drift.
ne concordant (discordant) sib pairs: pairs of siblings that are positively ated (concordant) or negatively correlated (discordant) for a trait.
s-related trait: a trait for which a change in value influences reproductive .
on: a state in which all members of a population are homozygous for a allele (which is then said to be fixed, with an allele frequency of 1).
ic drift: random fluctuations in gene frequency arising from a finite ve population size (N_e).
ne-wide mutation rate (U): the mean number of new deleterious mutant arising per individual each generation.
type: a combination of linked variants on a single chromosome.
cal-by-descent: alleles or genomic segments that are identical in one or ndividuals as a result of inheritance from a common ancestor.
ding coefficient (F): the probability that both copies of an allele are ed from a common ancestor (identical-by-descent).
ding depression: the detrimental effects of inbreeding, typically causing uction in the means of fitness-related traits, as a result of increased zygosity.
ding load: the proportional reduction in the value of a fitness-related ssociated with a unit increase in the inbreeding coefficient.
ion accumulation model: a genetic model of senescence in which rious alleles with effects restricted to later life reach higher frequencies in opulation than ones acting at an earlier age, assuming mutation–on balance.
ional target: the proportion of the genome capable of influencing a trait sult of *de novo* mutations.
ional variance: the genetic variance in a trait attributable to alleles ined by mutation in opposition to selection.
on–selection balance: a state of equilibrium between the input of c variants into a population by mutation and their elimination by natural on.
l alleles: see Selective neutrality.
enic: determined by a small number of genes of moderate effect.
opy: an effect of a genetic variant on more than one trait.
nic: determined by many genes of small effect.
orphism: a variant allele with a frequency greater than 0.01.
itative trait locus (QTL): any gene of small effect that contributes to tative variation in a trait.
ion coefficient: the reduction in fitness of a given genotype, measured e to the fitness of a standard genotype.
ive constraints: the elimination of variants from a population as a result ural selection.
ive disadvantage: see Selection coefficient.
ive neutrality: alleles with no effect on reproductive fitness.
cence: the decline with age in age-specific survival or other components roductive fitness.
ng genetic variation: the naturally occurring genetic variation within a opulation.
off (antagonistic pleiotropy) model: a genetic model of senescence in alleles show opposite pleiotropic effects on fitness-related traits in early er life.

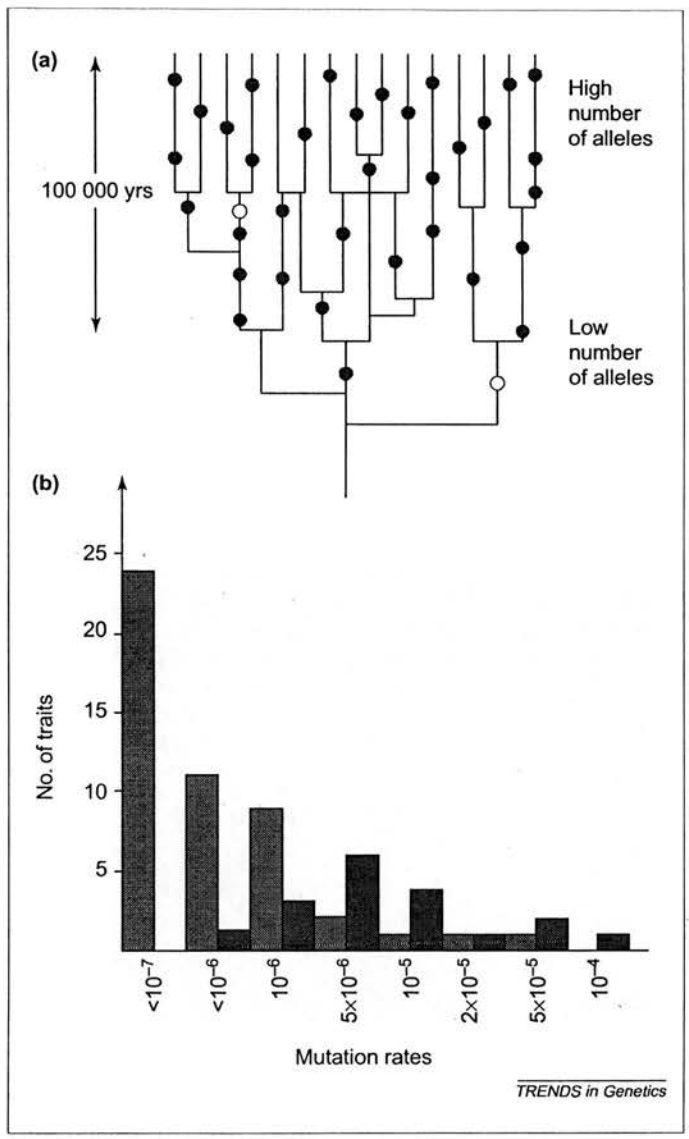
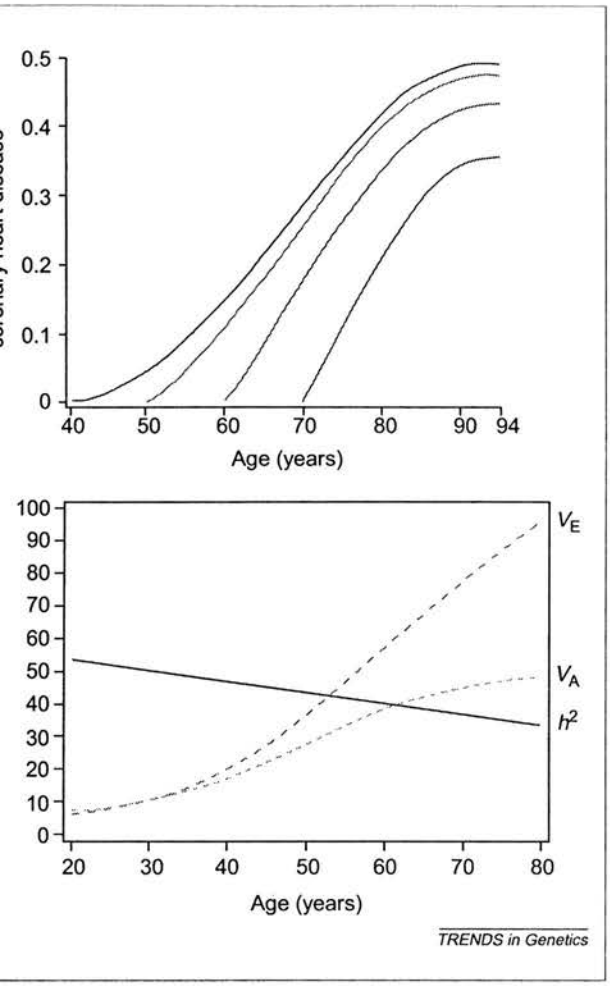


Fig. 1. (a) Diagram of human population expansion illustrating (1) the large number of 'young' mutations (filled circles) compared with polymorphic variants (open circles), most of which are ancient (>100 000 years old) but some of which might have become common within recent times as a result of a selective advantage; and (2) the low number of alleles in human founder populations and high number of alleles in large modern populations [11]. (b) Mutation rates and disease. Mutation rates for human monogenic diseases vary from $<10^{-7}$ to 10^{-4} per locus per generation. There appear to be more loci with low than with high mutation rates resulting in disease. High mutation rates in genes causing Duchenne muscular dystrophy (5×10^{-5} to 10×10^{-5}) and neurofibromatosis type 1 (1×10^{-4}) account for their high incidence (at mutation–selection balance). Similarly, genes with high mutation rates are predicted to contribute disproportionately to the genetic variance underlying common diseases [62]. The red boxes indicate accurately estimated mutation rates for more common diseases, the blue boxes indicate rates for rarer diseases, which are only known within an order of magnitude. The data are taken, with permission, from [63,64] (© Springer, 1986).

according to the Breast Cancer Information Core (BIC) database (http://research.nhgri.nih.gov/bic) the familial breast cancer genes, *BRCA1* and *BRCA 2*, show at least 1200 and 1400 different mutations of large effect, each of which are rare, except in some founder populations. This is despite the fact that well over half of these tumours occur post-reproductively, so that causal alleles might reach higher frequencies. Most of these mutations are likely to be either of recent origin or to show direct or indirect deleterious effects on fitness. By contrast, out of the five



) Cumulative lifetime risk of coronary artery disease in males [37], which almost one in two and shows a late-age levelling in disease risk, consist- some of the genetic models discussed. (b) Heritability changes with dis- e of onset. Possible changes in the additive component of genetic variance nge) in relation to the environmental variance (V_E, blue) and heratibility ple) with disease age of onset. Heritability is not an ideal measure of gen- ance in this context, because of the increase in V_E with age [1,27].

well-established common SNPs within *BRCA1* and *2* coding regions (coding SNPs or cSNPs), only one een shown to exert a marginal (1.3-fold) increase in cancer risk [13]. A small increase in risk in many e might account for a large fraction of cases. But this so if such effects occur within highly interactive c networks, with many other variants of similar or ite effect at varying frequencies in different popu- s, as expected under a CD/RV model. The pattern of ands of recent and rare mutations, many with large , and a small number of ancient cSNPs is predictable ut here it is argued that most cSNPs are common ely because they have little or no functional effect on disease or on reproductive fitness.

oding SNPs (comprising ~1.5% of all SNPs) are more to influence physiological function (hence disease) noncoding ones, are they any less common? The of cSNPs that change an amino acid and are also ted on structural grounds to be deleterious, occur at cantly lower frequencies than other SNPs, suggesting that they are indeed selected against [14,15]. Analyses of sequence divergence between humans and primates suggest that ~20% of all cSNPs are selectively neutral, most of which are common; of those predicted to be deleterious, over 80% are likely to be at frequencies below 0.01 (i.e. not truly polymorphic) [14].

In summary, the genetic variants that are most readily identified and studied in humans are SNPs, but most of these appear to have little or no effect either on reproductive fitness or on any sort of function. By contrast, the majority of deleterious variants, which are of most potential relevance to disease, are rare and accordingly difficult to study.

Genetic variation in late-onset traits

Is the pattern of genetic diversity likely to be different for variants influencing late-onset diseases? Unravelling the genetics of complex traits often requires indirect inferences about what are believed to be the many genes influencing them. Such 'polygenic' effects are thought to be too small and numerous to be measured individually, so their effects are measured collectively by partitioning the phenotypic variance into genetic and environmental components (Box 1).

How these components of genetic variance differ for late- versus early-onset traits has been examined in some detail theoretically. The intensity of selection on a gene with a late effect on fitness declines with the age at which it is expressed [16–19]. This implies that variants in such genes could reach higher frequencies, which would favour the CD/CV hypothesis. The 'mutation accumulation' (MA) model [17,19–21] extends this idea by suggesting that deleterious alleles with late effects accumulate in the genome, contributing to SENESCENCE and, by extension, to genetically influenced diseases that contribute to it. If these alleles have deleterious effects during reproductive life, they are still expected to be maintained at low frequencies, despite the diminishing force of selection with age. The 'TRADE-OFF' (TO) (or ANTAGONISTIC PLEIOTROPY) MODEL [18,19] provides an alternative, in which late-acting deleterious alleles can spread and even become universal in the population, if they also have favourable effects at an early age. Most genes are expressed before the end of reproductive life, and so are subject to selective scrutiny, but many show effects on different traits (PLEIOTROPY) at different times, with variable effects on fitness [19].

The expected higher frequencies of deleterious alleles with late-age effects are accompanied by increases in the components of genetic variance, because alleles with intermediate frequencies contribute more to these than do rare alleles [20,21]. Under both MA and TO models, the genetic variance components resulting from additive (V_A) and DOMINANCE (V_D) effects (Box 1) are expected to be larger for late- than for early-onset traits influencing fitness [20]. A late-age levelling in the rates of genetically influenced diseases, such as cancer, diabetes and cardiovascular disease, is also predicted by these models, much as observed (Fig. 2) [21,22]. Are the models supported by experimental data?

In *Drosophila melanogaster*, an increase in both V_A and V_D has been observed for several late-onset traits [20,23,24], suggesting that allele frequencies do indeed

x 1. Variance components and inbreeding effects

e-onset diseases can be considered to result when a threshold of antitatively varying risk or liability is exceeded [a]. Liability results m the net effect of many quantitative traits (QT), which are influenced genes and environment, often with small individual effects on risk, d hence are difficult to identify. These genetic effects can, however, be scribed collectively by analysing the components of genetic variance, ich are estimated from the resemblance between relatives for ease or QT. The total genetic variance (V_G) of a complex trait can partitioned into its components [a]:

Additive genetic variance (V_A): the component of variance due to genetic effects that are directly transmissible from parent to offspring, and which are the main causes of resemblance between relatives.

Dominance variance (V_D): the component of variance due to interactions (departures from additivity of effects) between alleles at the same locus, such as partial or complete dominance or recessivity.

Epistatic variance (V_I): the component of variance due to interactions between alleles at different loci.

The total phenotypic variance (V_P) in a trait is the sum of V_G and any ngenetic (environmental) effects (V_E), together with effects of eractions between genotype and environment (V_{GE}). The (narrow-nse) heritability is the ratio of V_A to V_P.

These components can be estimated from correlations between atives, such as parents and offspring or full-sibs. In practice, it is icult to separate V_D and V_I, and they are often treated as a single nponent of nonadditive variance. Late-onset traits are predicted to ow higher values of V_A and V_D [b,c].

Another source of information on genetic variation influencing ease is to measure the effects of inbreeding. Inbreeding can ntribute to disease [d,e] and to inbreeding depression [f]. This results m increased homozygosity of trait alleles, which either show essive effects on the trait acting in the same direction (DIRECTIONAL MINANCE; see Glossary) or show heterozygous advantage. *B*, the gative of the regression coefficient of trait mean on inbreeding efficient *F*, provides a useful summary statistic for the genetic damage t would occur if all deleterious recessives were made homozygous = 1); it is often called the INBREEDING LOAD (see Glossary) [a,f].

f survival to adulthood is measured on a scale of natural logarithms, *B* ovides an estimate of the number of deleterious mutations causing a netic death (lethal equivalents). In one human study, *B* was 0.7 per mete [e], suggesting that each (diploid) individual is heterozygous for lethal equivalents affecting juvenile mortality.

The value of *B* for lethal mutations is similar in a variety of species. cessive lethals contribute about half the inbreeding load for mortality *Drosophila* up to the adult stage, the rest coming from mutations with minor effects (detrimentals) [f]. A typical fly carries one lethal mutation in the heterozygous state. A recent study of two fish species suggests that, despite their greater genome size, a similar value applies to vertebrates [g].

Theory shows that the value of *B* due to deleterious mutations depends only on the genome-wide mutation rate and the dominance of individual mutations [f,h]. If all mutations have the same effects, the value of *B* for a disease-related trait that is positively correlated with fitness is:

$$B = U\alpha\{(1/h) - 2\}$$

where *U* is the mutation rate per diploid individual to deleterious alleles affecting the trait; *h* is the extent to which fitness is reduced in a heterozygous mutation, relative to its effect in homozygotes; and α is a constant of proportionality relating the effect of a mutation on the trait to its effect on fitness.

Relating the measurement of variance components and inbreeding effects to the predictions of models of the maintenance of genetic variation provides an important means of testing the models [f,h,i]. Variance component analysis is also used in quantitative trait locus (QTL) mapping.

References

a Falconer, D.S. and Mackay, T.F.C. (1996) *Introduction to Quantitative Genetics*, 4th edn, Longman

b Partridge, L. and Gems, D. (2002) Mechanisms of ageing: public or private? *Nat. Rev. Genet.* 3, 165–175

c Charlesworth, B. and Hughes, K.A. (1996) Age-specific inbreeding depression and components of genetic variance in relation to the evolution of senescence. *Proc. Natl Acad. Sci. U.S.A.* 93, 6140–6145

d Rudan, I. *et al.* (2002) Inbreeding and the genetic complexity of human hypertension. *Genetics* in press

e Bittles, A.H. and Neel, J.V. (1994) The costs of human inbreeding and their implications for variations at the DNA level. *Nat. Genet.* 8, 117–121

f Crow, J.F. (1993) Mutation, mean fitness and genetic load. *Oxf. Surv. Evol. Biol.* 9, 3–42

g McCune, A.R. *et al.* (2002) A low genomic number of recessive lethals in natural populations of Bluefin Killifish and Zebrafish. *Science* 296, 2398–2401

h Charlesworth, B. and Hughes, K.A. (1999) The maintenance of genetic variation in life-history traits. *Evolutionary Genetics: From Molecules to Morphology* (Vol. 1) (Singh, R.S., Krimbas, C.B. eds), pp. 369–392, Cambridge University Press

i Barton, N.H. and Keightley, P.D. (2002) Understanding quantitative genetic variation. *Nat. Rev. Genet.* 3, 11–21

ease for late-onset traits. Other evidence comes from suring the detrimental effects of inbreeding (INBREEDING DEPRESSION) (Box 1). Alleles with nonadditive effects h as recessives) are expected to cause increased eeding depression for late- compared with early-et traits, which is a unique prediction of the MA el. Age-related increases in V_D and in inbreeding ression have both been found for FITNESS-RELATED TS in *Drosophila* [24]. If this is also true in humans, gnificant fraction of the genetic variance underlying diseases with small effects on fitness could be due to alleles [20]. However, the MA and TO models are not ually exclusive and other *Drosophila* data support the e-off model, implying the presence of alleles at higher uencies as well [19]. The evidence from other species is ty, but an increase with age in heritability (Box 1) of human late-onset trait longevity [25] is consistent with the prediction of increased genetic variance underlying late-onset traits.

In short, the analysis of genetic variance components suggests that there is an increase in the frequencies of alleles influencing late-onset traits. But this does not imply that such variants are at high enough frequency to favour the CD/CV strategy; indeed, inbreeding effects suggest that many of them are at low frequencies. We now examine other evidence regarding the nature of such genetic variance.

Mutation and rare variants

Complex traits are influenced by many genes and so provide large MUTATIONAL TARGETS. Recent mutations provide a rich source of low-frequency variants, which account for a significant proportion of the STANDING GENETIC VARIATION in all organisms [10,26,27] (Box 2).

2. The fate of new mutations

verage time that new mutations persist in a population depends on SELECTIVE DISADVANTAGE, their dominance effects and the effective lation size (N_e) (see text) [a–c].

ost deleterious mutations are destined for ultimate loss from opulation, within a time in generations of the order of the al logarithm of N_e, unless their SELECTION COEFFICIENTS are smaller the reciprocal of the effective population size [d]. Assuming an N_e umans of ~10 000 [e,f] this gives a mean persistence time of nd 10 generations, with a very wide distribution around the n [d]. But the continual production of new mutations with each ration will lead to the maintenance of most deleterious variants given locus at low average frequencies close to those expected uilibrium in an infinitely large population (mutation–selection ce) [c].

hile they persist in the population, these alleles contribute genetic nce to the traits that they affect. The extent of the contribution nds on their mutation rates, their effects on the trait, the onship between trait and fitness, and on N_e [c,g]. Most of these ot known with any precision.

nulations suggest that loci with moderate to high *de novo* mutation contribute disproportionately to the genetic variance underlying an disease [h]. The over-representation of human disease loci with mutation rates is readily seen in monogenic diseases (see Fig. 1b in text) and this may be accentuated in more complex traits where the ational target' is considerably larger, including noncoding and atory sequences, many with very small effects [i,j]. NEUTRAL ES, such as the majority of SNPs, may reach higher frequencies re predicted to make a substantially smaller overall contribution to the disease variance, because they are likely to have very small effects on the trait.

References

a Slatkin, M. and Rannala, B. (2000) Estimating allele age. *Annu. Rev. Genomics Hum. Genet.* 1, 225–249

b Falconer, D.S. and Mackay, T.F.C. (1996) *Introduction to Quantitative Genetics*, 4th edn, Longman

c Bürger, R. (2000) *The Mathematical Theory of Selection, Recombination and Mutation*, John Wiley

d Kimura, M. (1983) *The Neutral Theory of Molecular Evolution*, Cambridge University Press

e Kaessmann, H. *et al.* (1999) Extensive nuclear DNA sequence diversity among chimpanzees. *Science* 286, 1159–1162

f Nachman, M.W. and Crowell, S.L. (2000) Estimate of the mutation rate per nucleotide in humans. *Genetics* 156, 297–304

g Charlesworth, B. and Hughes, K.A. (1999) The maintenance of genetic variation in life-history traits. *Evolutionary Genetics: From Molecules to Morphology* (Vol. 1) (Singh, R.S., Krimbas, C.B. eds), pp. 369–392, Cambridge University Press

h Pritchard, J.K. (2001) Are rare variants responsible for susceptibility to complex diseases? *Am. J. Hum. Genet.* 69, 124–137

i McKenzie, C.A. *et al.* (2001) Trans-ethnic fine mapping of a quantitative trait locus for circulating angiotensin I-converting enzyme (ACE). *Hum. Mol. Genet.* 10, 1077–1084

j Vafiadis, P. *et al.* (1997) Insulin expression in human thymus is modulated by *INS* VNTR alleles at the *IDDM2* locus. *Nat. Genet.* 15, 289–292

is especially so for populations, such as humans, nave undergone large and recent expansions. This TIONAL VARIANCE is a function of mutation rates, tional target size and population size. The EFFECTIVE LATION SIZE (N_e) refers to the number of individuals contribute genes to succeeding generations, and ols the rate of chance fluctuations in gene frequencies TIC DRIFT). Long-term N_e has been estimated to be 00 in humans, based on sequence diversity. Popu- expansion means, however, that the number of rare nts of recent origin is greater than predicted for an brium population with this value of N_e (Fig. 1). Such t and rare mutations are clearly major contributors man disease, and yet the inherent difficulty in ing them tends to overemphasize variants, such as , that are readily detectable.

st mutations are deleterious and destined to be lost n a few generations (Box 1). As a rule of thumb the ency of a mutation that reduces fitness by at least) is mainly controlled by selection rather than ic drift [28]. Therefore, even disease alleles with asurably small effects on fitness can be held at low ation frequencies [20,28]. The frequencies of new tions are therefore strongly influenced by their s on fitness, which may be substantial for mutations significant effects on disease, because the size of or indirect (pleiotropic) allelic effects on traits tends correlated with their effects on fitness [27,29,30]. For ticular disease susceptibility gene, the collective ency of new deleterious mutations is therefore ted to be low, and close to that expected under ION–SELECTION BALANCE (Box 2) [14–16].

The large size of the human genome results in a steady and substantial input of new mutations – estimated to be in the region of 175 per person per generation [31]. As a result, each of us carries a predicted 500–1200 deleterious mutations, most of which are rare (<0.01), in addition to those that are strongly deleterious and too rare to be detected in small sequence surveys [14].

Many variants with small phenotypic effects occur outside coding regions, so that GENOME-WIDE MUTATION RATES to deleterious alleles [32] and the number of deleterious mutations each person carries could be considerably higher than estimates based only on coding regions. Noncoding regions might contribute at least 60% as much as coding regions to the genome-wide mutation rate [15]. There are examples of noncoding mutations with large effects on disease [33], but those with small effects should be even more common, because there are many more ways for these to subtly alter expression or function. Indeed, there is an increasing number of examples of small-effect QUANTITATIVE TRAIT LOCUS (QTL) alleles located within noncoding regions, including ones relevant to human disease [34,35]. Mutations with essentially no effect on fitness can also influence late-onset traits, and should be at higher frequency. At the other extreme, only a small proportion of mutations show large phenotypic effects, but these tend to be over-represented in individuals with disease, as a result of selective ascertainment.

Summarizing the evidence

The possibility that a significant fraction of the genetic variation in complex traits is owing to rare alleles maintained by mutation–selection balance is supported

3. Estimating the number and effect size of genes influencing a complex trait

rinciple, it is possible to estimate the number of genetic loci encing a complex trait. In practice, estimates often have large pling variances and require assumptions that are not always met. classical method [a] is to cross inbred or natural populations that v a large difference in trait value and estimate the number of valent effect' genes accounting for the difference on the basis of the variances in the hybrid offspring (F1, F2 and backcross gener-s). This assumes that all 'high' alleles are fixed (homozygous) in parental line and 'low' alleles in the other. Zeng [b] has eliminated in these estimates by accounting for recombination between loci unequal gene effects. For tomato fruit weight, he found it difficult to nate the total number of genes, which varied from 17 to 1540 nding on the shape of the gene effect distribution, but a minimum loci accounted for 95% of the genetic variance regardless of the t distribution.

second method depends on the systematic genetic mapping of trait distinguishing a pair of lines, yielding direct estimation of gene bers [c]. This approach is constrained by lack of power to detect nts with small effects without unrealistically large samples but has useful in showing the leptokurtic or L-shaped distribution of gene t size using livestock and bristle number data [c,d].

ither of these methods provides a direct estimate of the number of segregating within a population; if the lines concerned have been ed from a natural population, they at least provide a lower bound to umber of such loci.

A third method can be applied directly to population data, including humans, and is useful for estimating the number of recessive or partially recessive loci contributing to a trait that shows inbreeding depression [e,f]. The effect of inbreeding, B, and the dominance variance, V_D, are measured (Box 1). A lower bound to the number of genes, n, affecting the trait is provided by:

$$n \geq B^2/V_D$$

References

a Falconer, D.S. and Mackay, T.F.C. (1996) *Introduction to Quantitative Genetics*, 4th edn, Longman

b Zeng, Z.-B. (1992) Correcting the bias of Wright's estimates of the number of genes affecting a quantitative character: a further improved method. *Genetics* 131, 987–1001

c Barton, N.H. and Keightley, P.D. (2002) Understanding quantitative genetic variation. *Nat. Rev. Genet.* 3, 11–21

d Hayes, B. and Goddard, M.E. (2001) The distribution of the effects of genes affecting quantitative traits in livestock. *Genet. Sel. Evol.* 33, 209–230

e Charlesworth, B. and Hughes, K.A. (1999) The maintenance of genetic variation in life-history traits. *Evolutionary Genetics: From Molecules to Morphology* (Vol. 1) (Singh, R.S., Krimbas, C.B. eds), pp. 369–392, Cambridge University Press

f Rudan, I. *et al.* (2002) Inbreeding and the genetic complexity of human hypertension. *Genetics* in press

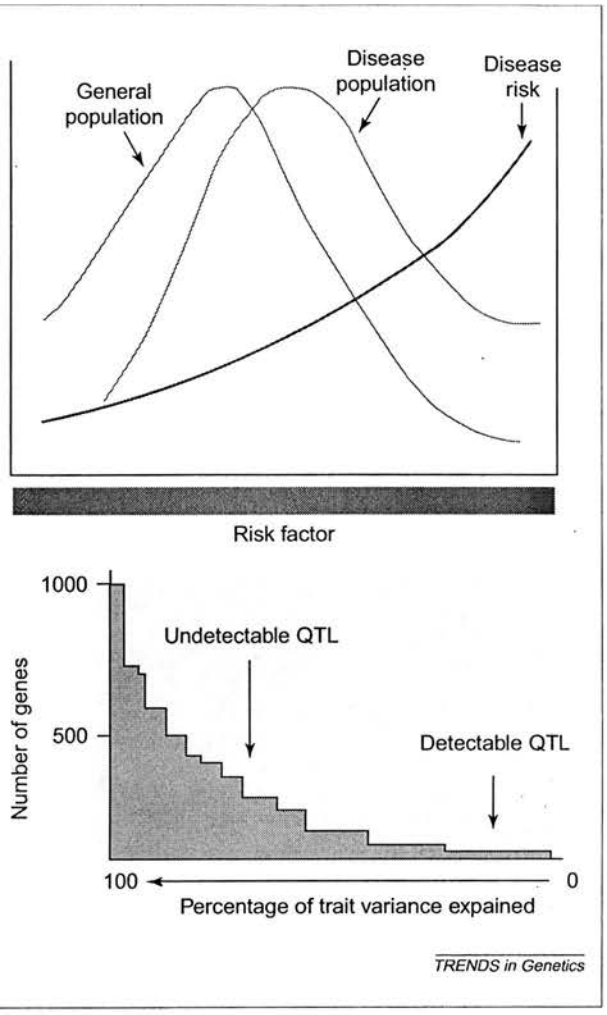
years of research into the genetics of quantitative (QTs) in *Drosophila*. A review of this extensive ture [27] suggests that deleterious alleles generated utation, and kept at low frequencies by selection, ibute between 33% and 67% of the genetic variation ypical trait with at least some effect on reproductive s. These are likely to include many of the QTs that ajor determinants of late-onset disease, whose effects isease begin well within reproductive life. The nder of the genetic variance appears to involve s at higher frequencies, maintained by some form of NCING SELECTION, such as heterozygote advantage or oral, spatial or frequency-dependent selection [1,27]. erall, allele frequencies might be higher for late- for early-onset traits, but this is consistent with ls favouring both low (MA model) and intermediate nodel) frequency alleles. We currently lack any firm n the size of these effects in humans, and on whether cting alleles that are at higher frequencies are the nost likely to be important in understanding disease. e contrary, the low predicted frequency of function- deleterious alleles suggests that most clinically icant disease susceptibility alleles will not be 'high' sense implied by the CD/CV hypothesis. Strategies entifying susceptibility alleles that are not robust in ce of a large and diverse group of rare alleles could fore fail.

umber and size of allelic effects

plicit assumption of most common disease mapping gies, including the CD/CV hypothesis [11], is that ex traits are oligogenic, so that susceptibility loci are tially detectable and informative for disease predic- r understanding. If, however, common diseases are truly polygenic, most individual effects will be too small to be useful. A limited number of loci appear able to exert strong monogenic effects on disease, suggesting that these influence rate-determining steps in disease pathways. The rarity of individual mutations under the CD/RV model, together with the large number of loci at which mutations can arise, means that the genetic basis of disease could vary greatly among individuals with the same disease.

The locus complexity of a trait depends on two factors. The first is its physiological complexity. Hirschsprung disease, for example, is a rare and highly specific lack of enteric ganglia, in which no more than eight loci account for most of the variance [36]. By contrast, coronary artery disease is a highly complex multidimensional trait with an exponentially increasing lifetime risk that is influenced by more than 280 risk factors [37,38] (Fig. 2a). The second factor is the distribution of gene effects. This is clearly important, because if most of the trait variance is determined by a handful of potentially detectable genes, the total number is irrelevant. An L-shaped or exponential distribution of mutation effect sizes has wide support [1,39], with many genes of small effect and fewer of large effect. However, even if a small number of genes account for the most extreme effects, the average patient with disease will not show large effects if there are many other determinants (Box 3).

Genetic models commonly assume a polygenic basis for complex traits, but the actual number of loci is hard to estimate [1]. Many loci with small effects will not be detectable using finite samples, and the effects of individual alleles are commonly overestimated. One way to estimate the number of loci is to combine estimates of dominance variance (V_D) and the INBREEDING LOAD (B) (Box 3). Because inbreeding depression results from increased homozygosity at loci influencing the trait, a regression of QT value (or



etection of clinically useful genetic effects in 'typical' patients with late-sease is likely to be impossible, but it may be feasible in those at the s of the disease or risk factor distribution. (a) Average patients have aver-etic effects. The relationship between risk factor distributions in the gen-ulation and in a population with disease shows that most disease events those within the normal range, because risk increases at all levels [41,42]. phasizes the need for genetic analysis of those at the extremes of the dis-. (b) Predicted L-shaped relationship between the number of quantitative (QTL) genes and the percentage of the trait variance explained. QTL con-to the left of the distribution will often be undetectable, and those contri-o the tail of the distribution on the right show larger effects and are ly detectable. The number of QTL might be too large for specific loci to be ally detectable in 'typical' patients, most of whom lie within the normal r individual risk factors. However, gene identification could be both tract-l heuristically useful in those at the extremes of the distribution (e.g. 10%) [40].

e prevalence) on the INBREEDING COEFFICIENT can e a lower limit for the number of segregating loci). If the minimum number of loci required to explain of the trait variance is large, then individual QTL might be too small to be detected with finite samples. *et al.* [40] came to exactly this conclusion in a study of ding effects on blood pressure. Assuming an L-shaped ution of allelic effect sizes and V_D/V_P of 0.33, the 25% of variation in systolic blood pressure could be ned by a minimum of 24–48 genes, and the upper 50% east 90–165 genes [40] (Fig. 3b). The total number of utory loci must be even larger, because this estimate es those with purely additive effects. These results suggest that loci underlying a trait such as hypertension in a 'typical' patient are likely to be so numerous and of such small effect that gene identification will be either impossible or unhelpful. The problem is compounded by the fact that most risk factors in common diseases operate at virtually all levels, so that the 'average' patient typically has a trait value within the normal range, and most individual effects could be too small to detect [41,42] (Fig. 3a).

Global environmental change

To what extent does recent human environmental change affect our increasingly elderly populations? The shift in age structure is itself a major environmental change to which our genomes are poorly adapted. However, a wide range of common diseases have shown large changes even in age-standardized prevalence within the past 50 years, especially in western societies [43]. Environmental changes almost certainly account for this, raising questions about the overall significance of heritable components (Box 4). Gene–environment interactions can be seen with late-onset diseases, such as type 2 diabetes [44], in which recent dietary changes appear to be a major contributor to disease [45]. Ethnic variation in susceptibility to such diseases is often taken as support for the role of genes, but it is equally plausible that the most relevant genetic factors are essentially invariant in such populations as a result of strong selective advantages in the past [46] – the COMMON DISEASE/FIXED VARIANT (CD/FV) HYPOTHESIS. Conventional mapping studies would then fail to detect them. A striking example of the power of selection is the relatively recent FIXATION of the Duffy blood group *FY*O* allele in sub-Saharan Africans, probably because of increased resistance to *Plasmodium vivax* malaria [47].

Other genes appear to have been under strong selection in certain populations without leading to fixation, resulting in alleles at intermediate frequencies [48,49] (Fig. 1a). Variants such as those found in certain MHC class I and II HAPLOTYPES show functionally significant effects and are therefore good candidates for the CD/CV approach. Traits exposed to major environmental change, such as dietary alteration at the time of urbanization, immunity to disease, response to starvation and some cultural behaviours are promising in this respect.

Gene identification under a polygenic hypothesis

The scenario of a major class of deleterious but individually rare alleles of recent origin underlying the heritable component of late-onset diseases has been largely ignored until recently. It poses seemingly intractable statistical problems for gene mapping and is therefore an unattractive investment for grant-awarding agencies and biotechnology companies. More attractive alternatives, such as the CD/CV hypothesis, appear to be driving the research strategy in spite of, rather than because of, the science.

An example is the proposal to use high-frequency SNP haplotypes to identify common disease determinants by population association studies [50,51]. A recent mutational origin for the variants underlying much of the heritability in such diseases means that many will be superimposed on ancient core haplotypes. The majority of new variants are deleterious [26,27] and multiple variants

4. Establishing a genetic basis for late-onset disorders

rding to the website Online Mendelian Inheritance in Man (http://
.ncbi.nlm.nih.gov/entrez/Omim/mimstats.html), fewer than 10%
uman transcribed sequences (including alternative transcripts)
v mutations capable of major (monogenic) effects that are relatively
pendent of context. By contrast, most common diseases are highly
ndent on both genetic and environmental context, although they
rally show significant heritability (≥ 0.1) (see Fig. 2b in main text).
ritability is assumed to be present if there is an increased trait
lation in relatives compared with unrelated individuals. Relatives
ecially twins) are, however, exposed to more similar environments
random controls; also, genetic and environmental effects fre-
tly interact and are correlated, so that in the absence of extensive
sib or adoptive twin data, basic assumptions may not be met,
ng to overestimation of heritability [a].
e magnitude of genetic effects in complex traits is frequently
fed by known environmental factors, leading some to question
current emphasis on genetics [b]. For example, diet, physical
ivity and tobacco have been proposed to account for 75% of new
s of cardiovascular disease [c]. In developed countries, tobacco
king alone causes one-third of all cancer deaths [d], 80% of
nic bronchitis [e] and 13% of coronary artery disease [e]. In
tion, recent environmental change has resulted in a global
emic of type 2 diabetes, with uncertain consequences for
ability. In coronary artery disease, heritability declines with
of onset [f], implicating either increased environmental variance
reduced total genetic variance (V_G), or both, in later life (see
2b in main text).
e contribution of individual genetic factors to complex disease is, in
cases, seldom greater than a few percent of the trait variance and a
l percentage of cases. For example, currently known breast cancer
s account for <2% of population risk [g]. The collective contri-
n of genetic variability, summarized by measures such as
ability or coefficients of genetic variation, is however both
tantial and almost universal in complex traits, for reasons that
iscussed in the text.
en with the standardized environments and reduced diversity of
ratory mice, mapping of QTL for obesity-related traits suggests
y genes of small effect [h]. Some QTL influencing human disease
v larger effects, such as *APOE*E4*, which accounts for up to 10% of
ariance in both plasma cholesterol and age of onset for Alzheimer
disease, although some of these effects might result from interactions (e.g. with dietary fat [i]). Opportunities for disentangling interactions between genes and environment are often limited [j].

The identification of genes accounting for rare monogenic forms of common diseases (e.g. breast and colon cancer, Alzheimer disease) has made a substantial contribution towards elucidating disease mechanisms and has led to therapeutic progress (e.g. statins, secretase inhibitors). It is less clear whether the identification of variants with small effects on disease risk will have a similar impact. Predictions based on experimental organisms suggest that the lack of control of genotype or environment in human studies, together with context-dependence of QTL effects, mean that individual QTL effects will generally be very small [k].

References

a Kamin, L.J. and Goldberger, A.S. (2002) Twin studies in behavioral research: a skeptical view. *Theor. Popul. Biol.* 61, 83–95
b Holtzman, N.A. and Marteau, T.M. (2000) Will genetics revolutionize medicine. *N. Engl. J. Med.* 343, 141–144
c Beaglehole, R. (2001) Global cardiovascular disease prevention: time to get serious. *Lancet* 358, 661–663
d Peto, J. (2001) Cancer epidemiology in the last century and the next decade. *Nature* 411, 390–395
e Vineis, P. *et al.* (2001) Misconceptions about the use of genetic tests in populations. *Lancet* 357, 709–712
f Marenberg, M.E. *et al.* (1994) Genetic susceptibility to death from coronary heart disease in a study of twins. *N. Engl. J. Med.* 330, 1041–1046
g Pharoah, P.D. *et al.* (2002) Polygenic susceptibility to breast cancer and implications for prevention. *Nat. Genet.* 31, 33–36
h Cheverud, J.M. *et al.* (2001) Genetic architecture of adiposity in the cross of LG/J and SM/J inbred mice. *Mamm. Genome* 12, 3–12
i Tikkanen, M.J. (1997) Apolipoprotein E polymorphism and plasma cholesterol response to dietary change. *World Rev. Nutr. Diet.* 80, 15–21
j Wright, A.F. *et al.* (2002) Gene–environment interactions – the BioBank UK study. *Pharmacogenomics J.* 2, 75–82
k Mackay, T.F.C. (2001) Quantitative trait loci in *Drosophila*. *Nat. Rev. Genet.* 2, 11–20

rise on the commonest haplotypes, so that association
ing is inefficient. Disease alleles might be missed
are not associated with ancient core haplotypes [52].
otype frequencies will be poorly matched to disease
frequencies, and disease associations will be hard to
t, especially if alleles of opposite effect arise on the
haplotype. But if susceptibility alleles are predomin-
rare and numerous, are there any strategies that will
?

ost complex human disease alleles identified to date
partially recessive [12], as predicted from other
isms [26,27,53]. To maximize the 'detectance' [54]
h alleles, ascertainment for large phenotypic effects
4] and for populations showing reduced environ-
al and genetic variance [3,55], and the use of large
le sizes [56] will all help. The use of QTs rather
disease *per se* also increases informativeness [3,57].
me CONCORDANT or DISCORDANT sibs with trait
s (or disease age-of-onset) close to the extremes of
istribution could also be helpful [57]. Large endoga-
populations might be especially valuable, because of
environmental uniformity and their ability to reveal
ical-by-descent (IBD) segments based on kinship
information (using relatively low-density genome scans) [3,36,55]. A search for homozygous IBD segments could help to detect rare recessive alleles of large effect. Many disease loci identified in this way could be 'local' to specific populations. If this provides new insights into disease pathways it will be invaluable, because new drug targets are in short supply. Finally, animal models, in which inbred or selected strains show differences in disease or trait value, are more tractable [58], because far fewer segregating loci are involved, although they represent a minute subset of the variation found in wild populations.

Genetic linkage is robust in the presence of allelic heterogeneity, whereas association studies are not, but both lack power in the face of a strongly polygenic basis for disease. The familial breast cancer genes were identified by linkage because some families showed large-effect alleles and locus heterogeneity was limited. The choice of study population can help to minimize disease complexity but the choice of disease-related phenotype is equally important [3,57].

Admixture mapping [59] has advantages, especially under a CD/FV model. Admixture between ethnic groups showing high and low disease prevalence, under currently

ar but differing historical environments, could be the way to identify moderate- or large-effect genes con- ting to such differences. The diversity of hazards in ent geographic and cultural environments would tend aggerate selective differences between populations, ugh demographic forces might have dispersed them ghout large urban populations.

large continental populations, the high levels of and allelic heterogeneity can be turned to ntage by the relative ease of identifying either familial forms of disease or rare survivors [25]. approach has produced disproportionate gains in rstanding disease mechanisms in monogenic dis- rs [12]. The recruitment of sufficiently large series of rios or quartets, especially when concordant or rdant for QT or early-onset disease, might only be ble in large national studies [60].

ıe proposed complexity of late-onset disorders sug- that identifying genes with the largest effects, which contribute most to the extremes of the se or trait distribution, might be the most robust oach – only one step away from the methods so ssfully applied to monogenic disorders. This will to avoid the illusion of large QTL effects that are osite artefacts of numerous but clustered variants or with individual effects so small as to be of value in elucidating disease mechanisms. The ulties encountered in identifying clinically rel- disease loci could reflect an overambitious goal of ng genes involved in the majority of patients with ease, fuelled by commercial forces. The scenario of ly polygenic basis for much of the heritability of lex traits might require both new approaches and urn to tried and tested ones.

owledgements

ıd H.C. were supported in part by an International Research pment Award from The Wellcome Trust. B.C. is supported by the Society. We thank A.S. Kondrashov, C.M. Steel, P.M. Visscher and Iastie for useful comments and suggestions, and S. Bruce for the k.

ences

rton, N.H. and Keightley, P.D. (2002) Understanding quantitative etic variation. *Nat. Rev. Genet.* 3, 11–21
ıder, E.S. (1996) The new genomics: global views of biology. *Science* , 536–539
ght, A.F. *et al.* (1999) Population choice in mapping genes for ıplex diseases. *Nat. Genet.* 23, 397–404
essmann, H. *et al.* (1999) Extensive nuclear DNA sequence diversity ong chimpanzees. *Science* 286, 1159–1162
hidanandam, R. *et al.* (2001) A map of human sequence variation taining 1.42 million single nucleotide polymorphisms. *Nature* 409, –933
eworski, M. *et al.* (2000) Adjusting the focus on human variation. ıds *Genet.* 16, 296–302
ıglyak, L. and Nickerson, D.A. (2001) Variation is the spice of life. . *Genet.* 27, 234–236
ck, M.E. *et al.* (2000) Patterns of genetic variation in Mendelian complex traits. *Annu. Rev. Genomics Hum. Genet.* 1, 387–407
ıura, M. (1971) Theoretical foundations of population genetics at molecular level. *Theor. Popul. Biol.* 2, 174–208
ch, M. *et al.* (1999) Perspective: spontaneous deleterious mutation. *lution* 53, 645–663
11 Reich, D.E. and Lander, E.S. (2001) On the allelic spectrum of human disease. *Trends Genet.* 17, 502–510
12 Wright, A.F. and Hastie, N.D. (2001) Complex genetic diseases: controversy over the Croesus code. *Genome Biol.* 2, 2007
13 Healey, C.S. *et al.* (2000) A common variant in *BRCA2* is associated with both breast cancer risk and prenatal viability. *Nat. Genet.* 26, 362–364
14 Fay, J.C. *et al.* (2001) Positive and negative selection on the human genome. *Genetics* 158, 1227–1234
15 Sunyaev, S. *et al.* (2001) Prediction of deleterious human alleles. *Hum. Mol. Genet.* 10, 591–597
16 Fisher, R.A. (1930) *The Genetical Theory of Natural Selection*, Oxford University Press
17 Medawar, P.B. (1946) Old age and natural death. *Mod. Quart.* 1, 30–56
18 Williams, G.C. (1957) Pleiotropy, natural selection and the evolution of senescence. *Evolution* 11, 398–411
19 Partridge, L. and Gems, D. (2002) Mechanisms of ageing: public or private? *Nat. Rev. Genet.* 3, 165–175
20 Charlesworth, B. and Hughes, K.A. (1996) Age-specific inbreeding depression and components of genetic variance in relation to the evolution of senescence. *Proc. Natl Acad. Sci. U.S.A.* 93, 6140–6145
21 Charlesworth, B. (2001) Patterns of age-specific means and genetic variances of mortality rates predicted by the mutation accumulation theory of ageing. *J. Theor. Biol.* 210, 47–65
22 Mueller, L.D. and Rose, M.R. (1996) Evolutionary theory predicts late-life mortality plateaus. *Proc. Natl Acad. Sci. U.S.A.* 93, 15249–15253
23 Shaw, F.H. *et al.* (1999) Toward reconciling inferences concerning genetic variation in senescence in *Drosophila melanogaster*. *Genetics* 152, 553–566
24 Hughes, K.A. *et al.* (2002) A test of evolutionary theories of aging. *Proc. Natl Acad. Sci. U.S.A.* 99, 14286–14291
25 Perls, T. *et al.* (2002) The genetics of aging. *Curr. Opin. Genet. Dev.* 12, 362–369
26 Simmons, M.J. and Crow, J.F. (1977) Mutations affecting fitness in *Drosophila* populations. *Annu. Rev. Genet.* 11, 49–78
27 Charlesworth, B. and Hughes, K.A. (1999) The maintenance of genetic variation in life-history traits. *Evolutionary Genetics: From Molecules to Morphology* (Vol. 1) (Singh, R.S., Krimbas, C.B. eds), pp. 369–392, Cambridge University Press
28 Kimura, M. (1983) *The Neutral Theory of Molecular Evolution*, Cambridge University Press
29 Keightley, P.D. and Hill, W.G. (1990) Variation maintained in quantitative traits with mutation–selection balance: pleiotropic side-effects on fitness traits. *Proc. R. Soc. Lond. Ser. B* B242, 95–100
30 Kruuk, L.E.B. *et al.* (2000) Heritability of fitness in a wild mammal population. *Proc. Natl Acad. Sci. U.S.A.* 97, 698–703
31 Nachman, M.W. and Crowell, S.L. (2000) Estimate of the mutation rate per nucleotide in humans. *Genetics* 156, 297–304
32 Kondrashov, A.S. (2001) Sex and *U. Trends Genet.* 17, 75–77
33 Kleinjan, D.A. *et al.* (2001) Aniridia-associated translocations, DNase hypersensitivity, sequence comparison and transgenic analysis redefine the functional domain of PAX6. *Hum. Mol. Genet.* 10, 2049–2059
34 McKenzie, C.A. *et al.* (2001) Trans-ethnic fine mapping of a quantitative trait locus for circulating angiotensin I-converting enzyme (ACE). *Hum. Mol. Genet.* 10, 1077–1084
35 Vafiadis, P. *et al.* (1997) Insulin expression in human thymus is modulated by *INS* VNTR alleles at the *IDDM2* locus. *Nat. Genet.* 15, 289–292
36 Gabriel, S.B. *et al.* (2002) Segregation at three loci explains familial and population risk in Hirschsprung disease. *Nat. Genet.* 31, 89–93
37 Lloyd-Jones, D.M. *et al.* (1999) Lifetime risk of developing coronary heart disease. *Lancet* 353, 89–92
38 Hopkins, P.N. and Williams, R.R. (1986) Identification and relative weight of cardiovascular risk factors. *Cardiol. Clin.* 4, 3–31
39 Hayes, B. and Goddard, M.E. (2001) The distribution of the effects of genes affecting quantitative traits in livestock. *Genet. Sel. Evol.* 33, 209–230
40 Rudan, I. *et al.* (2002) Inbreeding and the genetic complexity of human hypertension. *Genetics* in press
41 Rose, G. (1992) *The Strategy of Preventive Medicine*, Oxford University Press
42 Law, M.R. and Wald, N.J. (2002) Risk factor thresholds: their existence under scrutiny. *Br. Med. J.* 324, 1570–1576

rlton, J. and Murphy, M. (1997) *The Health of Adult Britain 1841–*
1, HMSO
ussin, E. *et al.* (1994) Effects of a traditional lifestyle on obesity in
a Indians. *Diabetes Care* 17, 1067–1074
l, J.V. (1962) Diabetes mellitus: a 'thrifty' genotype rendered
imental by progress? *Am. J. Hum. Genet.* 14, 353–362
ntorp, P. (2001) Thrifty genes and human obesity. Are we chasing
sts? *Lancet* 358, 1006–1008
nblin, M.T. *et al.* (2002) Complex signatures of natural selection at
Duffy blood group locus. *Am. J. Hum. Genet.* 70, 369–383
kin, M. and Rannala, B. (2000) Estimating allele age. *Annu. Rev.*
omics Hum. Genet. 1, 225–249
riman, T.R. and Todd, J.A. (1995) Genetics of autoimmune disease.
r. Opin. Immunol. 7, 786–792
h, D.E. *et al.* (2001) Linkage disequilibrium in the human genome.
ure 411, 199–204
l, N. *et al.* (2001) Blocks of limited haplotype diversity revealed by
-resolution scanning of human chromosome 21. *Science* 294,
9–1723
C. (2002) Irresistible force meets immovable object: SNP mapping
mplex diseases. *Trends Genet.* 18, 67–69
w, J.F. (1993) Mutation, mean fitness and genetic load. *Oxf. Surv.*
. Biol. 9, 3–42
ss, K.M. and Terwilliger, J.D. (2000) How many diseases does it
to map a gene with SNPs? *Nat. Genet.* 26, 151–157
ey, M. *et al.* (2002) Quantitative-trait homozygosity and associ-
ation mapping and empirical genomewide significance in large, complex pedigrees: fasting serum-insulin level in the Hutterites. *Am. J. Hum. Genet.* 70, 920–934
56 Dahlman, I. *et al.* (2002) Parameters for reliable results in genetic association studies in common disease. *Nat. Genet.* 30, 149–150
57 Gu, C. and Rao, D.C. (2001) Optimum study designs. *Adv. Genet.* 42, 439–457
58 Flint, J. and Mott, R. (2001) Finding the molecular basis of quantitative traits: successes and pitfalls. *Nat. Rev. Genet.* 2, 437–445
59 McKeigue, P.M. (1998) Mapping genes that underlie ethnic differences in disease risk: methods for detecting linkage in admixed populations, by conditioning on parental admixture. *Am. J. Hum. Genet.* 63, 241–251
60 Wright, A.F. *et al.* (2002) Gene–environment interactions – the BioBank UK study. *Pharmacogenomics J.* 2, 75–82
61 Visscher, P.M. and Haley, C.S. (1996) Detection of putative quantitative trait loci in line crosses under infinitesimal genetic models. *Theor. Appl. Genet.* 93, 691–702
62 Pritchard, J.K. (2001) Are rare variants responsible for susceptibility to complex diseases? *Am. J. Hum. Genet.* 69, 124–137
63 Stevenson, A.C. and Kerr, C.B. (1967) On the distribution of frequencies of mutation in genes determining harmful traits in man. *Mutat. Res.* 4, 339–352
64 Vogel, F. and Motulsky, A.G. (1986) *Human Genetics*, Springer-Verlag