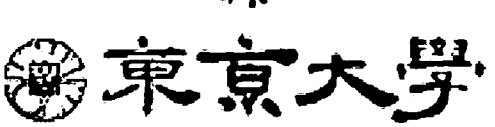
Developmental genetics and evolution of plant form in *Streptocarpus*

Cecily Jill Harrison

ROYAL
BOTANIC
GARDEN
EDINBURGH



The Gatsby Charitable Foundation

The British Council

Doctor of Philosophy
University of Edinburgh
2002

Declaration

I declare that this work is my own unless otherwise stated, and that it has not previously been submitted for a degree.

Abstract

The mechanisms by which genetic mutation translates to morphological variation between species are poorly understood, but there is increasing support for the involvement of homeobox genes in generating morphological novelty. Homologues of a *knotted*-like homeobox gene, that is involved in determining meristem identity in model plants, were isolated from different species of *Streptocarpus* which show marked morphological variation, and these were named *Sknox1* genes. Crosses were made between two species with different form, which suggested that their differences were determined by two loci, consistent with previous reports. To test whether *Sknox1* was involved in generating morphological differences between species of *Streptocarpus*, its expression was analysed in three species with different growth form. This showed conventional expression in a species with caulescent form, but novel patterns of expression in species with one-leafed or rosette-like form. Differences in expression of *Sknox1* therefore correlate with interspecific differences in form. Genetic analysis demonstrated that *Sknox1* did not directly cause morphological differences between species. *Sknox1* intron sequences varied substantially between species. Intron sequences showed evidence of concerted evolution and sometimes contained a repetitive element that may have arisen by gene conversion.

Table of Contents

Chapter 1: Introduction 16

Molecular basis of differences between species 16

Morphological differences between species of *Streptocarpus* 18

What is Streptocarpus? 18

Novel growth forms in subgenus Streptocarpus 18

Figure 2: aspects of *Streptocarpus* morphology... 20

How differences in *Streptocarpus* form are attained 21

Ontogeny of Streptocarpus 21

Seedling development 21

Meristem establishment 22

Manipulation of morphology 24

Phylogeny of *Streptocarpus* 24

Origin of acaulescent species 24

Developmental and genetic bases of differences in form 26

Genetic interactions in model species' meristems 26

Meristem establishment during Capsella *Embryogenesis* 27

Genes required for establishment of the SAM 27

Meristem structure 31

Meristem maintenance and organogenesis during post embryonic development 32

Genes that affect phase changes in model species 35

Genes that could be implicated in Streptocarpus *evolution* 35

Structure and function of knox homeodomain proteins 36

Meristem behaviour, *STM*, and the evolution of different forms of *Streptocarpus* 39

Changes in meristem behaviour in Streptocarpus 39
Parallels between knox *mutants and* Streptocarpus *morphology* 39
Aim of the PhD 40

Chapter 2: Materials and methods 41

Plant material 41

Molecular techniques 42

Nucleic acid extraction 42
PCR 44
Cleaved Amplified fragment Polymorphisms (CAPS) based screen for segregation of Sknox1 alleles 50
Gel electrophoresis of DNA 51
Gel purification of PCR products 51
Cloning 51
Sequencing 52
Sequence analysis, alignment and phylogenetic analysis 52
Southern blotting 53
Microscopy and histological techniques 56

Chapter 3: Inheritance of morphology in *Streptocarpus* 57

Crosses performed 57

Analysis of growth form in backcross plants 58

Results of the crosses 60
Some curiosities in form 64

Summary of results 67

Chapter 4: Characterisation of genes which could effect form in *Streptocarpus*.....68

Isolation of *Streptocarpus STM*-like genes..... 68

Amplification of 3' region..... 69

***Sknox* copy number 70**

Cloning 5' *Sknox* sequences.....72

Extension in the 5' direction using 5'RACE..... 72
Amplification to ELK domain 72
Amplification to intron 2 73
Loss of specificity in IPCR..... 73
Increased specificity of amplification using genome walking.....74
Amplification of the 5' end of Ssknox1.....*74*
Amplification of 5' coding regions from acaulescent species..... 75

Compilation of contiguous sequences 76

Identity of isolated genes.....76

Sknox1 *genes show conserved features of* knox *genes.....76*
Similarity between Streptocarpus knox *genes.....79*
Similarity of Streptocarpus *genes to other* knox *genes.....79*
Differences in structure between Sknox1 *sequences, and possible functional implications..... 80*
Phylogeny..... 81
Homeodomain nucleotide sequence phylogeny 81
Amino acid sequence phylogeny..... 84
Trends shown 84
Sdknox2 *is the result of a duplication that occurred after the divergence of* Streptocarpus *and* Antirrinum *86*

Expression of *Sknox1 genes* 87

Experimental design 87
Possible outcomes 88
Outcome 88
Meristematic properties in Streptocarpus *phyllomorphs* 90
Further meristem-like features of the phyllomorph 90

Further genes isolated from *Streptocarpus* 92

Relationships between CLV1*-like genes* 92
NAC domain genes 93
Isolation of a Streptocarpus *NAC gene* 94
Affinity of SdNAC1 *to other NAC genes* 96
Transposons 96

Record of sequence discussed 97

Record of further sequence not included in this thesis 97

Summary of results 97

Discussion 98

What could be causing expression of Sknox1 *in the phyllomorph?* 98
When and where is STM *first expressed in* Streptocarpus*?* 99
Redundancy of gene duplicates 99

Chapter 5: Effects of inheritance and expression of *Sknox1* on *Streptocarpus* morphology 101

A CAPS based screen for segregation of *Sknox1* alleles 101

Results of a screen using intron 3 alone 101
Screen using coding sequence 5' of intron 3 102
Inheritance of Sknox1 *alleles was Mendelian* 104
Correlation to morphology? 105
Could Sknox1 *affect other aspects of form?* 106

A CAPS based screen for inheritance of SNAC1 106
Other genes could affect morphology 107

Summary of results 108

Discussion 108

Could Sknox1 *genes be regulated by* cis*-acting elements?* 108
Major differences in growth form between species 108
Differences in degree of leaf gnarling between species 110

Chapter 6: Phylogenetic analysis and evolution of *Sknox* genes 112

Phylogenetic analysis of intron 3 sequences 112

Taxon sampling 112

PCR and sequencing results 113

Conserved motifs in intron 3 and sequence alignment 115
Alignment 115
Rooting 116
Phylogenetic analysis 116

Repetitive elements were found in some species 125

Summary 128

Discussion 128

Repeat evolution 128
Evolution of repeats in Streptocarpus 129

Chapter 7: Discussion 132

Morphological change appears rapid in *Streptocarpus* ... 132

Inflorescence meristem development in acaulescent species 133
Possible means of identification of loci that determine form 133

Mechanisms of evolution 134

Differences of plant growth and evolution to animals *136*
Evidence of links between genetic and phenotypic variation in plants *136*
Streptocarpus *as an evolutionary and developmental case study* *138*

Phylogenetic utility of Sknox gene introns 138

Genetics *139*
Molecular biology *139*

References 142

Acknowledgements 155

Appendices 156

Appendix 1: Morphological data 156

Appendix 1.1: Backcross family 1 morphology 156

Appendix 1.2: Morphology of other backcross plants 157

Appendix 1.3: Leaf gnarling phenotypes 161

Appendix 2: *Sknox* sequences 162

Appendix 2.1: cDNA sequences of *Streptocarpus STM*-like genes obtained by 3'RACE 162

Appendix 2.1A: Ssknox1 *3' region* *162*
*Appendix 2.1B:*Srknox1 *3' region* *162*
Appendix 2.1C: Sdknox1 *3' region (a genomic contaminant from 3' RACE)* *162*
*Appendix 2.1D:*Sdknox2 *3'region* *162*

Appendix 2.2:Partial genomic products 163

Appendix 2.2A: Sdknox1 *sequences* 163
Appendix 2.2B: Srknox1 *sequences* 164
Appendix 2.2C: Swknox1 *sequences* 164

Appendix 2.3: Partial cDNA products 165

Sdknox1 165
Srknox1 165

Appendix 2.4: Deduced amino acid sequences of *Streptocarpus* homologues of *STM* 166

Appendix 2.4A: Sdknox1 166
Appendix 2.4B: Srknox1 166
Appendix 2.4C: SSknox1 166

Appendix 2.5:Ribosomal cDNA sequences amplified by 5' RACE 167

Appendix 2.5A: Streptocarpus rexii *5.8s ribosomal subunit* 167
Appendix 2.5B: Streptocarpus rexii *40s ribosomal subunit* 167

Appendix 2.6: Record of sequences amplified 168

Appendix 3: *Sknox1* genotypes for backcross plants 169

Appendix 4: Intron 3 sequence data 171

Appendix 4.1: *Sknox* gene intron 3 sequences 171

Appendix 5: Properties of data matrices and phylogenies presented in this thesis. 179

Appendix 6: GenBank accession numbers of knox nucleotide sequences used in phylogeny 180

List of figures

Figure 1: A: the Ultrabithorax-Extradenticle-DNA ternary complex 17
Figure 2: aspects of *Streptocarpus* morphology 20
Figure 3: Germination and seedling development in *Streptocarpus* 21
Figure 4: Anisocotyly in *Streptocarpus* seedlings 22
Figure 5: Main clades of *Streptocarpus* shown in an ITS sequences phylogeny 26
Figure 6: Stages of embryonic development in *Arabidopsis thaliana* 27
Figure 7: Mutants that do not fully establish a SAM 30
Figure 8: Structure of an *Arabidopsis* SAM 32
Figure 9: Expression domains of *CLV1*, *CLV3*, and *WUS* in the shoot meristem 34
Figure 10: Effects of *knox* mis-expression in various species 38
Figure 11: Predicted segregation of alleles and morphology in an F2 population 57
Figure 12: *S. dunnii*, F1 hybrid, and *S. rexii* 58
Figure 13: Leaves showing different degrees of leaf gnarling 63
Figure 14: Production of secondary leaves from the groove meristem 65
Figure 15 A: Further phenotypes in backcross plants 66
Figure 16: Further leaf phenotypes in *S. wittei* (*S. rexii* x *S. wittei*) backcross plants 66
Figure 17: Region of *Streptocarpus knox* homologues amplified. 69
Figure 18: Diagrammatic representation of region amplified by 3' RACE 70
Figure 19: At least two restriction fragments from each species hybridise to a 3' *Sknox1* probe in Southern analysis. 71
Figure 20. Location of primers for 5' RACE 72
Figure 21: Diagrammatic representation of regions amplified using IPCR 74
Figure 22: Region amplified using genome walking 75
Figure 23: Region amplified using RT- 75
Figure 24: Genomic sequence of a *Streptocarpus saxorum STM* homologue, *Ssknox1* 78
Figure 25: Diagrammatic representation of *Ssknox1* gene structure 79
Figure 26: Percentage similarity of *STM* to *knotted1* and *Ssknox1*. 80
Figure 27: *Knox* amino acid sequence alignment 80
Figure 28: Homeodomain nucleotide sequence phylogeny of *knox* and MEIS genes 83
Figure 29: Phylogeny of *KNOX* proteins 85
Figure 30: Hypothetical phylogenies to investigate the timing of gene duplication. 86

Figure 31: Affinity of *Sdknox2* to other *knox* genes 87
Figure 32: RT-PCR analysis of expression of *Streptocarpus STM* orthologues 89
Figure 33: Sagittal section through *S. rexii* seedling stained with DAPI and BrdU 91
Figure 34: Diagrammatic representation of *CLV*1 93
Figure 35: Phylogeny of *CLV1*- like genes 93
Figure 36: Diagrammatic representation of *NAM* 94
Figure 37: DNA sequence of *SdNAC1* 95
Figure 38: Diagrammatic representation of structure of *SdNAC1* 95
Figure 39: Phylogeny of NAC domain genes 96
Figure 40: Diagram of *SSknox1*, showing position of primers used to amplify fragments for a CAPS screen in acaulescent species 102
Figure 41: CAPS based screen for segregation of *Sknox1* alleles in backcross populations. 103
Figure 42: Inheritance of *Sknox1* alleles 104
Figure 43: Diagrammatic representation of *SdNAC1* showing position of primers 107
Figure 44: Basis of a CAPS based screen for the segregation of *SNAC1* alleles 107
Figure 45: Abridged ITS phylogeny redrawn from Möller and Cronk 113
Figure 46: Example of product migration in a minigel to illustrate PCR amplification of multiple products from some species of *Streptocarpus*. Arrows indicate long (top) and short (bottom) products amplified from a single species. 114
Figure 47: Phylogenetic hypotheses relating to *Sknox* gene duplication. 119
Figure 48: Alignment of coding sequences amplified from *Sknox* genes 121
Figure 49: Intron 3 alignment 123
Figure 50: Phylogenetic analysis of intron 3 data 124
Figure 51: Phylogenetic hypotheses relating to *Sknox* gene duplication 119
Figure 52: Dot plots showing frequency of repeats 125
Figure 53: Alignment of repetitive elements. 126
Figure 54: Phylogeny of intron repeats in two species of *Streptocarpus* 127
Figure 55: Models of repeat evolution 131

List of tables

Table 1: *Streptocarpus* species used in this study 41
Table 2: Plants used to generate backcross populations 42
Table 3: Details of backcrosses in *Streptocarpus*. 42
Table 4: Primers referred to in this text, and their source 46
Table 5: Form during juvenile development in backcross plants 59
Table 6: Form of backcross plants at flowering. 60
Table 7: Summary of the results of Oehlkers' crosses. 62
Table 8: Summary of scores of leaf gnarling in backcross plants. 64
Table 9: Sequences isolated from *Streptocarpus* 97
Table 10: Number of *Sknox1* homozygotes and heterozygotes in *Streptocarpus* backcrosses 105
Table 11: Scores allocated to family 1 backcross plants. 156
Table 12: Number of phyllomorphs of backcross plants scored at different times 160
Table 13: leaf gnarling phenotypes in backcross plants. 161
Table 14: Accession numbers of fragments amplifed 168

Abbreviations

ARP	AS1 Rough Sheath2 and phantastica-like genes
bp	Base pairs
BrdU	5-bromodeoxyuridine
BSA	Bovine serum albumin
CAPS	Cleft amplified fragment polymorphisms
cDNA	Complimentary DNA
CSPD	Chemiluminescent substrate
CTAB	Hexadecyltrimethylammonium bromide
CZ	Central zone
DAPI	4'-6'-diamidino-2-phenylidole
DEPC	Diethyl pyrocarbonate
DNA	Deoxyribo-nucleic acid
dNTPs	Deoxynucleoside triphosphate mix
EDTA	Ethylene di-amino tetra-acetic acid
FUdR	5-fluoro-2'-deoxyuridine
GA	Gibberellic acid
GSP	Gene specific primer
HEPES	N-2-hydroxyethylpiperazine-N'-2-ethanesulfonic acid
ICAPB	Institute of cell and population biology
ICMB	Institute of cell and molecular biology
IPCR	Inverse PCR
IPTG	isopropylthio-β-galactoside
ITS	Internal transcribed spacer
kbp	kilo base pairs
knox	Knotted-like homeobox
L1	The outermost meristem layer
L2	The layer subtending the L1
L3	The innermost cell layers of the meristem
LB	L Broth
NEB	New England Biolabs
OLB	Oligonucleotide Labelling Buffer

PCR	Polymerase chain reaction
PZ	The peripheral zone of the meristem
QTL	Quantitative trait locus
RACE	Rapid amplification of cDNA ends
RBGE	Royal Botanic Garden Edinburgh
RNA	Ribo-nucleic acid
RZ	The rib zone of hte meristem
SAM	Shoot apical meristem
SDS	sodium dodecyl sulphate
SSC	sodium aline, sodium citrate
TALE	Three amino acid loop extension
TBE	Tris/ boric acid/ EDTA buffer
TE	Tris/ EDTA buffer
Tm	Melting temperature
TWEEN 20	polyethylenesorbitan monolaurate
UV	Ultra violet light
WT	Wild type
X-gal	5-bromo-4-chloro-3-indoyl-β-D-galactopyranoside

Chapter 1: Introduction

Molecular basis of differences between species

Morphological variation is a prerequisite of Darwinian evolution by natural selection. Variation may be continuous or discrete and species are differentiated and described on the basis of both. Continuous variation is thought to be generated by the action of many different loci together, whereas some discrete aspects of variation are associated with genetic differences at only one or two loci. Which aspects of genetic variation translate most directly to the production of phenotypic variation, and hence to the generation of new species?

Recent studies in evolution have suggested that an alteration in expression of transcriptional regulators, that control and co-ordinate the activity of other genes throughout development, can have a profound effect on morphology (Doebley and Lukens 1998, Stern 1998, Wang *et al.* 1999, Sucena and Stern 2000).

One class of transcriptional regulators that have been strongly implicated in morphological evolution in animals are the homeobox genes that encode homeodomain proteins. Homeodomain proteins are transcription factors that co-ordinate expression of genes necessary for the implementation of different developmental pathways (Scott 1999). The homeodomain is a 60 amino acid conserved region that unifies these proteins, comprising three alpha helices that bind DNA in a sequence specific manner. DNA binding specificities overlap between different homeodomain proteins, and to mediate the specificity of genes targeted, homeodomain proteins may act with co-factors. In animals, the co-factor can be another homeodomain protein. Some homeodomain proteins dimerise, and DNA is clasped between the two component proteins (Bürglin *et al.* 1997). This is illustrated by the interaction between *Ultrabithorax* (*Ubx*) and *Extradenticle* (*Exd*), two *Drosophila* homeodomain proteins belonging to Hox and PBC classes respectively and interaction is mediated by the YPWM sequence, a conserved motif upstream of the homeodomain (figure 1, overleaf; Passner *et al.* 1999, Piper *et al.* 1999).

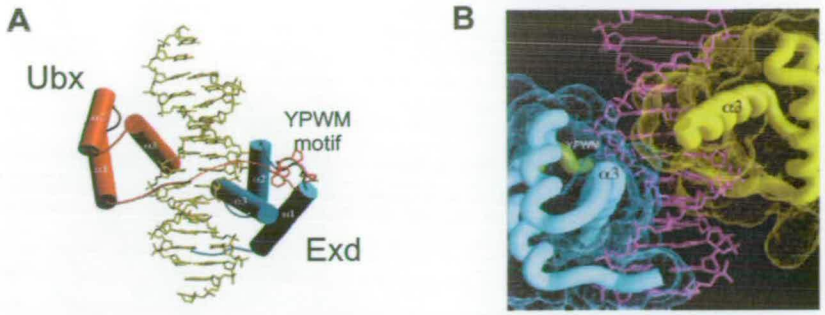


Figure 1: A: representation of the Ultrabithorax-Extradenticle-DNA ternary complex, illustrating the arrangement of the three alpha helices of the homeodomain in each protein, and the interaction between the proteins mediated by the YPWM motif. B: reverse view of A. Ubx; yellow, Exd; blue. Reproduced from Passner *et al.* (1999).

Due to their role in controlling genetic pathways, mutation at homeobox loci tends to lead to discrete and abrupt morphological change, such as production of a second thorax in *Ultrabithorax* mutants of *Drosophila* (Lewis 1985), or the production of a leg in place of an antenna in flies carrying dominant *Antennapedia* mutations (Schneuwly *et al.* 1987). Naturally occurring changes at homeobox loci are linked to major changes in metazoan body plan, including the divergence between insects and crustaceans (Averoff and Akam 1995, Ronshaugen *et al.* 2002). More subtle morphological variation, in patterns of hair growth that distinguish species of *Drosophila*, has also been linked to regulatory changes at homeobox loci (Stern 1998, Sucena and Stern 2000).

In plants, examples of how changes in gene regulation affect morphology in evolution are more limited. One example is provided by the domestication of maize, in which increased apical dominance, reduced lateral branching, and increased prevalence of ears on lateral branches, amongst other traits, were strongly selected for by man. Quantitative Trait Locus (QTL) mapping showed that one locus affecting differences in form in maize was *teosinte branched1* (*tb1*), a transcriptional regulator in the *cycloidea* gene family. *tb1* protein sequences are invariant between teosinte and maize, implying that changes in the traits selected for are linked to regulatory changes at this locus (Doebley *et al.* 1998, Doebley and Lukens 1998).

The effects of changes in homeodomain gene function are well documented in developmental studies in plants, but not evolutionary studies. Therefore this thesis aims to look at homeobox gene function in relation to the evolution of gross morphological differences between closely related plant species. *Streptocarpus* is a genus that

encompasses substantial variation, and its species relationships are well characterised in phylogeny.

Morphological differences between species of *Streptocarpus*

What is Streptocarpus?

Streptocarpus Lindley is a genus of about 130 species in the family Gesneriaceae (subfamily Cyrtandroidae) from Africa, Madagascar and the Comoro islands. The genus is united by its twisted fruits, after which it was named. *Streptocarpus* is morphologically a particularly diverse genus which comprises plants with very unusual growth forms. Broadly these are classed into two types, corresponding to the taxonomic split of *Streptocarpus* into two subgenera. Subgenus *Streptocarpella* includes caulescent species (Hilliard and Burtt 1971), and these species are united by a chromosome number of n=15 (Lawrence 1940). In subgenus *Streptocarpella*, opposite and decussate leaves arise from a shoot apical meristem, and inflorescences arise in leaf axils (figure 2a).

Novel growth forms in subgenus Streptocarpus

Species of subgenus *Streptocarpus* have radically different morphology, with basic chromosome number n=16 (Lawrence 1940, Hilliard and Burtt 1971). In subgenus *Streptocarpus*, some Madagascan species have conventional plant form; the leaves arise from a shoot apical meristem, and the inflorescences are axillary. Most members of the subgenus do not have conventional form, and their plant body is built in an entirely different way.

Phyllomorphic construction

In acaulescent species, the basic unit of construction shows properties of both stems and leaves, and due to this peculiarity is termed a phyllomorph. Phyllomorphs are comprised of a stem/petiole-like structure (the petiolode) that supports the lamina. Laminar growth is supplied by meristematic regions at the base and sides of the phyllomorph termed the basal meristems in previous studies (Jong 1970, Jong and Burtt 1975). Midrib and petiolode growth are supplied by an intercalary meristem in the petiolode. Inflorescences arise acropetally along the base of the midrib, in a large meristematic bulge called the groove meristem.

Adaptive value of acaulescence

The capacity for continued growth in leaves, production of roots from the petiolode, and formation of inflorescences at the junction of the lamina and the petiolode all distinguish

phyllomorphs from ordinary leaves (Jong 1970). Another way in which phyllomorphs differ from ordinary leaves is that they can lose part of their lamina. Within the Gesneriaceae, plants are suited primarily to moist, warm, and aseasonal climates (Burtt 1998). Most acaulescent *Streptocarpus* species grow in Southern Africa where there is a seasonal climate. In these species, an abscission line forms across the phyllomorph at the start of the dry season, and as the season progresses the distal part of the lamina dies. The thick fleshy veins may serve a water storage capacity (Burtt 1998). New laminar tissue is formed from the basal meristem when favourable conditions return.

Acaulescent species' form

Within *Streptocarpus*, phyllomorphic construction has been recruited to give different species widely divergent form (figure 2 B-E). In some species, termed unifoliate (figure 2 B-D), one phyllomorph comprises the entire photosynthetic body of the plant, and it can be up to 70 cm long (Hilliard and Burtt 1971). The phyllomorph arises by continued growth of one of two cotyledons, and the plants are monocarpic.

In other species, termed rosulate (figure 2E), successive phyllomorphs are iterated from the petiolode, so as it matures, the plant gains a rosette-like form. Each phyllomorph is monocarpic, but the plant is perennial. Some rosulate species have regular spiral arrangement of phyllomorphs around a central axis, and these are called centric rosulates, and in others (ex-centric rosulates) the arrangement appears haphazard (Jong 1970, 1978).

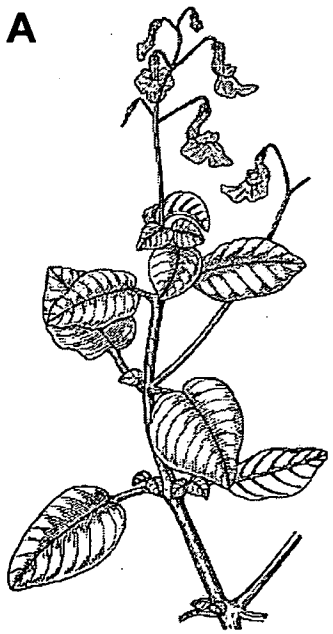
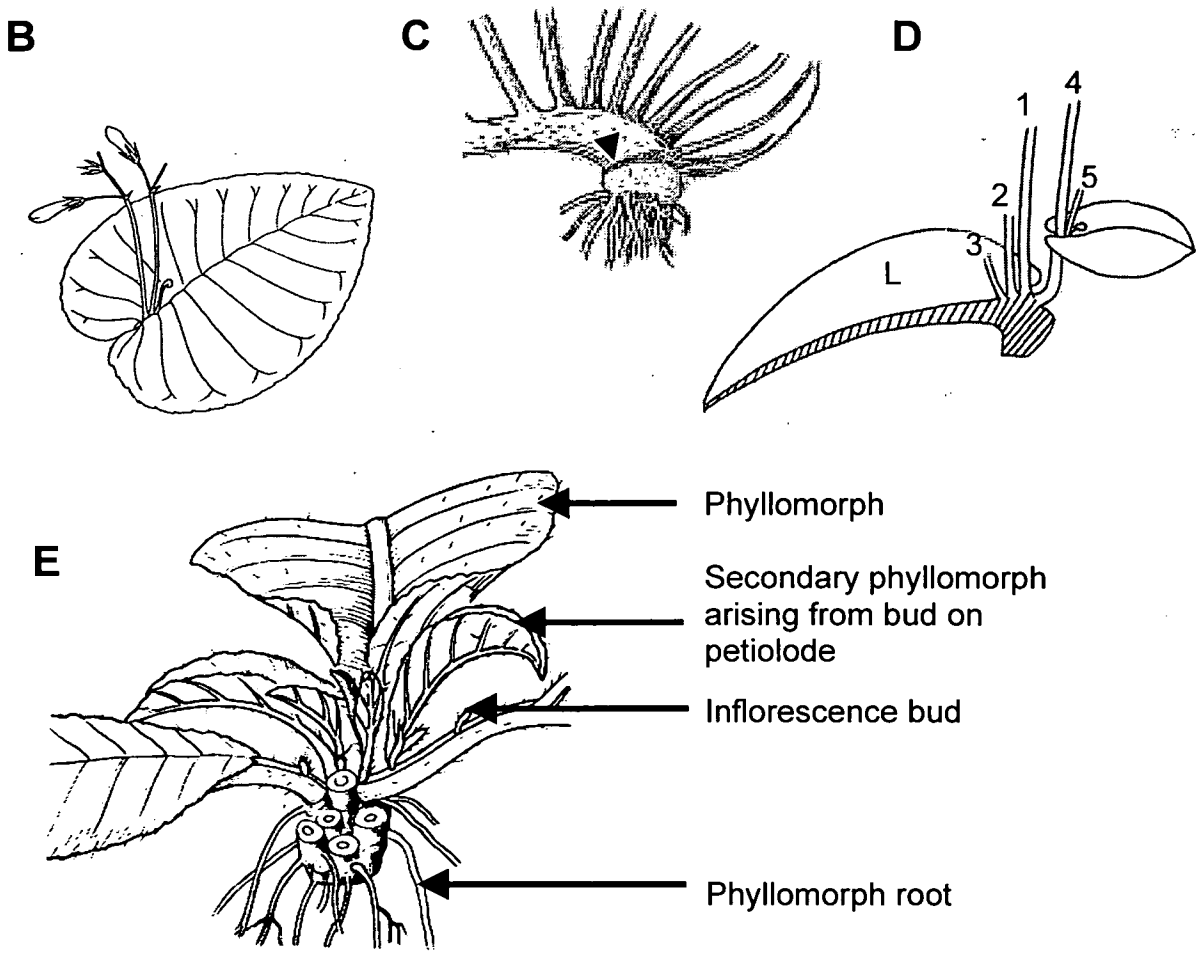


Figure 2: aspects of *Streptocarpus* morphology.
A: *S. holstii*, a caulescent species, showing opposite decussate phyllotaxy, and inflorescences arising from leaf axils. Picture taken from Hilliard and Burtt (1971).
B-D: phyllomorphic form exemplified by unifoliate species.
B: drawing to illustrate unifoliate form.
C: the base of a *S. grandis* phyllomorph showing location of the petiolode meristem (arrow) and roots growing from the petiolode.
D: diagram of bisected plant, showing inflorescences arising acropetally up primary phyllomorph. Shading indicates the cut surface, L indicates the lamina, Numbers indicate inflorescences, arising in sequence. B redrawn from Hilliard and Burtt (1971), C and D taken from Jong (1970).
E: *S. baudertii*, a centric rosulate species, showing iteration of secondary phyllomorphs from petiolodes. Redrawn from Jong (1970).



How differences in *Streptocarpus* form are attained

Ontogeny of Streptocarpus

Streptocarpus species are differentiated largely on the basis of differences in vegetative development, which reflect different patterns of phyllomorphic construction. How then do these differences in construction arise during development? Although there have been no studies of embryogenesis in *Streptocarpus*, there have been morphological studies on mature embryos and seedlings. Seeds are tiny (not more than 1.6 mm long and 0.5 mm wide), and bullet shaped (the embryo does not fold back on itself as in *Arabidopsis*). Germination of most *Streptocarpus* species is epigeal, although there are a few exceptions (figure 3A). Following emergence of the hypocotyl a ruff of unicellular epidermal hairs grows from its swollen base (figure 3B and C). The development of the primary root is delayed (Jong 1970).

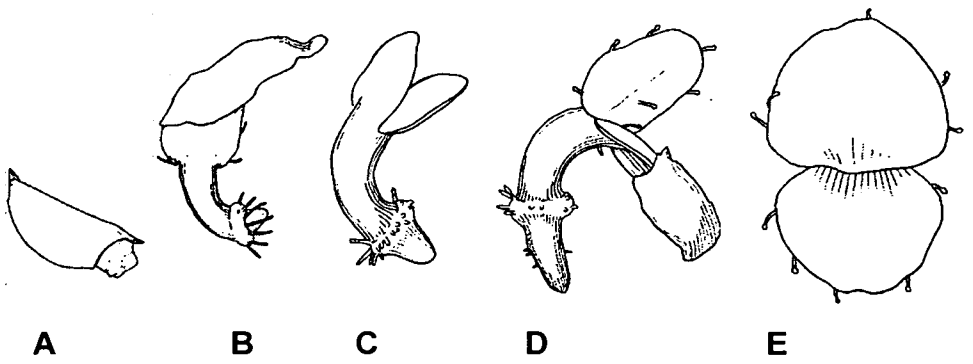


Figure 3: Germination and seedling development in *Streptocarpus*. A: Germinating *S. grandis* seed. B-E: *S. gardenii* seedling emergence and development. Scale bar =400μm, redrawn from Jong (1970).

Seedling development

In all species, when the cotyledons first start to grow they are of equal size, sessile and tightly adpressed, and there is no visible plumule (Burtt 1970, Ponniah 1971, Imaichi *et al.* 2000). In sections through mature embryos of unifoliate and rosulate species, there is little evidence of a presumptive shoot apical meristem (SAM) (Jong 1978, Imaichi *et al.* 2000). Gradually as growth of both cotyledons proceeds, one cotyledon becomes larger than the other, a phenomenon called anisocotyly (figure 4: Caspary 1858, Crocker 1860, Hill 1938, Oehlkers 1956, Burtt 1970). In another one-leafed member of the Gesneriaceae, *Monophyllaea*, surgical experiments suggested that anisocotyly is

When one of the cotyledons was removed before the onset of anisocotyly, the other always continued to grow. If removal was after the onset of anisocotyly, continued growth only occurred if the smaller cotyledon was removed (Tsukaya 1997). After the onset of anisocotyly in *Streptocarpus* the smaller of the two cotyledons may wither away. The adaptive significance of anisocotyly and delayed plumular development may be that they confer a greater photosynthetic capacity on tiny seedlings when they first germinate earlier than would otherwise be possible, compensating for the lack of reserves in the seed. The larger of the two cotyledons has a capacity for continued growth, and assumes different fate in the different growth forms. Anisocotyly may have predisposed the Gesneriaceae to evolve such unusual growth forms (Burtt 1970, Burtt 1994).

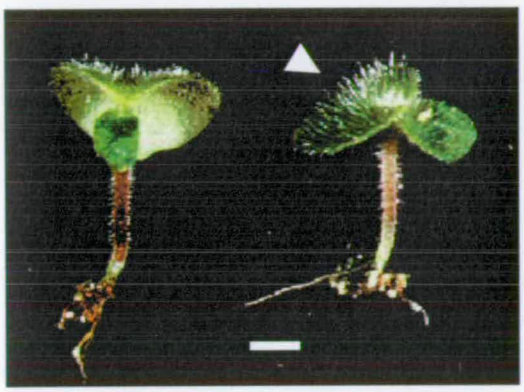


Figure 4: Anisocotyly in *Streptocarpus* seedlings. Arrow indicates larger cotyledon, scale bar ~1 mm, photo ©RBGE.

Meristem establishment

SAM establishment in caulescent species

Following the onset of unequal growth, the larger of the two cotyledons is often raised above the smaller by a stalk of variable size, termed the cotyledonary petiolode (Jong 1970). In all species looked at so far, a region of densely staining cells arises from this region at an angle to the petiolode, and displaced towards the larger cotyledon. In caulescent species this develops into a conventional apical bud, with layered and zonal organisation. This SAM forms leaf primordia in pairs, the first of which is decussate with respect to the cotyledons.

Groove meristem establishment in acaulescent species

In acaulescent species, this region is the source of the groove meristem. As for the caulescent species, a pool of densely staining, dividing cells becomes apparent after the onset of anisocotlylous growth, which is displaced towards the larger cotyledon and is at an incline with respect to the petiolode. The meristem remains at the junction between the lamina and the petiolode as the cotyledonary petiolode grows, eventually becoming concealed in a groove between the two blades. It is therefore termed the groove meristem and has conventional meristem structure, with discernible outer layers and an inner corpus. In unifoliate species the groove meristem gives rise to inflorescences and after flowering the plant dies (Jong and Burtt 1975, Imaichi *et al.* 2000).

In rosulate species the densely staining pool of cells arises as for the other main growth forms, but then splits, the smaller distal part giving rise to a groove meristem as in unifoliate species. The larger proximal part forms a large phyllomorph primordium, and iterates further phyllomorphs throughout the life of the plant. It also has an L1, L2, corpus arrangement, resembling a conventional SAM (Jong 1970, Jong 1978).

Basal and petiolode meristem structure

Densely staining regions of cells that are actively dividing, and that were termed basal and petiolode meristems by Jong (1970) and Jong and Burtt (1975), supply laminar and petiolodar tissue to the phyllomorph. The 'basal meristems' are present initially in both cotyledons, and also supply laminar growth to secondary phyllomorphs, but do not have discernible cell layers (Jong 1970, Jong and Burtt 1975, Tsukaya 1997, Imaichi *et al.* 2000). 'Petiolode meristems' form as a ring at the boundary between the petiolode and the midrib. They resemble intercalary meristems of other flowering plants in that they supply growth to the stem-like petiolode.

Inheritance of form

Acaulescent species of *Streptocarpus* have common chromosome numbers, and some species form fertile hybrids when crossed; hybridisation has probably been an important factor in speciation in *Streptocarpus* (Hilliard and Burtt 1971). In the last century Oehlkers (1938, 1940, 1941, 1964) investigated the genetic basis of morphology in subgenus *Streptocarpus* more formally. In crosses between rosulate and unifoliate species, he found evidence that two dominant genes, one coming into effect before the other, determined the rosulate form (Oehlkers 1938, 1941, 1964), and this evidence is examined further in chapter 3.

Manipulation of morphology

Differences in form between species can also be manipulated experimentally. In some species of the Gesneriaceae (*S. nobilis*, *Chirita* spp., *Rynchoglossum* spp.) that are usually caulescent, phyllomorphic structure can be induced if the plants are grown in short days rather than long days (Burtt 1970, Ponniah 1971). In two unifoliate species formation of supplementary phyllomorphs can be induced by exogenous application of gibberellic acid (GA)(Dubuc-Lebreux 1978). Also, in some phyllomorphic species, a stem of fairly conventional morphology can be induced if seedlings are grown on a GA supplemented medium, although leaf phyllotaxy may be disrupted (Rosenblum and Basile 1984). In acaulescent species, plumule development can also be induced experimentally by coating the cotyledons with a seal, and this effect has been attributed to a disruption in auxin transport brought about by trapped ethane (Rosenblum and Basile 1984).

In species from both subgenera, meristem formation has been demonstrated on leaf explants. In *S. rexii* (acaulescent) explants, meristems formed by division of basal cells of glandular hairs, and in some cases lead to flowering growth, but the form of the plants was not specified (Viémont 1980). A similar phenomenon has been reported for African violets (*Saintpaulia*), which form a monophyletic group nested within *Streptocarpus* (Möller and Cronk, 1997a). In African violet explants, cells adjacent to glandular hair basal cells begin to divide, and these cells eventually form conventional SAMs capable of iterating shoots and leaves (Ohki 1994).

Phylogeny of *Streptocarpus*

Origin of acaulescent species

The origin of acaulescence within the Cyrtandroidae has been investigated in morphological and phylogenetic studies. These suggest that acaulescence originated several times in disparate genera within the subfamily, and that caulescence is plesiomorphic (Möller and Cronk 2001a). Within *Streptocarpus* the situation is not quite so clear; while most acaulescent species probably had caulescent ancestors, it is also probable that some caulescent species had acaulescent ancestors (Hill 1938, Hilliard and Burtt 1971, Möller and Cronk 2001a). A example of this is *S. decipiens* which falls in a clade of South African unifoliate and rosulate species in ITS sequence phylogeny, and forms an abscission line on its cotyledonary phyllomorph like African acaulescent

species, but develops a shoot system with opposite and decussate leaves (Hilliard and Burtt 1971, Möller and Cronk 2001a).

Relationships between the different forms of *Streptocarpus* based on ITS sequence data and phylogeny are summarised from Möller and Cronk's (2001a) analysis in figure 5, overleaf. By comparison to outgroup taxa this shows that caulescent form is plesiomorphic with respect to the genus as a whole, but within subgenus *Streptocarpus* the plesiomorphic state is not clear: the plesiomorphic groups of the subgenus *Streptocarpus* clade contain species of varied form. The acaulescent species are paraphyletic, and there are several independent origins of unifoliate and rosulate forms from each other.

The centre of diversity for *Streptocarpus* is Madagascar where there are many forms that deviate from the three main categories mentioned here, and it is possible that this is where *Streptocarpus* arose (Möller and Cronk 2001b). Representatives from both subgenera occur in both Africa and Madagascar, and it is unlikely that this distribution arose by seed dispersal as the seeds are so tiny. It is more likely that the genus colonised Madagascar by a land bridge around the mid Eocene around 46-25 million years ago, in conjunction with many animal species (McCall 1997, Möller and Cronk 2001b).

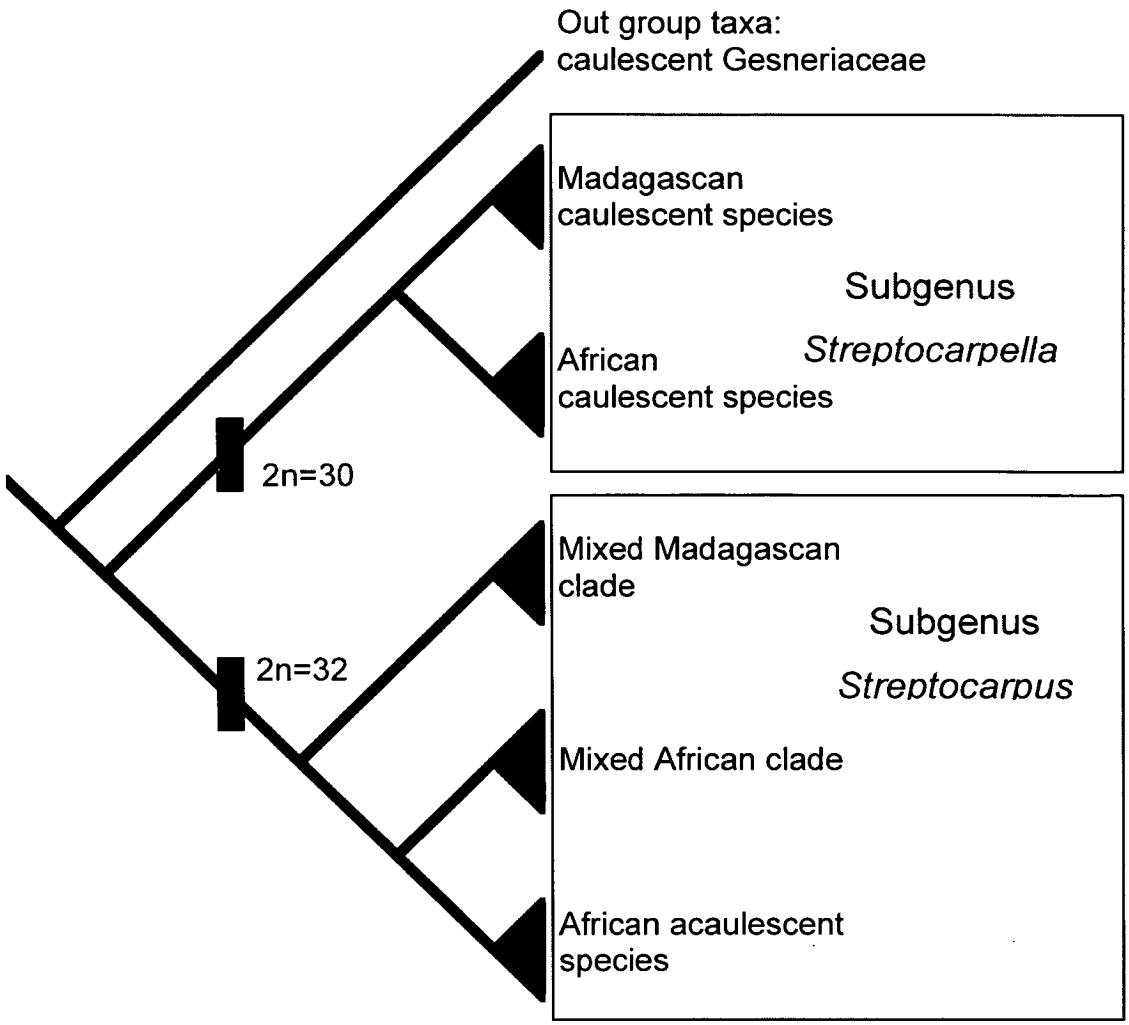


Figure 5: Main clades of *Streptocarpus* shown in an ITS sequences phylogeny (Möller and Cronk 2001a).

Developmental and genetic bases of differences in form

Genetic interactions in model species' meristems

The location of meristematic regions, behaviour of meristems and growth of acaulescent *Streptocarpus* is highly unusual. In model species mutagenic and genetic studies have led to the isolation of many gene families that affect meristems in different ways and could therefore have had a role in the evolution of novel growth forms in *Streptocarpus*. Some of these have known roles in establishing the meristem during embryogenesis in model species.

Meristem establishment during Capsella *Embryogenesis*

In angiosperms, fertilisation is double: one male nucleus fuses with the two polar nuclei to form the triploid endosperm, and the other fertilises the female gamete. In *Capsella,* once the endosperm is established the zygote undergoes a series of stereotypical cell divisions outlined in figure 6. The first of these is an unequal, transverse division, producing a smaller terminal cell, and a larger basal cell. The basal cell divides transversely again, and the two cells resulting from this division divide further to form the suspensor. The terminal cell divides to form the embryo proper in several stages. During the globular stage the terminal cell cleaves to give a 2-, 4-, 8-, 16-, and 32- celled roughly spherical embryo. Dermal, ground and vascular tissues begin to differentiate. During the heart stage differential cell division gives rise to two cotyledon primordia at the end away from the suspensor. Elongation of the embryo axis and cotyledons occurs, and the suspensor is crushed as the embryo grows (Schultz and Jensen 1968a, 1968b).

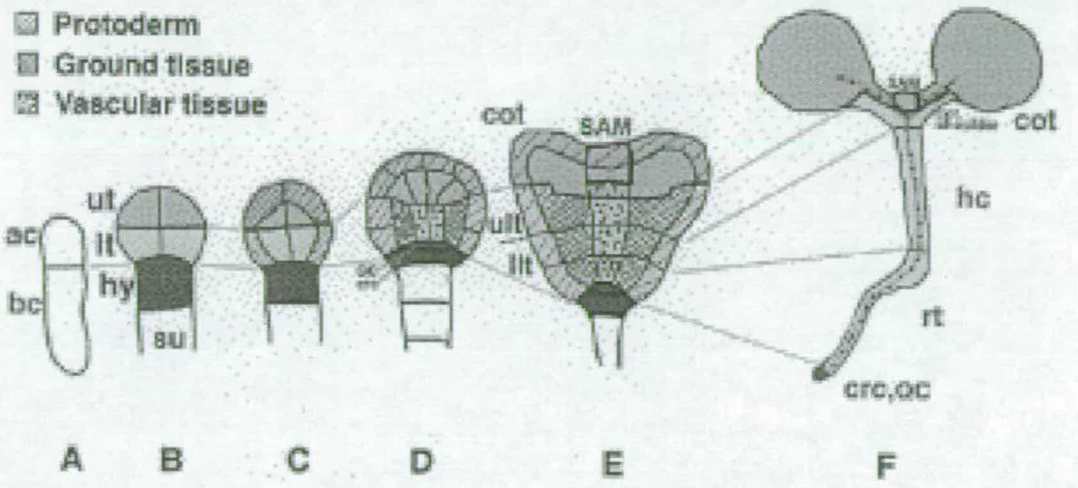


Figure 6: Stages of embryonic development in *Arabidopsis thaliana.* A: one-celled stage, B: eight celled stage, C: dermatogen stage, D: late globular stage, E: heart stage, F: seedling. Taken from J. Haseloff's website (www.plantsci.cam.ac.uk/Haseloff/teaching).

Genes required for establishment of the SAM

The SAM first becomes histologically apparent in the late heart-torpedo stage embryo, but genetic markers for meristem activity indicate that it initiates during the globular stage. The cotyledons also initiate during the globular stage, and it is not yet clear whether they derive from SAM cells (the cotyledons would then be the first lateral organs initiated by the SAM), or are independently derived.

Embryonic expression of genes marking meristem identity

Genes whose expression is necessary for establishment of the SAM during *Arabidopsis* embryogenesis have been revealed by loss-of-function mutants that do not form a SAM, or whose SAM differentiates prematurely (figure 7). These include *pinhead/zwille* (*pnh/zll*: McConnell and Barton 1995, Moussian *et al.* 1998, Lynn *et al.* 1999, 7 A-D), *wuschel* (*wus*: Laux *et al.* 1996, 7 H), *shootmeristemless* (*stm*: Barton and Poethig 1993, 7 F, H, L, M), *cup-shaped cotyledon1* (*cuc1*: Aida *et al.* 1999, Takada *et al.* 2001) *cup-shaped cotyledon2* (*cuc2*: Aida *et al.* 1999), and *argonaute* (*ago*: Bohmert *et al.* 1998, Lynn *et al.* 1999). *PNH/ZLL* appears to be first expressed earlier than other meristem determining genes throughout the 4 cell embryo. As the embryo develops, different levels of expression are seen in different parts. Intense expression is detected in the hypodermal cells, which mark vascular precursors. Reduced expression marks the adaxial sides of the presumptive cotyledons and shoot apical meristems. Due to its early expression it has been suggested that *PNH/ZLL* may confer embryonic competence to initiate a SAM (Moussian *et al.* 1998, Lynn *et al.* 1999). *WUS* expression appears next, in the apical hypodermal cells of the 16 cell embryo. As embryonic cells continue to divide, expression is confined to a single inner cell layer at the centre of the apex. Once the meristem is established, *WUS* expression is confined to a small group of cells in the L3 of the central zone (Mayer *et al.* 1998). The earliest evidence of expression of the remaining genes is in the globular embryo, and their expression appears to initiate at about the same stage (Aida *et al.* 1999, Long *et al.* 1996, Long and Barton 1998, Moussian *et al.* 1998, Lynn *et al.* 1999). *CUC1* expression is first detected in two domains perpendicular to the presumptive cotyledons. *CUC2* is first expressed in the central protodermal cells in the apical half, and its expression spreads to form a strip between the two cotyledons by the heart stage. By the torpedo stage, expression is lost from the central region, and forms a band around the SAM (Aida *et al.* 1999). Jointly, *CUC1* and *CUC2* expression are required for establishment of *STM* expression. *STM* is first expressed in a single off centre hypodermal cell in the apical half of the globular embryo. As the embryo develops, expression extends to comprise a strip between the cotyledons, and through the outer cell layers (Long and Barton 1998). *AGO* is expressed evenly throughout the globular embryo, and expression is also detected in the suspensor (Bohmert *et al.* 1998, Lynn *et al.* 1999).

Embryonic expression of genes marking lateral identity

Expression of genes implicated in organ formation and lateral identity is in complementary domains to the meristem determining genes. Expression of *ASYMMETRIC LEAVES1* (*AS1*), and other genes marking organ identity is first detected in two hypodermal foci, which correspond to the site of cotyledon initiation. Expression continues to mark lateral organ identity throughout development (Byrne *et al.* 2000). In *stm-1* mutants, *AS1* is expressed throughout the apical part of the embryo, blocking SAM development, and it appears that one role of *STM* is to repress *AS1* expression. (Byrne *et al.* 2000).

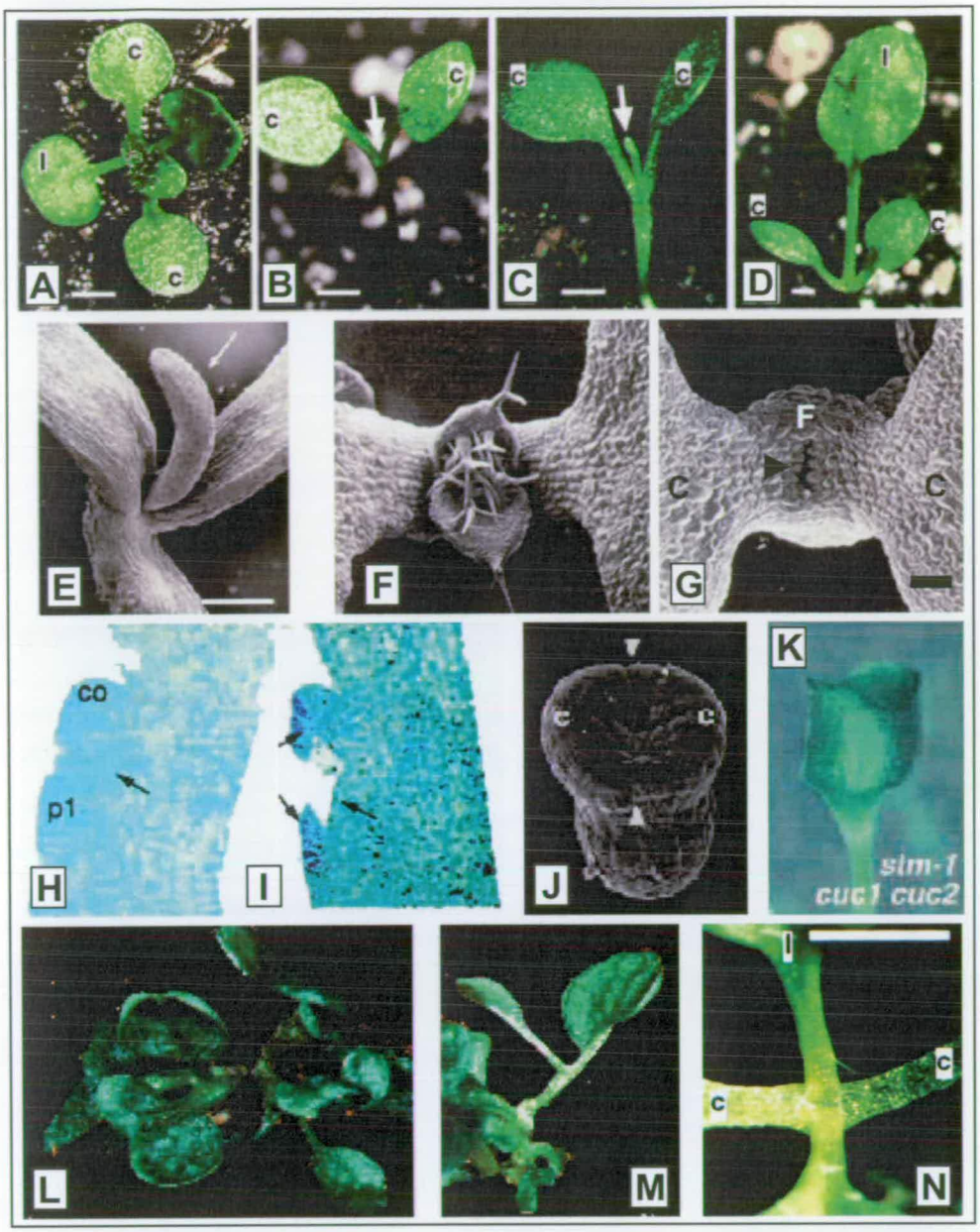


Figure 7: Mutants that do not fully establish a SAM. A: wild type *Arabidopsis* seedling, showing cotyledons (c), with shoot and leaves (l) emerging between them. B-D: *pinhead/zwille* mutant seedlings showing lack of SAM formation (B), or termination of the meristem in a filamentous structure (C) or a single leaf (D), reproduced from Moussian *et al.* (1998). E: *argonaute* mutant seedling, showing termination of the SAM with a single leaf-like organ, reproduced from Lynn *et al.* (1999). F: SEM of wild type *Arabidopsis* seedling, as in A. G: strong loss-of-function *stm* mutant with no SAM (arrow) and fused (F) cotyledons (C) reproduced from Long *et al.* (1996). H: section through wild- type maize embryo, showing the embryonic SAM (arrow) between the coleoptile (co) and first leaf primordium (p1). I: section through a maize *kn1* mutant embryo, showing the coleoptile and leaf primordium as in H, but no bulging SAM between them. Reproduced from Vollbrecht *et al.* (2000). J: SEM of *cuc1/cuc2* double mutant embryo, showing initiation of cotyledons (c) in a ring. K: *cuc1/cuc2/stm* mutant seedling, showing cup shaped phenotype that results from fusion of the cotyledons. J and K reproduced from Aida *et al.* (1999). L, M: weak loss-of-function *stm* mutants reiterating leaf-like organs from the hypocotyls (Clark *et al.* 1996). N: *wus* mutant seedling with apical development arrested after the formation of two leaves (Mayer *et al.* 1998).

Developmental zones established in the embryo

The combined expression patterns of these genes delimits different developmental domains within the globular embryo. The first of these is a central strip which gives rise to the SAM and is marked by *CUC2* and *STM* expression. This is flanked by peripheral domains marked by *AS1* expression from which the cotyledons form. A boundary between domains is marked by overlapping expression of *CUC2*, *STM*, and *AS1*.

Meristem structure

Expression of all of the genes mentioned above continues throughout plant development, and implicates them in the regulation and maintenance of meristems. After their establishment during embryogenesis, meristems maintain the capacity to form new organs, and to renew themselves throughout plant development. This is reflected in their structure (figure 8). In previous studies, the meristem has been subdivided in two ways. The first of these separates groups of cells according to their rate of division and their cytoplasmic density into three zones (figure 8a). These are (1) the central zone, which contains a self-renewing pool of stem cells, and supplies cells the other zones, (2) the peripheral zone, where cells are recruited into developing organs, and (3) the rib zone, which is the origin of ground tissue within the stem, and subtends the other zones. The second way by which dicot meristems have been subdivided is on the basis of layers of clonally related cells, the outermost L1, middle L2, and innermost L3 (figure 8b); in monocots there may only be two layers. It is not yet known whether these subdivisions of the meristem reflect functional differences between parts, but they provide a useful point of reference.

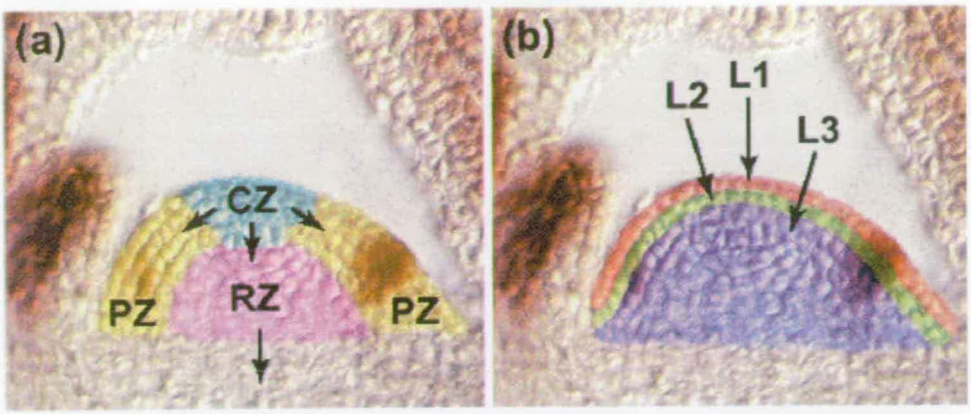


Figure 8: Structure of an *Arabidopsis* SAM (taken from Bowman and Eshed 2000). CZ=central zone, PZ=peripheral zone, RZ=rib zone. L1=outermost layer of clonally related cells, L2= middle layer of clonally related cells, L3 = inner layer of clonally related cells. L1+L2=tunica, L3=corpus. Brown staining indicates expression of *FILAMENTOUS FLOWER*, which is restricted to lateral organs and their precursors (Sawa *et al.* 1999, Eshed *et al.* 1999).

Meristem maintenance and organogenesis during post embryonic development

Organogenesis

In post-embryonic development, meristems are characterised partly by repetitive iteration of lateral organs from their flanks. Iteration of lateral organs is marked by down regulation of *STM* expression at the flanks of the meristem, and expression of *AS1* in organ primordia. *AS1* is therefore implicated in establishment of lateral fate, and in determinacy. Homologues of *AS1* in other species are *rough sheath2* (*RS2*) in maize and *PHANTASTICA* (*PHAN*) in *Antirrhinum*. Collectively they are termed the ARP genes. Both *RS2* and *PHAN* appear to function similarly to *AS1*; again expression marks lateral fate (Waites *et al.* 1998, Timmermans *et al.* 1999, Tsiantis *et al.* 1999). *PHAN* is also implicated in dorsoventral asymmetry in the leaf: some leaves of *phan* mutants are radial and ventralised.

Meristem maintenance: indeterminacy

Conversely to the ARP genes, *STM* expression provides a marker for meristem identity and seems involved in maintaining indeterminacy. Double mutants for *as1/stm-1* form a comparatively normal shoot, indicating that *STM* expression is only required for meristem maintenance in the presence of *AS1* expression. In the absence of both, indeterminacy may be maintained by other *STM*-like genes, which are also down

regulated at the site of organ initiation (Byrne *et al.* 2000). *STM* is a homeodomain protein, and in animals the roles of homeodomain proteins are better characterised. Further possible roles for *STM* and other plant homeodomain proteins are discussed from page 36.

Boundaries between organs and the meristem

Organs are produced in an ordered pattern and are physically separate from each other and from the meristem. This separation is absent in *cuc1/cuc2* mutants, and *No Apical Meristem* (*NAM*) mutants of *Petunia*, which have fused cotyledons (Aida *et al.* 1999, Souer *et al.* 1996). The genes affected by these mutations all encode similar proteins, the NAC domain proteins. In adult development in *Arabidopsis, CUC1* expression in a strip marks the boundaries between organs and meristems, and meristems with different fate (inflorescence versus floral meristems: Takada *et al.* 2001). In *Petunia*, *NAM* expression also marks the boundaries between the meristem and organ primordia (Souer *et al.* 1996).

Maintenance of a stem cell population and size

Another feature of meristems is that they maintain their size, and this has to be achieved by balancing loss of cells to the peripheral zone for tissue formation with replenishment of cells within the central zone throughout the life of the plant. Several mutants which affect meristem size have been isolated.

Mutants with reduced meristematic activity

One of these, whose implication in embryonic meristem formation has already been discussed, is *WUS*. *wus* mutants have small meristems in comparison with WT *Arabidopsis*. In strong *wus* mutants, development arrests after seedlings have iterated the first two leaves and the stem cells of the CZ differentiate. In weaker mutants, development continues to flowering, but floral development terminates prematurely in the formation of a single central stamen (Laux *et al.* 1996, Mayer *et al.* 1998).

Mutants with enlarged meristems

A converse phenotype is seen in the *clavata* mutants (*clv1*, *clv2*, and *clv3*) which all have enlarged meristems and club-like siliques. This enlargement has been interpreted as a result of an increase in the number of stem cells accumulating in the apex. This could either result from an increase in rates of cell division, or a decrease in rates of cell recruitment into other compartments (Clark *et al.* 1993, Clark *et al.* 1995, Clark *et al.*

1997, Laufs *et al.* 1998, Fletcher *et al.* 1999, Jeong *et al.* 1999). In comparison to WT meristems, cell division is slowed in the CZ of *CLV* mutants (Laufs *et al.* 1998), which favours the latter hypothesis.

CLV1 and *CLV3* encode a receptor kinase and its small peptide ligand respectively (Clark *et al.* 1997, Fletcher *et al.* 1999), and *CLV2* encodes a protein with a kinase domain, but no CLV1-like receptor domain, thought to form a complex with *CLV1* and *CLV3* (Jeong *et al.* 1999). Receptor-like kinases and associated proteins are commonly components of signalling cascades. It is therefore thought that the *CLV*-genes are involved in a signalling pathway that restricts the size of the central zone.

wus mutations are epistatic to *clv* mutations; double mutants have a *wus* phenotype. *wus* and *clv1/3* mutants show complementary phenotypes, and are expressed within complementary domains of the meristem (figure 9). Ectopic *WUS* expression is sufficient to induce a *clv* phenotype, and *WUS* can induce *CLV3* expression. It is therefore likely that *CLV1/3* and *WUS* act on the same pathway, with *WUS* promoting stem cell identity, and *CLV1/3* countering this promotion by repressing its expression (Clark *et al.* 1993, Clark *et al.* 1995, Clark *et al.* 1997, Laufs *et al.* 1998, Schoof *et al.* 2000).

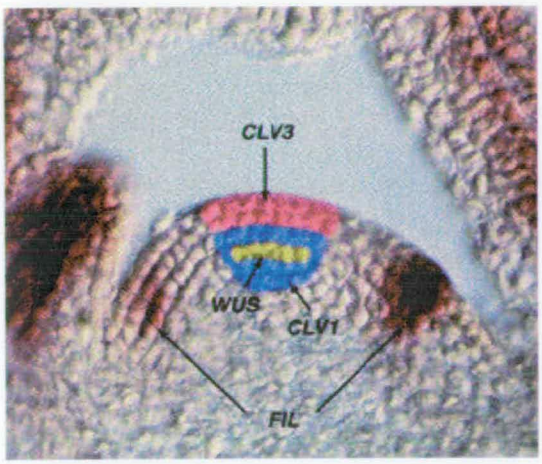


Figure 9: Expression domains of *CLV1*, *CLV3*, and *WUS* in the shoot meristem. Taken from Bowman and Eshed (2000).

A further gene which affects meristem size in the *CLV/WUS* pathway is *POLTERGEIST* (*POL*). The *pol* single mutant phenotype is similar to WT, but *pol* suppresses the weak *clv1* phenotypes in double mutants. *pol* also partially represses *clv2/3* phenotypes in

double mutants. *pol* enhanced *wus* phenotypes, suggesting that it functions redundantly with *WUS* in promoting stem cell identity (Yu *et al.* 2000).

Genes that affect phase changes in model species

In flowering plants, the meristem behaves differently at different stages of development. During embryogenesis the meristem and cotyledons are established, during vegetative development the meristem iterates leaves and is self perpetuating, and during reproductive development apical meristems switch from making leaves to making bracts and floral organs, and they may become determinate. Vegetative development may also comprise a juvenile and an adult phase. Therefore development is characterised by distinct phases of growth, in which the meristem behaves differently.

In model species there are many mutants with defective phase change at different stages in their development. One example is the *embryonic flower* (*emf*) mutant of *Arabidopsis* in which the SAM initiates as an inflorescence meristem, forgoing vegetative development (Sung *et al.* 1992, Yang *et al.* 1995). As these mutants express characteristics appropriate to one stage of development during another stage of development, they are termed heterochronic mutants. Heterochronic mutations do not disrupt specific aspects of morphology, but can effect total body plan without impairing an organism's viablity. They are therefore considered to have broad scope for generating evolutionary novelty (Gould 1977). This may be particularly relevant to *Streptocarpus* as, at least in the unifoliate species, the main body of the plant develops from an embryonic structure; the cotryledon.

Genes that could be implicated in Streptocarpus evolution

In model species there are now many genes that are known to affect meristem behaviour, and morphological change in *Streptocarpus* potentially implicated many of these. *Streptocarpus* homologues of some genes involved in meristem function were cloned as a tool to investigate meristems and morphological change in *Streptocarpus*. When I started my PhD, the *STM*-like genes provided the best characterised marker for meristem identity in other species, one reason that this study focussed on *STM*. Before discussing further reasons for focussing on *STM*, I will introduce further background on *STM*, related genes, and their functions.

Structure and function of knox homeodomain proteins

STM encodes a homeodomain protein. Although homeodomain proteins all have a DNA binding domain comprised of three alpha helices, they also have structural differences, and therefore fall into several classes. One class is the Three Amino acid Loop Extension (TALE) superclass, characterised by a 63 amino acid homeodomain, rather than the typical 60 amino acid homeodomain, and TALE proteins are found in yeast, worms, insects, mammals and plants. The first family of plant homeobox genes to be found belong to the TALE superclass, and they are the *Knotted1*-like (*knox*) genes (Vollbrecht *et al.* 1991). *Knox* genes also share several highly conserved domains outwith the homeodomain. From N to C terminus, the conserved domains of *knox* genes are: (1) the MEINOX domain which is split into KNOX1 and KNOX2 sub-domains, and may be required for protein dimerisation, providing some evidence that knox proteins may function in plants similarly to homeodomain proteins in animals (Bürglin 1997, Bürglin 1998, Bellaoui *et al.* 2001, Nagasaki *et al.* 2001) (2) the GSE box, (3) the ELK domain, and (4) the homeodomain (Kerstetter *et al.* 1994, Bürglin 1997, Bürglin 1998). The function of GSE and ELK domains are not yet clear.

Knox genes are themselves separated into two classes on the basis of structural differences. The Class 2 genes have an intron in the ELK domain, which is absent from Class 1 genes. Class 1 and Class 2 genes also have different expression patterns, and probably differ in their developmental roles (Kerstetter and Hake 1997, Hake *et al.* 1995, Bharathan *et al.* 1999, Reiser *et al.* 2000). Class 2 genes show delocalised expression that may be ubiquitous, and as there are currently no mutants in Class 2 *knox* genes known, their role(s) remain(s) unclear (Reiser *et al.* 2000).

Class 1 *knox* gene expression marks meristem identity

In almost all flowering plants studied so far, Class 1 *knox* genes provide markers for meristem identity. They are expressed in different sub-domains of the meristem and developing internodes but, with the exception of the compound leaves of tomato, are not expressed in lateral organs from when they first start to develop at the flanks of a meristem (Jackson *et al.* 1994, Lincoln *et al.* 1994, Hareven *et al.* 1996, Nishimura *et al.* 1999, Sentoku *et al.* 1999).

Loss-of-function Class 1 *knox* mutants

Loss-of-function mutants in Class 1 *knox* genes can affect meristem initiation and maintenance (Barton and Poethig 1993, Kerstetter *et al.* 1997, Reiser *et al.* 2000,

Vollbrecht *et al.* 2000). In strong *stm* mutants, no apical meristem forms, and the cotyledons are fused (figure 7 G). In weak *stm* mutants, organs form initially from the region of cotyledonary petiolar fusion (figure 7 L, M). Subsequent leaves are formed from the petioles of these leaves, in the region that would be an axillary meristem in a true leaf, and these leaves are terminal (Clark *et al.*1997). In strong *kn1* loss-of-function mutants in maize, phenotypes are more variable and depend on the genetic background. Strong mutations in some backgrounds again result in absence of a shoot meristem in the embryo (figure 7 H, I). As there is always some evidence of leaf initiation, it is not clear whether the meristem establishes and terminates prematurely, or fails to establish (Kerstetter *et al.* 1997, Vollbrecht *et al.* 2000). Defects in these two mutants can be interpreted as resulting from a loss of indeterminacy in meristematic tissues.

Gain-of-function *knox* mutants

Mis-expression of Class 1 *knox* genes in leaves disrupts leaf development (Smith *et al.* 1992, Lincoln *et al.* 1994, Chuck *et al.* 1996, Hareven *et al.* 1996 Frugis *et al.* 1999, Muehlbauer *et al.* 1999. Byrne *et al.* 2000, Pautot *et al.* 2001, Golz *et al.* 2002). Phenotypes associated with *knox* expression in leaves are diverse (figure 10), but there are some common themes. These are (1) blurring of the boundary between proximal and distal regions of the leaf (figure 10b, c, e-h), (2) gnarling, knotting, and wrinkling of the leaf (figure 10b, c, e-g, h, k, m), (3) ectopic formation of meristems or shoots from leaves, (4) reduced apical dominance.

In some dicots *knox* expression in leaves is also associated with leaf lobing (figure 10d-f, h) seen in *Arabidopsis*, lettuce and tobacco plants with 35S::*KNAT1* expression, *KNAT2* mutants, and *as1* mutants, and there also appears to be a marginal effect in *Hirz* mutants (Lincoln *et al.* 1994, Chuck *et al.* 1996, Frugis *et al.* 1999, Byrne *et al.* 2000 Golz *et al.* 2002).

In tomato, 35S::*Kn1* expression leads to a much higher degree of compounding than in wild type leaves (figure10 i and j: Hareven *et al.* 1996).

Taken together, these effects can be interpreted as a loss of determinacy and differentiation in leaves, the converse to what is seen in two loss-of-function *knox* mutants.

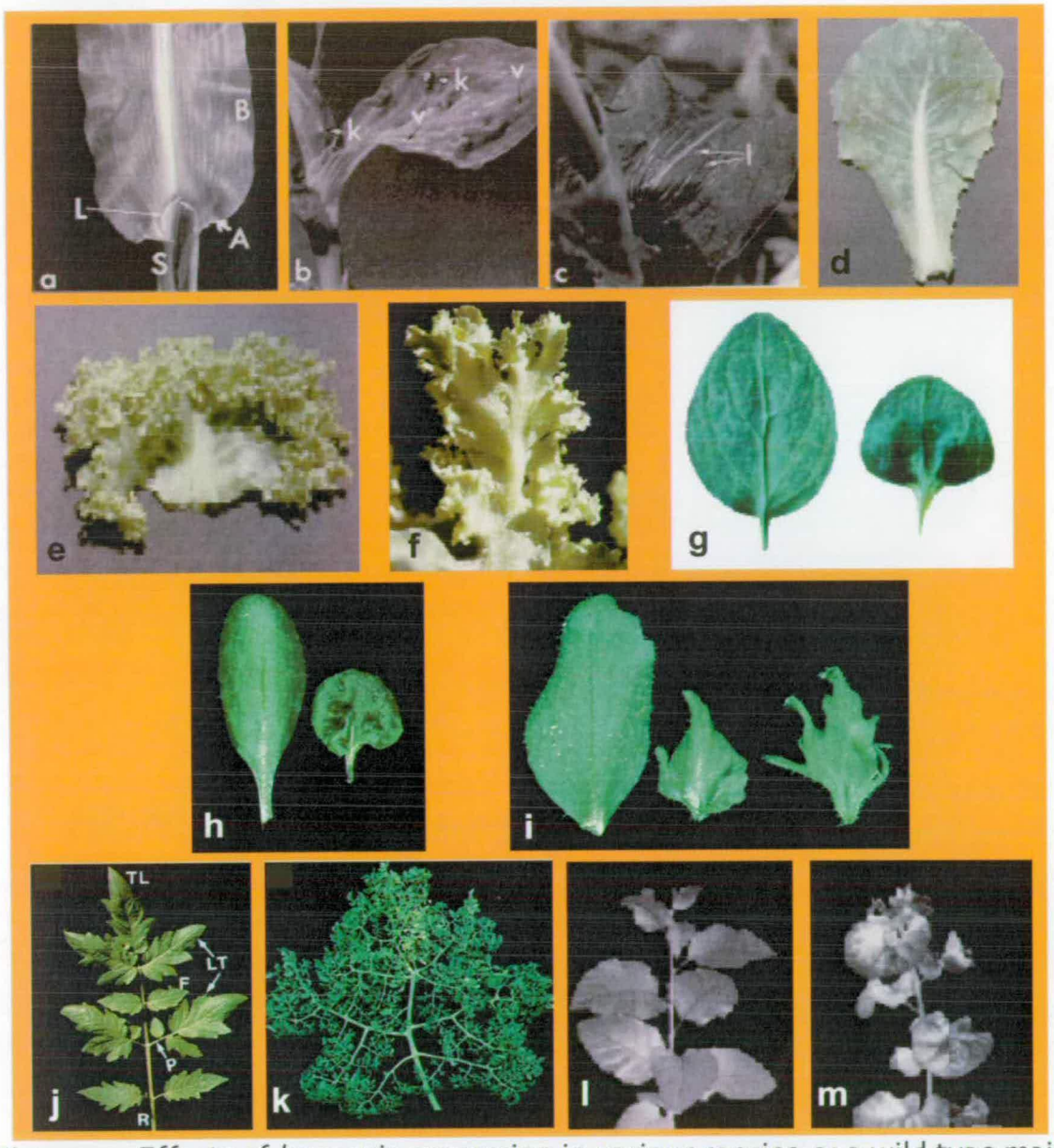


Figure 10: Effects of *knox* mis-expression in various species. a: a wild type maize leaf showing sheath (S), ligule (L), auricle (A) and blade (B). b and c: maize *KNOTTED* mutant leaves showing gnarling caused by ectopic knotted expression in the leaf blade, and displacement of the sheath/laminar boundary. Knots (K) are may be localised to veins (V). d-f: *knox* mediated morphological changes in lettuce (*Lactuca sativa*). d: wild-type lettuce leaf, e-f plastocyanin::*KNAT1* expressing lettuce leaves reproduced from Frugis *et al.* (1999). g: Wild type (left) and *Hirz* (right) mutant leaves from *Antirrhinum*. In *Hirz* mutants an *STM*-like gene is expressed in the leaf (Golz *et al.* 2002). h, k: Effects of the *as1* mutation in *Arabidopsis*. h: wild type (left) and *as1* (right) rosette leaves, and i: *as1* cauline leaves. In *as1* mutants, *KNAT1* and *KNAT2* are misexpressed in the leaf, and the leaves are rumpled and lobed in later development. Sometimes ectopic meristems form on the leaves. Reproduced with permission of R. Barley (University of York). j, k: wild type (left) and ultra-compounded 35S::*Kn1* expressing tomato leaves. Reproduced from Hareven *et al.* (1996). L, m: wild type (left) and 35S:: *NTH1* (right) tobacco plants. Reproduced from Nishimura *et al.* (2000).

Meristem behaviour, *STM*, and the evolution of different forms of *Streptocarpus*

Changes in meristem behaviour in Streptocarpus

The three main morphological transitions in *Streptocarpus* are (1) the switch from caulescent to acaulescent form, (2) the switch from unifoliate to rosulate form, and (3) the switch between rosulate to unifoliate form. It is probable that the switch between caulescent and acaulescent form only occurred once, and that the switches between unifoliate and rosulate form have occurred several times during *Streptocarpus* evolution (Möller and Cronk 2001a). It is not clear which of the acaulescent forms first arose. The switch from caulescent form to acaulescent form appears to have involved a loss in capacity of the apical meristem to form a shoot and leaves, and this may result from incorrect initiation of the SAM during embryogenesis. The switch from unifoliate to rosulate form appears to involve a gain in meristem capacity; instead of iterating inflorescences and then dying, the meristem iterates successive phyllomorphs. Conversely, the switch from rosulate to unifoliate implicates a loss-of-capacity. This variation in form may be effected by changes in meristem behaviour (Jong and Burtt 1975), and therefore may implicate meristem determining genes. Of the genes affecting meristem behaviour mentioned so far, the Class 1 *knox* genes appear the most useful tool to investigate *Streptocarpus* (Cronk and Möller 1997).

Parallels between knox *mutants and* Streptocarpus *morphology*

Several features of *STM,* and Class 1 *knox* loss- and gain-of-function mutants are highly reminiscent of *Streptocarpus*. These are (1) strong *stm* mutants lack an embryonic SAM (Barton and Poethig 1993), as do *Streptocarpus*; (2) *stm* mutants iterate leaf-like organs from a region of fused cotyledonary petiolar tissue, and then from subsequent leaf axils or petioles (Barton and Poethig 1993, Clark *et al.* 1997), in a manner analogous to rosulate *Streptocarpus*; (3) gain-of-function *knox* mutants show meristematic activity in leaves, and disrupted leaf morphology (Lincoln *et al.* 1994, Chuck *et al.* 1996, Hareven *et al.* 1996, Byrne *et al.* 2000), properties shared by phyllomorphic *Streptocarpus*.
A further line of evidence supporting the case for involvement of *STM* in morphological change in *Streptocarpus* is that a tobacco *STM* homologue has been implicated in repressing the GA biosynthetic pathway (Sakamoto *et al.* 2001), and unifoliate *Streptocarpus* phenocopy caulescent *Streptocarpus* when they are treated with GA (Rosenblum and Basile, 1984).

Aim of the PhD

The primary aim of my PhD was therefore to test whether *Streptocarpus* homologues of *STM*, and other meristem determining genes, could be implicated in morphological innovation and evolution in *Streptocarpus* by altering meristem behaviour through changes in their expression.

Chapter 2: Materials and methods

Plant material

All plant material was supplied from the RBGE *Streptocarpus* collection. A full list of species used, their RBGE accession numbers and authorities is given in table 1. Crosses between unifoliate and rosulate species were used to investigate the inheritance of growth form. Parent, hybrid, and backcross accession numbers are given in tables 2 and 3.

Species	Authority	Accession number	Growth form
Streptocarpus beampingaratrensis	Humbert	19972887	caulescent
Streptocarpus burundianus	Hilliard and Burtt	herbarium S137a	rosulate
Streptocarpus dunnii	Hooker	19972909	unifoliate
Streptocarpus hirticapsa	Burtt	19932793	rosulate
Streptocarpus ibityensis	Humbert	19932867	rosulate
Streptocarpus itremensis	Burtt	19972889	unifoliate
Streptocarpus kentaniensis	Britten and Story	19961876	rosulate
Streptocarpus levis	Burtt	19982242	caulescent
Streptocarpus modestus	Britten	19943058	rosulate
Streptocarpus pallidiflorus	Clarke	19691211	caulescent
Streptocarpus papangae	Humbert	19972886	caulescent
Streptocarpus pallidiflorus	Clarke	19691211	caulescent
Streptocarpus pentherianus	Fritsch	19972034	plurifoliate
Streptocarpus primulifolius	Gandoger	19912192	rosulate
Streptocarpus rexii	(Hook.) Lindl.	19870333	rosulate
Streptocarpus saxorum	Engler	19711885	caulescent
Streptocarpus schliebenii	Mansfeld	herbarium S136a	caulescent
Streptocarpus spp.	-	19972893	caulescent
Streptocarpus stomandrus	Burtt	19711392	caulescent
Streptocarpus thompsonii	Brown	19982253	caulescent
Streptocarpus thompsonii	Brown	19941334	caulescent
Streptocarpus thysanotus	Hilliard and Burtt	no. 18 S129a	caulescent
Streptocarpus venosus	Burtt	19982247	caulescent
Streptocarpus wendlandii	Sprenger	19982266	unifoliate
Streptocarpus wittei	Hilliard and Burtt	19981673	unifoliate
Saintpaulia ionantha	Wendl.	19970092	caulescent
Saintpaulia tongwensis	Burtt	19970096	caulescent
Saintpaulia velutina	Burtt	19872179	caulescent

Table 1: *Streptocarpus* species used in this study, their RBGE accession numbers, authorities and growth forms.

Generation	Taxon	Accession number
Parent	*S. dunnii*	19941745A
Parent	*S. rexii*	19870333/3
Parent	*S. wittei*	19871695
F1	*S. rexii* x *S. dunnii*	19972904
F1	*S. wittei* x *S. rexii*	19972906

Table 2: Plants used to generate backcross populations, and their RBGE accession numbers.

Family number	Cross	Accession number
Family 1	(*S. rexii* x *S. dunnii*) x *S. dunnii*	19980107
Family 2	(*S. rexii* x *S. wittei*) x *S. wittei*	19982603
Family 3	(*S. rexii* x *S. wittei*) x *S. wittei*	19982606
Family 4	(*S. rexii* x *S. wittei*) x *S. wittei*	19982607
Family 5	(*S. rexii* x *S. wittei*) x *S. wittei*	19982610
Family 6	(*S. rexii* x *S. wittei*) x *S. wittei*	19982614
Family 7	(*S. rexii* x *S. wittei*) x *S. wittei*	19982619
Family 8	(*S. rexii* x *S. wittei*) x *S. wittei*	19982620
Family 9	(*S. rexii* x *S. wittei*) x *S. wittei*	19982621
Family 10	(*S. rexii* x *S. wittei*) x *S. wittei*	19982622
Family 11	(*S. rexii* x *S. wittei*) x *S. wittei*	19982623
Family 12	(*S. rexii* x *S. wittei*) x *S. wittei*	19982624

Table 3: Details of backcrosses in *Streptocarpus*.

Molecular techniques

Where percentage measures are used in techniques described, the measure is w/v for a solid or v/v for a liquid.

Nucleic acid extraction

RNA extraction

RNA for RACE experiments was extracted according to Martin and Northcote (1981). To prevent degradation of RNA by atmospheric RNases, gloves were worn, and for each experiment a new bag of Eppendorf tubes and pipette tips was opened and used. All

solutions were made up with DEPC treated sterile distilled water. 0.5-1 g tissue was ground in 800 µl buffer (50 mM Tris.HCl pH 9, 150 mM NaCl, 5 mM EDTA) and 200 µl 10% SDS. This suspension was transferred to a 1.5 ml Eppendorf tube and 0.5 ml phenol/chloroform pH8 was added and mixed by vortexing. Samples were centrifuged at 15,000 rpm for 2 minutes in a benchtop centrifuge. 800 µl supernatant was recovered using a pipette, transferred to a new tube, and treated again with phenol/chloroform. Samples were centrifuged again as before, and 800 µl of the supernatant was transferred into a new tube. 200 µl 12 M LiCl was added to the supernatant to precipitate the RNA, and mixed in by shaking. RNA was allowed to precipitate overnight at 5°C. To pellet the RNA, samples were centrifuged at 15,000 rpm, and the supernatant was removed. RNA was dissolved in 200 µl DEPC treated water at 80°C, with intermittent vortexing to loosen the pellet. After 2-3 minutes, samples in tubes were transferred to chill on ice, and were then centrifuged as before to remove debris from the supernatant. The supernatant was transferred to a new tube, and RNA was precipitated using 1/10 volume 3 M Sodium acetate and 2.5 volumes 70% ethanol. Samples were centrifuged as before to pellet the RNA, and the salt was washed out of the pellet using 70% ethanol. After the first wash, samples were centrifuged, washed and spun again, and allowed to air dry at room temperature. RNA was resuspended in 100 µl water, and samples were frozen rapidly in liquid nitrogen and stored at -80°C. For RT-PCR experiments, RNA was extracted using TRIzol reagent (Gibco BRL.) according to manufacturer's instructions. In both cases 13 µl RNA was used in cDNA synthesis reactions primed using a primer consisting of oligod(T) and a unique anchor sequence, Q_t, according to Frohman (1995) for 3' RACE, and according to manufacturer's instructions for 5' RACE .

Small scale DNA extraction for PCRs

Two discs punched from leaves using a 1.5 ml Eppendorf tube yielded sufficient DNA for use in PCRs, and DNA was extracted as in Doyle and Doyle (1987). Discs were ground in 750 µl CTAB extraction buffer (100 mM tris-HCl pH8.0, 1.4 M NaCl, 20 mM EDTA, 2% CTAB) with 0.4% β-mercaptoethanol pre-heated to 65°C. The suspension was transferred to an Eppendorf tube, and left to incubate at 65°C in a heating block for 15-20 minutes. 750 µl chloroform/isoamyl alcohol (24:1) was added and mixed in by shaking. Samples were then centrifuged at 10,000 rpm at room temperature for 5 minutes, and the aqueous phase was transferred to a new tube. An equal volume of

isopropanol was added to the supernatant and mixed in by gentle shaking. DNA was allowed to precipitate at -20°C for 30 minutes, and was pelleted by centrifugation at 10,000 rpm for 5 minutes. DNA pellets were washed twice and dried as described for RNA extraction, and DNA was dissolved in TE or water and stored at -20°C or 5°C.

Large scale DNA extractions

Conventional *Antirrhinum*, *Arabidopsis* and maize DNA extraction protocols did not yield sufficient quantities or quality of DNA from *Streptocarpus* for restriction analysis. A combination of several protocols was devised. 4 g of young leaf tissue was frozen and ground to a powder in liquid nitrogen in a mortar. The leaf powder was mixed with 6 ml of a urea based extraction buffer (7 M urea, 5 M NaCl, 2 M tris-HCl pH8.0, 0.5 M EDTA pH 8.0) in a 30 ml centrifuge tube. This suspension was shaken vigorously with an equal volume of tris-HCl pH8.0 equilibrated phenol and centrifuged at 10,000 rpm in a Sorvall RC-5B Automatic Superspeed Refrigerated Centrifuge for 10 minutes to pellet protein and cell debris. The supernatant was treated in the same way with phenol/chloroform pH 8, and then chloroform. DNA was precipitated using an equal volume of isopropanol, and pelleted by centrifugation at 5,000 rpm for 5 minutes as above. The pellet was washed with 1 ml 70% ethanol and transferred to a 1.5 ml Eppendorf tube. The DNA was dissolved in 400 µl water overnight. To remove carbohydrate, 200 µl 7M sodium acetate was added to the tube, mixed by shaking, and incubated overnight at 4°C. Samples were centrifuged at 25,000 rpm for 20 minutes in Eppendorf tubes in an International Equipment Company Micromax bench top centrifuge. The supernatant was carefully pipetted into a new tube. 1 ml ethanol was added to precipitate DNA, and DNA was pelleted as above. The pellet was washed twice in 70% ethanol and dissolved in water or TE to the desired volume.

PCR

Principle of PCR

The PCR (Mullis and Faloona 1987) uses short known sequences of DNA (primers) flanking a target region, to prime DNA synthesis across that region in both directions. The target strand is heat denatured, then the temperature is dropped so that primers anneal to their target sequence. Addition of further bases from dNTPs included in the reaction mix to the primer in the 3' direction is catalysed by a heat stable polymerase, also included in the mix. This cycle is repeated many times, amplifying logarithmically the target DNA sequence.

Primer design for PCR

All degenerate primers were designed by aligning sequences available on GenBank and from Andrew Hudson's laboratory, from as many species as were available for that gene. Regions of consensus were used to locate primer sites, and to minimise the need degeneracy. Primers designed by JH were analysed using the NetPrimer program, available on the internet (www.premierbiosoft.com). A list of all the primers used in this study, their sequences and who designed them is given in table 4. The position of each of the primers used with respect to *STM* protein sequence is shown in figures in the main text.

Primer	Source	Sequence (5'-3')
AAP	Gibco BRL	GGCCACGCGTCGACTAGTACGGGIIGGGIIGG
AP1	Siebert *et al.* (1995)	GGATCCTAATACGACTCACTATAGGG
AP2	Siebert *et al.* (1995)	AATAGGGCTCCGAGCGGC
Clavata forward	R. W.	GGIAACGGIGGIGCIGGIATHGTITA
Clavata reverse	R. W.	TTIGCIARICCRAARTCIGCIAC
CUC1	J. H.	ATGGGIGARAARGARTGGTA
CUC2	J. H.	TGGGTIATGCAYGARTA
cuc p1	J. H.	CATCCATATCGAATGGCCCAGCAT
cuc p2	J. H.	GCGTTTTGGAGTCGCTGGAA
CUCwalk1	J. H.	GCCCAGCTAACGTGGCCCAGTCGAAATT
CUCwalk2	J. H.	GGTCAGCGATCCGTTCCAGCGACTCCAAAA
ELK UP	J. H.	GCTTGGGGRTCAATGAAACTGT
FC	M. M. & I. O.	TGGAGCCGCCACTACAAATG
FE	M. M. & I. O.	AAATGGCCTTATCCTTCGGTA
First exon up	J. H.	GCAGAGTGAGACAAAGTAAGGGCT
GR22	M. M. & I. O.	TGAACCAGTTGTTGATYTGCTT
GSP1	J. H.	CCACGAACTGCATATC
GSP2	J. H.	GCGCTTCCTTTGGTTAATGAAC
GSP3	J. H.	CCAGTTGTTTATCTGCTTCTGTT
GSP4	J. H.	CGAGGGCGAAATATC
GSP5	J. H.	GAATGGAGTGCCCATGAAATTG
GSP6	J. H.	GCTTCCAATGGCGCTTCCAAA
KNAT1	J. H.	ATGCARTTYATGGTIATGGAYGG
MEGSD	J. H.	CACGCAGTAGTGTGTGTAATGGAG
M13-40	Promega	TTTTCCCAGTCACGACGTT
New rex down	J. H.	GCTCGGGGTGACATTAGTAT
New rex up	J. H.	GCATCAAATATAGAATAAGAA
New sax down	J. H.	CTCGTTGTAAGTTGCAGCTTGT
New sax up	J. H.	GTCAGCCGACTCCAAGCCAA
PHAN	J. H.	GGTGGGARGTITTYAARGA
Probe1	J. H.	GGAAGCCGTCCGAGGATAT

Probe2	J. H.	GCGTTCAATATGTTGTACTA
5Pseq	J. H.	GGCATCTCCGTCTCCTTCCAT
Qi	Frohman (1995)	GAGGACTCGAGCTCAAGC
Qo	Frohman (1995)	CCAGTGAGCAGAGTGACG
Qt	Frohman (1995)	CCAGTGAGCAGAGTGACGAGGACTCGAGCT
Rex down	J. H.	TTTCTTGTTTAGTTCCTGTT
ribosomal down	J. H.	CGACACACCCCTGGTACTTT
ribosomal up	J. H.	CCTGCAACTTGGTGGTACGGTA
Sax down	J. H.	CGGAAGAAACAATTGAGATGG
Sax 2 down	J. H.	CAATTTCTTGTTTAGTTCCTTGTAATGC
Sax far down	J. H.	CAGCTCGGAAAACATATA
Sax further	J. H.	CTGGGTCGTTGGATTCATGGT
Sax further up	J. H.	ATCTACTACTTCTTGA
SFD	J. H.	CGTAGTAGTGTATGGATATGATGGTGTAGAC
SFU	J. H.	CGGGTGCCAGAAGATGAAGTAG
SH3D	J. H.	TCTCCGGTCTTAATTTCTGGTGTT
SH3U	J. H.	CCCGCGCCACGGATATCTAA
Sax up	J. H.	CTGGGATTCCTACATTAGT
Screen up	J. H.	GCTTCTGAGATTCCTAGATT
SP6	Promega	CATACGATTTAGGTGACACTATAG
T7	Promega	TAATACGACTCACTATAGGGAGA
UAP	Gibco BRL	CUACUACUACUACUAGGCCACGCGTCGACT
Upstream P1	J. H.	CGCCACAAGCTGCGACATAAA
WALK 1	J. H.	CGAGCCTGTCTACACCATCATATCCATAC
WALK 2	J. H.	CCAACCATGAATCCAACGACCCAGAT

Table 4: Primers referred to in this text, and their source. J. H.: Jill Harrison, M. M.: Michael Möller (RBGE), R. W: Richard Waites (University of York), I. O.: Ian Oliver (University of Edinburgh).

PCR conditions

Several PCR based techniques were used to amplify genes involved in meristem function.

Reaction and cycling conditions depended on the template amplified, and were varied according to need. Typical conditions for the main types of PCR used are given below. All PCRs were carried out in 0.2 ml thin walled tubes using a Hybaid Omn-E thermal cycler or an MJ Research PTC-200 DNA Engine.

Standard PCR

For amplification of fragments less than 1 kbp, standard PCR conditions were used. These were optimised for use with Bioline BIOTAQ polymerase in a 50 µl reaction.

Reaction components:		Cycling conditions:	
DNA (5 ng/µl)	2 µl	(94°C	30 s

10 x reaction buffer	5 µl	55°C	30 s
50 mM $MgCl_2$	2 µl	72°C	1 min)
10 mM dNTPs	1 µl	30 cycles	
10 mM forward primer	1 µl		
10 mM reverse primer	1 µl		
sterile deionised water	37.5 µl		
BIOTAQ polymerase	0.5 µl		

Degenerate primer PCR

Degenerate primers to homeodomain sequences, Fc (sense) and GR22 (antisense), which flank intron 3 of *STM* were used to amplify sequences from several *Streptocarpus* species in a 50 µl PCR.

Reaction components:		Cycling conditions:	
DNA (5 ng/µl)	2 µl	94°C	2 min
10 x reaction buffer	5 µl	(94°C	15 s
50 mM $MgCl_2$	2 µl	50°C	15 s
10 mM dNTPs	1 µl	72°C	30 s)
10 mM forward primer	5.1 µl	35 cycles	
10 mM reverse primer	5.1 µl	72°C	5 min
sterile deionised water	29.3 µl		
BIOTAQ polymerase	0.5 µl		

3' Rapid Amplification of cDNA Ends (3' RACE)

First round reaction components:		First round cycling conditions:	
cDNA	2 µl	94°C	1 min
10 x reaction buffer	5 µl	(94°C	15 s
50 mM $MgCl_2$	2 µl	50°C	30 s
10 mM dNTPs	1 µl	72°C	2 min)
10 mM forward primer	5.1 µl	35 cycles	
10 mM reverse primer	1 µl	72°C	5 min
sterile deionised water	33.4 µl		
BIOTAQ polymerase	0.5 µl		
Second round reaction components:		Second round cycling conditions:	
1:500 first round product	2 µl	94°C	1 min
10 x reaction buffer	5 µl	(94°C	15 s

50 mM $MgCl_2$	2 µl	58°C	30 s
10 mM dNTPs	1 µl	72°C	1 min 30 s)
10 mM sense primer	5.1 µl	35 cycles	
10 mM reverse primer	1 µl	72°C	5 min
sterile water	33.4 µl		
BIOTAQ polymerase	0.5 µl		

5' Rapid Amplification of cDNA Ends (5' RACE)

5' RACE experiments were done using a Gibco kit, the 5' RACE System for Rapid Amplification of cDNA Ends (Version 2.0). Gene specific primers were designed against sequence obtained in 3' RACE (GSP1, 2 and 3 complement *S. saxorum*; GSP4, 5, and 6 complement *S. dunnii* and *S. rexii*), and their sequences are given in table 4.

Inverse (I) PCR

IPCR (Ochman *et al.* 1988, Triglia *et al.* 1988) was used to lengthen genomic *STM* sequences, with primers designed against cDNA sequences. Genomic DNA was digested with *Hin*D III, *Eco*R I.

Reaction components:		Reaction Conditions:
DNA (5 µg)	10 µl	overnight at required
10 x buffer	2 µl	temperature
5 mM spermidine	0.5 µl	
100 x BSA	0.2 µl	
sterile distilled water	6.8 µl	
Enzyme	0.5 µl	

Reactions were heat inactivated at 70°C for 10 min. DNA fragments were circularised using DNA ligase (NEB) as below.

Reaction components:		Reaction conditions:
DNA (0.5-1 µg)	50 µl	overnight at room
10 x NEB ligase buffer	40 µl	temperature
sterile deionised water	308.5 µl	
NEB T4 DNA ligase	1.5 µl	

1 ml of Qiagen buffer QG was added, followed by 200 µl isopropanol, and the DNA was purified on a Qiagen gel purification column. DNA was eluted in 100 µl water, and quantified by gel electrophoresis. 50 ng was amplified in a 50 µl PCR using conditions overleaf.

Reaction components:		Cycling conditions:	
DNA	50 ng	94°C	2 min
10 x Expand reaction buffer	5 µl	(94°C	30 s
10 mM dNTPs	1 µl	55°C	30 s
10 mM sense primer	1 µl	68°C	3 min)
10 mM antisense primer	1 µl	10 cycles	
sterile deionised water	q. v.	(94°C	30 s
Roche Expand Taq polymerase	0.5 µl	55°C	30s
		68°C	4 min
		+5 s extension time per cycle)	
		30 cycles	
		72°C	10 min

Genome walking

Known *STM* sequences were also extended in the 5' direction using genome walking. The protocol used was adapted from a Clontech protocol (PT3042-1 version PR03300) and the original protocol of Siebert *et al.* (1995) by G. Ingram and K. Coenen (University of Edinburgh). 2.5 µg Genomic DNA was digested with *Eco*R V or *Pvu* II in a 100 µl reaction as below.

Reaction Components:		Reaction conditions:
DNA (2.5 µg)	10 µl	overnight at 37°C
10 x buffer	2 µl	
5 mM spermidine	0.5 µl	
100 x BSA	0.2 µl	
sterile distilled water	6.8 µl	
Enzyme	0.5 µl	

The reaction products were extracted with phenol/chloroform, and precipitated with 0.1 volume Sodium acetate and 2.5 volumes of ethanol followed by centrifugation at 12,000 rpm in a bench top centrifuge. The pellet was washed in 70% ethanol, air dried, and resuspended in 20 µl water. Digested DNA was ligated to an adaptor (Siebert *et al.* 1995) in the reaction below.

Reaction components:		Reaction conditions:
DNA	5 µl	Overnight at room
10 x NEB ligase buffer	1 µl	temperature

NEB T4 DNA ligase	1 µl
25 mM adaptor	2.4 µl
sterile deionised water	0.6 µl

The ligase was denatured by incubation at 70°C for 5 minutes, and the reaction volume was made up to 100 µl with TE for use in PCR. 1 µl of the adaptor-ligated DNA pool was used in nested PCRs.

First round reaction components:		Cycling conditions:	
DNA	1 µl	94°C	1 min
10 x Expand reaction buffer	5 µl	(94°C	25 s
10 mM dNTPs	1 µl	72°C	3 min)
10 mM adaptor primer 1	1 µl	5 cycles	
10 mM gene specific primer 1	1 µl	(94°C	25 s
sterile deionised water	40.8 µl	67°C	3 min)
Roche Expand Taq polymerase	0.2 µl	20 cycles	
		67°C	7 min
Second round reaction components:		Cycling conditions:	
1/50 diluted round 1 product	5 µl	as above	
10 x Expand reaction buffer	5 µl		
10 mM dNTPs	1 µl		
10 mM adaptor primer 2	1 µl		
10 mM gene specific primer 2	1 µl		
sterile deionised water	36.8 µl		
Roche Expand Taq polymerase	0.2 µl		

Cleaved Amplified fragment Polymorphisms (CAPS) based screen for segregation of Sknox1 alleles

PCR products were amplified from *S. dunnii*, *S. rexii* and *S. wittei* using primers *Screen up* and *new rex down* under standard cycling conditions. PCR fragments were precipitated using Sodium acetate and ethanol, washed in 70% ethanol, resuspended in water and normalised for concentration on a miniscus gel. 200 ng DNA was cleaved in a 20 µl overnight restriction digest using *Sau*3 AI at 37°C as overfleaf.

DNA	200 ng
buffer	2 µl
*Sau*3 AI	0.5 µl
sterile deionised water	q. v.

Cleaved products were separated by agarose gel electrophoresis using 2% gels, and gels were visualised as described below.

Gel electrophoresis of DNA

Except for use in Southern blotting (described below), all DNA fragments were separated by electrophoresis in agarose miniscus gels in 0.5 x TBE (Sharp *et al.* 1973). DNA solutions were mixed with one tenth volume loading dye (0.4% bromophenol blue, 0.4% xylene cyanol, 25% ficoll in water (Maniatis *et al.* 1982) . For PCRs with expected products of 500 bp or less, 2% gels were used. For larger products 1% gels were used. Gels incorporated 0.5 µg/ml ethidium bromide. DNA size markers (1 kbp ladder, and low mass ladder) were from Boehringer Mannheim and Gibco BRL respectively. Gels were run at 100 V until the dye front had migrated a sufficient distance towards the anode for fragment resolution. Gel pictures were taken and stored digitally using the Enhanced Analysis System (E.A.S.Y., version 4.19, Scotlab, Coatbridge, Lanarkshire, UK).

Gel purification of PCR products

All PCR products to be sequenced or cloned were gel purified using the Qiagen gel extraction kit according to manufacturer's instructions.

Cloning

All DNA fragments to be cloned were ligated into the pGEM-T or pGEM-T EASY vector (Promega) overnight at 4°C. When the plasmid ligates to the desired insert, the insert disrupts β-galactosidase activity that is dependent on the plasmid's lacZ reporter gene. β-galactosidase catalyses cleavage of galactose from X-gal contained in agar, causing colonies to stain blue. When the plasmid ligates to itself, the reporter gene is functional, and colonies remain white. *E. coli* strain TOP10 [Invitrogen] or JM101 were used. For TOP10 cells transfection was by heat shock according to manufacturer's instructions. For JM101 electrocompetent cells, cells were pulsed at 200 Ω with 1.8 kV in a sterile 1 mm cuvette, and immediately resuspended in 1 ml SOC medium. Transformed cells were grown for one hour at 37°C in SOC medium in a shaking

EX BIBL. UNIV.

incubator before being spread onto LB agar containing 100 mg/ml ampicillin, 80 µg/ml X-gal and 0.5 mM IPTG to select and screen for bacteria containing plasmid with insert. White colonies were picked and cultured in 5 ml LB broth with ampicillin overnight at 37°C in a shaking incubator. DNA was extracted from 3 ml of bacterial culture using the Qiagen miniprep kit to manufacturer's instructions. Plasmid DNA was screened for length and direction of insert by digestion with *Eco*R I in a 20 µl reaction as below or by standard PCR.

Reaction components:		Reaction conditions:	
DNA	2 µl	37°C	1 h
buffer	2 µl		
*Eco*R I	0.2 µl		
sterile deionised water	16.8 µl		

Sequencing

All cycle sequencing (Sanger and Coulson 1975) by JH and Nicola Wright was done using the ABI PRISM d-rhodamine terminator cycle sequencing kit with AmpliTaq DNA polymerase. Reaction conditions were adapted from manufacturer's instructions.

Reaction components:		Cycling conditions:	
Terminator Ready Reaction Mix	3 µl	96°C	30 s
template if PCR product	40-100 ng	(96°C 30s	
template if plasmid DNA	200 ng	50°C	15 s
10 mM primer	1 µl	60°C	4 min)
sterile deionised water	q.v.	25 cycles	
total volume	10 µl		

Products were precipitated as recommended by the manufacturer. Products were electrophoresed by N. Preston, L. Evenden, H. Williamson (ICMB) or J. Lovell (ICAPB).

Sequence analysis, alignment and phylogenetic analysis

Electropherograms were viewed and edited using Factura (Applied Biosystems) or Chromas version 1.62 (Technelysium). Sequences for the intron phylogeny were further edited and aligned using ClustalX, CLUSTALW (Thompson *et al.* 1994, Thompson *et al.* 1997), or manually in Sequence Navigator version 1.01 (Perkin Elmer). Aligned sequences were analysed using parsimony in PAUP version 3.1.1 or 4.0b8 (Swofford

1993, 1998). Simple addition and tree bisection-recombination were used for all analyses. Sequence homology was determined by BLAST searching all available databases (http://blast.genome.ad.jp/; option TBLASTN).

Protein sequences for phylogenetic comparison to *Streptocarpus* were taken from GenBank (accession numbers given in table below), and aligned in ClustalX or CLUSTALW using default parameters. Phylogenetic analysis was done in PAUP using default parameters.

Southern blotting

For Southern transfer and hybridisation (Southern 1975) with genomic DNA, 5 µg DNA was digested with different restriction enzymes, identified where relevant in the main text. DNA fragments were precipitated with 0.1 volume 3 M Sodium acetate and 2.5 volumes ethanol, and dissolved in 20 µl water. Fragments were separated by electrophoresis in 0.8% agarose gels in 0.5 x TBE overnight at 25 V. Gels were stained with 10 µg/ml ethidium bromide in 0.5 x TBE and photographed. For Southern transfer and hybridisation of PCR products, products were separated as usual by gel electrophoresis. Gels were treated with 0.2 M HCl for 10 min to nick the DNA. Gels were transferred to 0.5 M NaOH, an alkaline denaturing solution for 30 min. Gels were buffered in 1 M Tris.HCl pH7 + 1.5 M NaCl for 30 min. Gels were inverted onto a supported porous wick soaked in 20 x SSC in a baking dish and covered by Hybond N nylon membrane (Amersham International plc) trimmed to size and wetted in 20 x SSC. DNA was transferred to the membrane by flow of 20 x SSC through the wick, gel and membrane to weighted dry paper towels placed on top of the filter overnight as described in Maniatis *et al.* (1982). Filters were rinsed in 0.2 x SSC, and DNA was fixed to the membrane by UV cross linking with 40 J/cm^2 UV light.

Probe radiolabelling

Blots were probed with DNA incorporating radioactive dCTα^{32}P, or with a Dig labelled riboprobe. For radioactive probes DNA was labelled in a random priming reaction (Feinberg and Vogelstein 1983) as below.

Reaction components:

DNA (100 ng)	18 µl
5 x oligonucleotide labelling buffer (OLB)	6 µl

(OLB: 250 mM Tris-HCl pH 8.0, 25 mM $MgCl_2$, 5 mM β -mercaptoethanol, 2 mM dATP, 2 mM dGTP, 2 mM dTTP, 1 M HEPES (N-[2-hydroxethyl]piperazine-N'-[2-ethanesulfonic acid]), adjusted to pH 6.6 with 4 M NaOH.

Boehringer Klenow polymerase	2 µl
BSA (10 mg/ml)	1 µl
dCTα^{32}P (Amersham International plc)	3 µl (30 µCi)

The reaction was incubated at 37°C for 45 min. Unincorporated nucleotides were separated from labelled probe by passing the labelling reaction through a Sephadex G-50 column as described in Maniatis *et al.* (1982). Immediately before use, the probe was heat denatured by boiling for 5 min in an Eppendorf tube in a water bath.

Riboprobe synthesis

PCR products were amplified from fragments cloned into pGEM-T or pGEM-T EASY vectors, using T7-40 and a probe specific primer, so that products incorporated the T7 promoter. Riboprobes incorporating Dig-labelled UTP were transcribed in vitro from these products.

Reaction components:		Reaction conditions:	
300 ng template	11 µl	37°C	1 h
5 mM ATP	2.5 µl		
5 mM CTP	2.5 µl		
5 mM GTP	2.5 µl		
1 mM Dig-UTP	2.5 µl		
15 units RNAguard RNase inhibitors (Amersham Pharmacia Biotech inc.)	0.5 µl		
RNA polymerase buffer	2.5 µl		
Promega T7 RNA polymerase	1 µl		

Reactions were diluted to 100 µl and then treated with DNase to digest the DNA template at 37°C for an hour.

Reaction components:	
tRNA (100 mg/ml)	1 µl
Boehringer Mannheim DNase	0.5 µl
DEPC treated sterile deionised water	75 µl

RNA was precipitated for an hour at -70°C using an equal volume of 3.5 M Ammonium acetate and two volumes of ethanol. Salt was washed out of the pellet twice using 70% ethanol, and the probe was resuspended in 40 µl 50% formamide to prevent RNA degradation.

Hybridisation and autoradiography

To prevent non-specific binding of probe, blots were rocked in pre-hybridisation solution. Conditions differed depending on the probe used. For radio-labelled probes this was 4 x SSC, 1% SDS, and 0.5% milk in water, and filters were incubated in 50 ml in a sealed sandwich box at 65°C for 15 min in a Hybaid hybridisation oven. For Dig-labelled riboprobes pre-hybridisation solution was 4 x SSC, 0.1% N-lauryl sarcosine, 0.02% SDS, and 1% milk powder in water, and filters were incubated for 2h at 65°C. Denatured probe was added to the mix. Blots were probed overnight at 65°C. Hybridisation mix was tipped off and stored at -20°C. For radio-labelled probes, excess mix was washed off in 250 ml 2x SSC with 0.1% SDS for 15 min at 65°C and the filter was then drained of excess fluid. The filter was placed face down on Saran Wrap, enfolded, and placed in a cassette face up against X-ray film (Sterling Diagnostic Imaging inc). Film was exposed overnight at -70°C and developed in a automatic X-Omat processor. For Dig-labelled probes, hybridisation fluid was again tipped off and stored at -20°C for further use. Filters were washed twice at room temperature for 20 minutes in 2x SSC with 0.1% SDS, and then twice at 68°C in 0.5% SSC with 0.1% SDS. Filters were then soaked in 50 ml of maleic acid blocking buffer (100 mM maleic acid, 150 mM NaCl; pH 7.5, 1% Boehringer blocking reagent) to block against non-specific antibody binding for 45 minutes at room temperature. Filters were then incubated in 20 ml maleic acid blocking buffer with 2 µl anti-Dig antibody for 30 minutes. To remove excess antibody, filters were washed twice for 15 minutes in maleic acid buffer (100 mM maleic acid, 150 mM NaCl; pH 7.5) with 0.3% TWEEN-20. Filters were transferred to detection buffer (100 mM Tris-HCl pH 9.5, 100 mM NaCl) and allowed to equilibrate to pH 9.5 for 5 minutes at room temperature. Antibody detection was using a chemiluminescent substrate (CSPD). Filters were sealed in a bag with 0.5 ml CSPD, for 3 minutes at room temperature. Excess fluid was drained off, bags were resealed, and filters were incubated at 37°C for 15 minutes. Filters remained in the bag and were used

to expose X-ray film (as above) for 2-10 minutes, or until a signal was perceived when the film was developed as above.

Microscopy and histological techniques

Identification of DNA synthesis with BrdU

Seeds were germinated and grown on damp filter paper in Petri dishes for 10 days at 27°C in constant white light. One day prior to fixing, dishes were infiltrated with 1 ml 1 mM 5-bromodeoxyuridine (BrdU)/ 1 µM 5-fluoro-2'-deoxyuridine (FUdR) in water and incubated for a further 24h. Seedlings were fixed in 4% paraformaldehyde and dehydrated through an ethanol series to 100% ethanol and embedded in Technovit 7100 resin (Kulzer and Co.) as described by Tsukaya *et al.* (1993) and Fujie *et al.* (1994). 1 µm sections were cut using a glass knife in an ultramicrotome, and mounted on coverslips. Mounted sections were stained with a BrdU specific mouse monoclonal antibody, which was detected with Alexa 488 (Molecular Probes Inc.) as described by Tsukaya (1997). Sections were then double stained with 5 µl 10 mM 4'-6'-diamidino-2-phenylidole (DAPI), which localises to DNA. Slides were examined using a Leica DAS Mikroskop DMR microscope, and photographs were taken using Konica 400 ASA colour film.

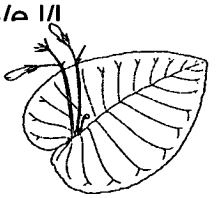
Chapter 3: Inheritance of morphology in *Streptocarpus*

Crosses performed

Previous studies on the inheritance of growth form in *Streptocarpus* suggested that two loci determined the inheritance of morphology between rosulate and unifoliate species, one acting early and one late, and that the rosulate character was dominant with respect to unifoliate (Oehlkers 1938, Oehlkers 1941, Oehlkers 1964).

If Oehlkers' predictions are correct, the rosulate parent in his crosses should have the genotype E/E L/L, and the unifoliate parent, should have the genotype e/e l/l, where E and L are the dominant alleles at the early and late acting loci respectively. The F1 hybrid should have the genotype E/e L/l, and develop rosulate morphology early. F2 populations should contain 9 different genotypes as outlined in figure 11 below. In this case 3/4 of plants (those carrying the E allele) should show rosulate morphology early in development. A further 3/16 (those carrying the L allele) of plants should develop it later. The initial ratio of rosulate: unifoliate plants should shift from 3:1 to 15:1.

rosulate parent: E/E L/L x unifoliate parent: e/e l/l



F1 hybrid E/e L/l

gametes: E/L, E/l, e/L, e/l

F2 genotypes:

	E/L	E/l	e/L	e/l
E/L	E/E L/L	E/E L/l	E/e L/L	E/e L/l
E/l	E/E L/l	E/E l/l	E/E L/l	E/e l/l
e/L	E/e L/L	E/e L/l	e/e L/L	e/e L/l
e/l	E/e L/l	E/e l/l	e/e L/l	e/e l/l

Figure 11: Predicted segregation of alleles and morphology in an F2 population where one early and on late acting dominant locus determines form. Early rosulates carry the E allele, late rosulates carry the L allele.

To verify Oehlkers' observations Q. Cronk and M. Möller crossed unifoliate and rosulate species available at the RBGE to form an F1 hybrid (figure 12). As predicted, hybrids always had more than one phyllomorph like the rosulate parent, although these were often larger than those found in the rosulate parent. Backcrosses to the unifoliate parents, proposed to carry recessive alleles at both loci, were made so that the full range of morphological types possible would be displayed, and in the backcrosses morphology segregated.

Figure 12: From left to right; *S. dunnii* (a unifoliate species), F1 hybrid, and *S. rexii* (a rosulate species).

Analysis of growth form in backcross plants

Accession numbers of the species used in the crosses, and of resulting hybrid plants are given in chapter 2. Sixteen *S. dunnii* x (*S. rexii* x *S. dunnii*) and 116 *S. wittei* x (*S. rexii* x *S. wittei*) backcross plants were grown in the greenhouse at the RBGE by S. Scott. To look for segregation of morphological types, growth form was monitored from seedling to flowering stages. A spectrum of types was found, and distinction between the unifoliate and rosulate characters was not always clear. For the *S. dunnii* x (*S. rexii* x *S. dunnii*) back cross (family 1) which matured first, I developed an intuitive numerical scoring system based on the degree of rosulate character in the plants. The scores used and the states they represent are given below. Plants were scored at flowering.

1: single leaf only.

1.5: two cotyledons (occasionally in unifoliate species one cotyledon does not out-compete the other, and both grow to flowering).

2: one main leaf, with some very small leaves.

3: two main leaves and some small.

4: more than two main leaves, but not fully rosulate.

5. fully rosulate.

Scores of 2 or over were classified as rosulate. This scoring system was tested independently by two people (J. Harrison and Q. Cronk) on a single population of plants (family 1), and gave consistent results between users shown in table 11 appendix 1.1, and summarised in table 6 below.

I felt that this scoring system was inflexible in that it did not allow me to compare objectively scores taken at different stages of development. I therefore recorded the actual number of leaves present, and their dimensions, at different stages in the life cycle for the remaining crosses (families 2-12, tables 5 and 6, and given in full in tables 11 and 12; appendices 1.2 and 1.3). The rosulate character was scored consistently as any plant growing more than one leaf at the time or scoring. To investigate the apparent stage dependence of the action of one of the genes affecting morphology seen by Oehlkers, plants were scored at different times. Initially plants were grown under uniform conditions in a growth room, and plants were scored when secondary phyllomorphs first appeared in the population. Plants with a single cotyledonary phyllomorph, or those with two initial cotyledons were scored as unifoliate, and plants with secondary phyllomorphs were scored as rosulate. Plants were scored again at flowering, also taking into account whether they had two initial cotyledons.

Family number	Unifoliate plants	Rosulate plants	Total
Family 1	ns	ns	-
Family 2	5	3	8
Family 3	7	3	10
Family 4	8	3	11
Family 5	5	5	10
Family 6	2	1	3
Family 7	5	3	8
Family 8	5	7	12
Family 9	6	7	13
Family 10	11	7	18
Family 11	8	4	12
Family 12	8	3	11
Total	70	46	116

Table 5: Form during juvenile development in backcross plants. Ns: plants were adult when I commenced my PhD, and were therefore not scored during juvenile development.

Family number	Rosulate plants	Unifoliate plants	Total
Family 1	12	4	16
Family 2	8	0	8
Family 3	10	0	10
Family 4	9	2	11
Family 5	9	1	10
Family 6	3	0	3
Family 7	5	3	8
Family 8	11	1	12
Family 9	11	2	13
Family 10	16	2	18
Family 11	12	0	12
Family 12	9	2	11
Total	115	17	132

Table 6: Form of backcross plants at flowering.

Results of the crosses

S. dunnii x (*S. rexii* x *S. dunnii*) backcross

If, as proposed by Oehlkers, rosulate morphology was dependent on dominant alleles at two loci, the rosulate *S. rexii* parent should have the genotype E/E, L/L, and the *S. dunnii* parent e/e, l/l. E and L are the dominant alleles at the early and late acting loci respectively. The F1 *S. rexii* x *S. dunnii* hybrid should therefore have the genotype E/e L/l, and develop rosulate morphology early. The backcross population should consist of four different genotypes: E/e L/l; e/e L/l; E/e l/l; e/e l/l; in equal proportions. In this case half the plants (those carrying the E allele) should show rosulate morphology early in development, and a further quarter should develop it later. The initial ratio of rosulate to unifoliate plants should shift from 1:1 to 3:1 as the plants matured.

In the *S. dunnii* x (*S. rexii* x *S. dunnii*) back cross, plants were adult when I commenced my PhD, and were therefore not scored during juvenile development. Morphology was scored at flowering, giving a 3:1 ratio of rosulate to unifoliate form, consistent with Oehlkers' proposal.

S. wittei x (*S. rexii* x *S. wittei*) backcross

In the *S. wittei* x (*S. rexii* x *S. wittei*) back cross, form was scored at two stages in development. During juvenile development, 70 plants were scored as rosulate and 46 plants as unifoliate. If a single dominant allele determines early morphology one would expect 58 of these to be unifoliate and 58 to be rosulate. The χ^2 value for the difference between expected numbers in each class and observed numbers in each class was calculated as 4.96 with one degree of freedom, and the *p* value was between 0.05 and 0.025. If one early acting dominant locus determines form during juvenile development such a deviation from a 1:1 ratio should only occur in between 2.5% and 5% of cases, so the results from this cross during early development are unlikely to be consistent with a 1:1 ratio.

During adult development, 103 plants were scored as rosulate, and 13 as unifoliate. The expected numbers of plants in each class were 87 and 29 respectively. The χ^2 value for the difference between expected numbers in each class and those observed was 11.3, and the P value was >0.005 with one degree of freedom. If two dominant loci determine form during adult development a deviation from a 3:1 ratio should occur less than 0.5% of the time.

At a 5% significance level, and in the case of equal viability of alleles in plants, we should reject Oehlkers hypothesis that rosulate form is determined by two dominant loci, one of which comes into effect later than the other.

Possible reasons that my results deviated from mendelian ratios

In the *S. dunnii* x (*S. rexii* x *S. dunnii*) backcross (family 1), we found a ratio of 3:1 at flowering, consistent with Oehlkers' reports. Contrary to Oehlkers' reports, in the *S. wittei* (*S. rexii* x *S. wittei*) backcross morphology segregated in a ratio of 7.92:1. This is significantly different to the 3: 1 ratio expected in a backcross if two genes determined morphology. There are several possible reasons for this:

(1) more than two genes affect morphology.

To address this I checked the likelihood of misinterpretation of Oehlkers' results using the χ^2 test. Where the plants were scored early (crosses 1 and 4), the predicted ratio is 3:1, and where they were scored late, the predicted ratio is 15:1 (crosses 2 and 3).

Cross	Time of scoring	Total number of plants	Total rosulate	Total unifoliate	χ^2	P value	Source: Oehlkers
1. *S. rexii* x *S. grandis* F2	6 months	186	145	41	0.867	10-50%	1941, 1964
2. *S. rexii* x *S. grandis* F2	9 months	186	169	17	2.645	10-50%	1941, 1964
3. *S. wendlandii* x *S. rexii* F2	10 months	51	48	3	0.011	90-97.5%	1938
4. *S. wendlandii* x *S. rexii* F2	8 months	438	318	120	1.34	10-50%	1938

Table 7: Summary of the results of Oehlkers' crosses.

In each cross and at each time of scoring this supported his inference. Also in the *S. dunnii* x (*S. rexii* x *S. dunnii*) backcrosses, a 3:1 ratio of rosulate to unifoliate form was seen, suggesting an alternative explanation.

(2) alleles may not have equal viability in backcross plants.

It is possible that unifoliate form was under-represented because seeds/seedlings carrying unifoliate alleles were less viable than those carrying rosulate alleles. This could be addressed by comparing percentage progression to maturity between seeds from unifoliate and rosulate species. Genetic viability could also be impaired if chromosomal inheritance in backcross plants was not mendelian, and this could be verified by karyotyping representatives.

(3) genetic background may have a substantial effect on phenotype.

As two different species were crossed to make the F1, heterozygosity in F1 and backcross plants should be high, and may encompass many variable genes that have an effect of morphology. In other species *knox* phenotypes appear particularly dependent on background. For example, failure to establish a meristem in KNOTTED loss-of-function mutants in maize is background dependent (Vollbrecht *et al.* 2000).

(4) my scoring system underestimated the number of unifoliate plants.

In some species of *Streptocarpus* classified as unifoliates, extra phyllomorphs do develop at the base of the petiolode, but these are generally subsidiary with respect to the cotyledonary phyllomorph. This would suggest that the unifoliate character is

sometimes plastic, and the overall form of these species still appears unifoliate. It is possible that the scoring system I developed under-estimated the number of unifoliate plants, as all plants with more than one phyllomorph were scored as rosulate, regardless of the size of secondary phyllomorphs. Out of the 103 plants scored as rosulate, there were ~17 plants which only had two phyllomorphs, and the second phyllomorph was substantially smaller than the first. Re-adjusting the numbers allocated to each class to account for this resulted in a nearly exact 3:1 ratio of rosulate to unifoliate phenotype (86 and 30 plants respectively). Out of the different possibilities, it seems most likely that scoring should take into account plasticity of the unifoliate character, as for scoring the *S. dunnii* backcross plants, and if my results are adjusted in accordance with this, they support Oehlkers' assertion that two loci determine inheritance of the rosulate character.

Analysis of leaf gnarling in backcross plants

Another trait that appeared to segregate in the back crosses was the degree of leaf gnarling. *S. rexii* has quite flat leaves, whereas *S. wittei* has gnarled leaves. This was of particular interest as *knox* gene expression has been linked to gnarling phenotypes in several other species (figure 10, introduction). A qualitative valuation of the degree of gnarling was developed. Four classes were chosen: (1) leaves flat as in *S. rexii*, (2) an intermediate class, with leaves mainly flat, (3) an intermediate class, with leaves mainly gnarled, and (4) leaves extremely gnarled (figure 13).

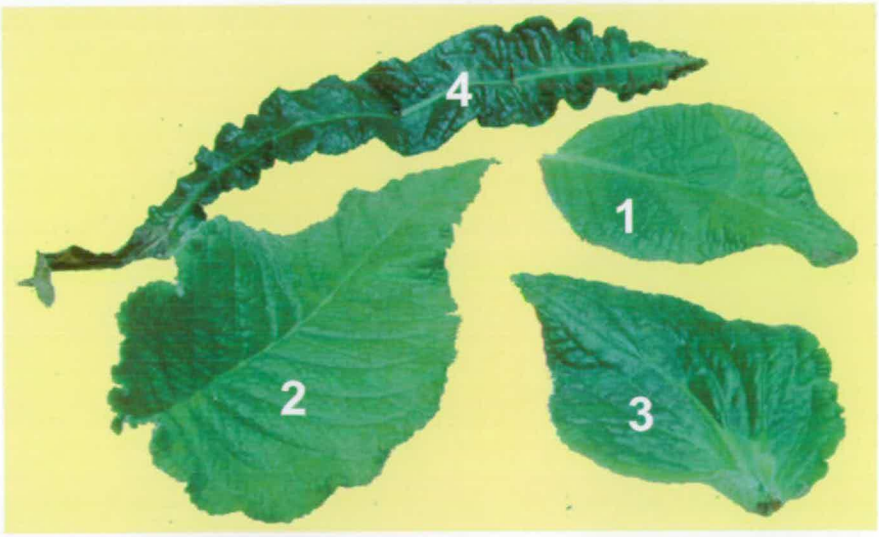


Figure 13: Leaves showing different degrees of leaf gnarling in the *S. wittei* (*S. rexii* x *S. wittei*) backcross. Character states were (1) leaves flat, (2) an intermediate class, with leaves mainly flat, (3) an intermediate class, with leaves mainly gnarled, and (4), leaves extremely gnarled.

In total 73 plants were scored of which 52 had flat leaves and 20 had gnarled leaves (table 8 below, results given in full in table 13, appendix 1). This approximates to a 3:1 ratio of flat: gnarled leaves, and in a χ^2 goodness of fit test the value was 1.736. The *p* value was between 0.5 and 1, i.e. such a deviation from the 3:1 ratio should occur between 10% and 50% of the time. This is therefore consistent with the flat leafed trait being determined by two loci, and the flat character being dominant with respect to gnarled.

	leaves flat	intermediate (2)	intermediate (3)	leaves gnarled
number of plants	7	46	15	5

Table 8: Summary of scores of leaf gnarling in backcross plants.

Some curiosities in form

As well as segregating for unifoliate and rosulate character, and leaf gnarling, the *Streptocarpus* backcross plants showed other interesting phenotypes. After flowering many plants started to grow leaves prolifically at the base of the inflorescences, from the groove meristem. These leaves were clearly distinguishable from leaves formed up to flowering, by their more rounded shape and erect stance (figure 14). Their growth is interesting as inflorescence meristems in *Streptocarpus* typically do not generate leaves. Out of 116 *S. wittei* (*S. rexii* x *S. wittei*) backcross plants, 78 (67%) grew extra leaves after flowering. Of the remaining 33%, 17% did not form extra leaves, and 16% were either dead, or had not yet flowered. Oehlkers did not report this phenomenon, and it may be that his plants did not survive past flowering.

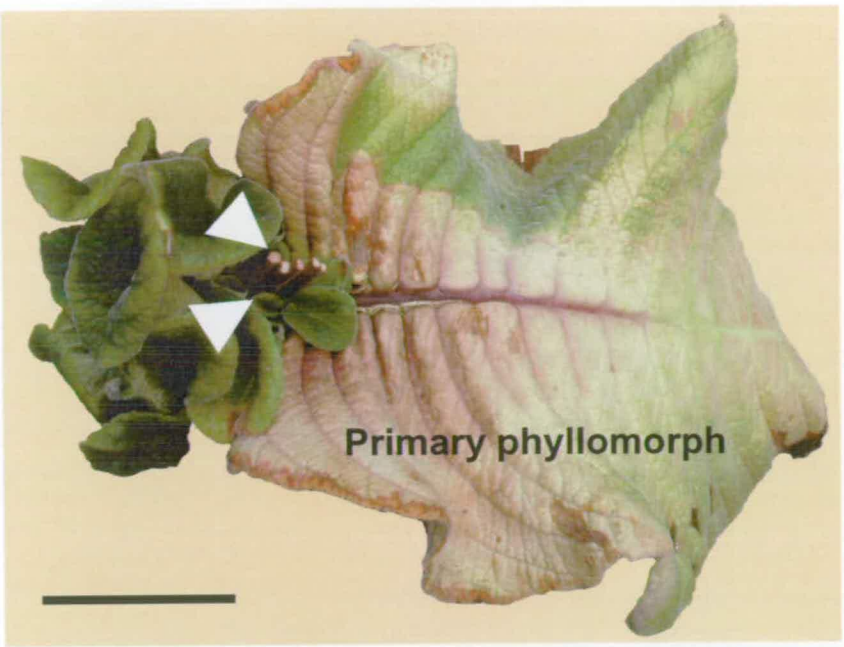


Figure 14: Production of secondary leaves from the groove meristem after flowering. Top arrow indicates cut off stem of the first inflorescence, bottom arrow indicates a new leaf arising from the groove meristem. Scale bar =5 cm.

As well as production of leaves after flowering plants occasionally produced long petiolate, peltate leaves (figures 15 A) which were reminiscent of leaves produced in *cupiliformis* mutants of *Arabidopsis*. Some plants also grew many small leaves all along the midrib (figure 15 B). This phenotype is strongly reminiscent of *Arabidopsis KNAT1* mutants, which sometimes also grow many ectopic leaves along the midrib. These unusual phenotypes were each only seen in 2-3 of the backcross plants, but despite their disrupted form, plants still flowered. A more prevalent phenotype was an increase in the amount of midrib-like tissue in leaves. This was accompanied by the disruption of leaf form (figure 16), which was seen in about one third of backcross plants, but not in parent species.

Figure 15 A: backcross plant with a long petiole (top arrow), and recurved lamina, which are reminiscent of leaves in *cuplilformis* mutants. B: inflorescences (top arrow) and secondary phyllomorphs arising from the midrib of a primary phyllomorph in a backcross plant.

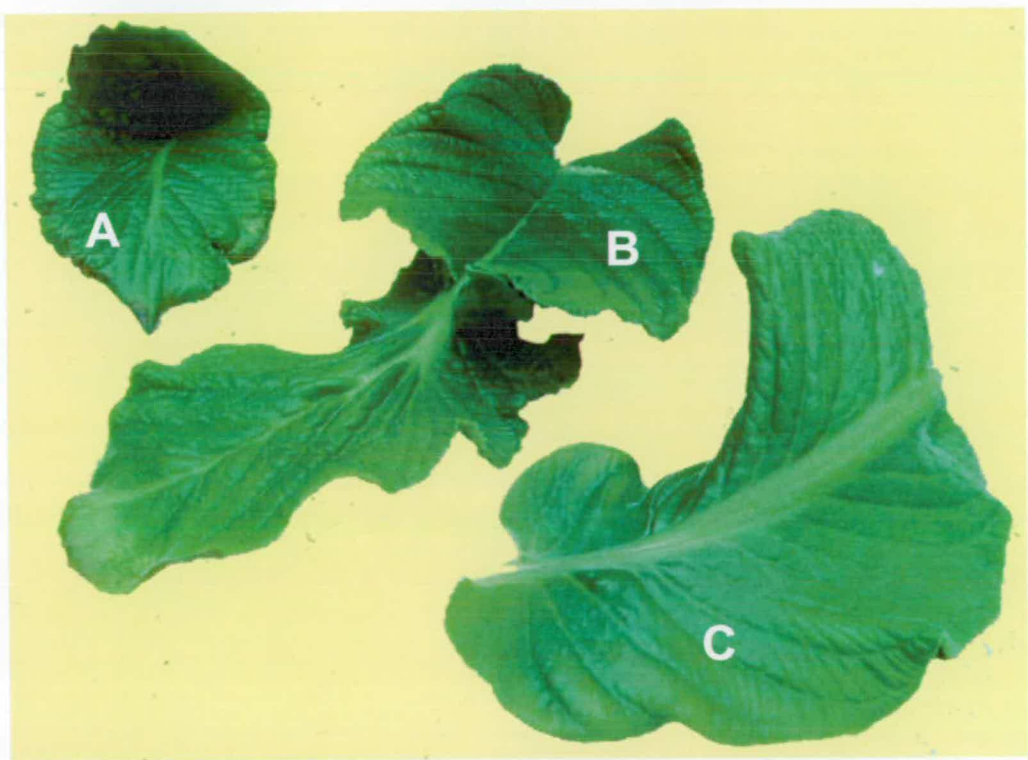


Figure 16: Further leaf phenotypes in *S. wittei* (*S. rexii* x *S. wittei*) backcross plants. (A) a peltate leaf, (B) and (C), leaves with midvein overgrowth and disrupted form.

Summary of results

This work supports Oehlkers' observations that two dominant loci determine *Streptocarpus* form if a fairly lax definition of unifoliate form is adopted. If a more rigorous definition is adopted, it seems that more than two loci affect form. This study did not clearly support the early action of one locus. In back crosses a leaf gnarling phenotype also segregated in a 3:1 ratio that was suggestive of determination by two loci that were partially co-dominant. In the back crosses form was sometimes extremely perturbed, but this did not disrupt the plants' capacity to flower.

Chapter 4: Characterisation of genes which could effect form in *Streptocarpus*

To investigate the possible effect of genes important in meristem function in model species on *Streptocarpus* morphology, I first needed to clone their *Streptocarpus* homologues, and to examine where they were expressed.

Isolation of *Streptocarpus STM*-like genes

The genes that I was most interested in when I started my project were the *STM*-like *knox* genes for reasons explained in the introduction. At the outset of my research, partial *Streptocarpus STM* sequences had been isolated from three species, which were representatives of the three major growth forms of *Streptocarpus,* by M. Möller (RBGE). These species were *S. dunnii* (unifoliate), *S. rexii* (rosulate) and *S. saxorum* (caulescent), which are grown in the RBGE *Streptocarpus* collection.

The most conserved part of *knox* genes is the homeodomain, and it is interrupted by an intron (intron 3). The primers used by M. Möller (Fc and GR22) were designed against the consensus between available *Arabidopsis* and *Antirrhinum* homeodomain sequences, and the region amplified encompassed a short stretch of coding sequence and intron 3, amounting to about 200 bp in total (figure 17).

The first step of my research was to repeat Michael Möller's experiments with primers that he kindly donated, and I amplified products spanning intron 3 from genomic DNA in the same three species used by Michael. From some species I obtained PCR products of different lengths. One fragment from all species matched Michael's 200 bp fragment in size, and the other was about 450 bp. In *Antirrhinum*, *STM* orthologues are duplicated (Golz *et al.* 2002), and it seemed that this might also be the case for *Streptocarpus*. However, the *knox* gene family is large, and comprises two classes with divergent function. To ensure that I specifically amplified *STM*-like genes and to discriminate between amplified genes, I needed to amplify sequence from less conserved parts of the genes.

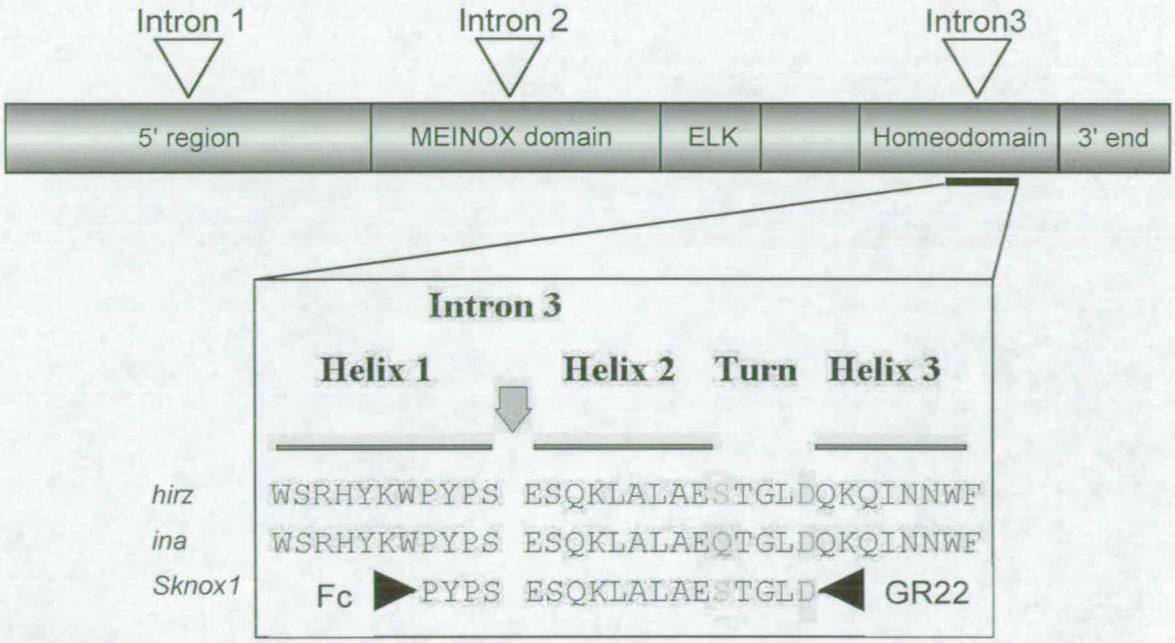


Figure 17: Diagram to show region of *Streptocarpus knox* homologues amplified in initial experiments on *Streptocarpus* (bar). Arrows indicate primer positions and direction of extension.

As there are no genetic libraries for *Streptocarpus* which could be screened for *STM*-like genes, I decided to use PCR based techniques to achieve this. PCR is dependent on location of primers in opposition to one another, so it is usually dependent on some knowledge of a gene's sequence for primer design. This presented a problem, as I had only a short stretch of 3' *Streptocarpus* sequence from which to design primers, and conservation between *knox* genes outwith the homeodomain was not sufficient to allow further degenerate primers to be designed. One way of getting round this was using Rapid Amplification of cDNA Ends in the 3' direction (3' RACE).

Amplification of 3' region

cDNA gene transcripts are polyadenylated at their 3' end, so antisense oligod(T) primers designed to complement the poly-A tail can be used in conjunction with a sense gene specific primer to amplify unknown 3' cDNA sequences (Frohman *et al.* 1988, Frohman 1995). Because RACE involves amplification from cDNA derived from poly(A) RNA, it has the further advantage of being likely to detect expressed sequences rather than pseudogenes. As *STM*-like genes are expressed in meristems in other species, I isolated RNA from shoot apical meristems in the caulescent species, and from inflorescence

meristems in acaulescent species. cDNA synthesis was primed using Q_t , a primer consisting of oligod(T) and a unique anchor sequence. Reverse transcription was followed by PCR with primers Fc and Q_o, which is equivalent to part of the anchor sequence of Q_t (figure 18). Where I did not have any products in the first round of amplification, a nested amplification was performed with a second gene specific primer, Fe, and a nested anchor primer Q_i. PCR products (about 330 bp) were gel purified and sequenced directly, or cloned and then sequenced.

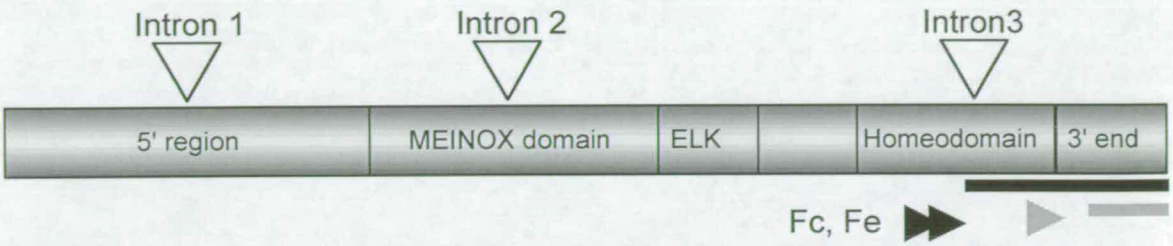


Figure 18: Diagrammatic representation of region amplified by 3' RACE. Black bar spans region amplified for *Streptocarpus knox1* genes, grey arrow indicates position of *KNAT1* primer, grey bar spans region spanned for *knox2*.

Sequences were used to BLAST search all databases available, and were compared to each other. This indicated that I had cloned the 3' region of one *STM*-like gene from all three species and the genes have been called *Sdknox1*, *Srknox1*, *Ssknox1*. I also isolated a second *STM*-like gene from *S. dunnii* using *KNAT1* primer in place of Fc, in an attempt to clone *KNAT1* homologues from *Streptocarpus*, and this was called *Sdknox2*. Evidence that the sequences were from cDNA was found for *Srknox1*, *Ssknox1* and *Sdknox2*; the sequences terminated in a poly(A) tract, and contained a long open reading frame encoding a protein that was very similar to STM. *Sdknox1* had no poly(A) tract, and also had stop codons and a 3' consensus splice site at its 5' end, indicating that it was probably a genomic contaminant. All of these sequences are given in appendix 2.1.

Sknox copy number

Because I had occasionally amplified two products using degenerate primers and there are two *STM* homologues in *Antirrhinum* (a member of a sister family to the Gesneriaceae), I wanted to investigate the copy number of *Streptocarpus STM* homologues. To achieve this, I used the 3' sequences amplified in RACE as a probe for Southern hybridisation at high stringency. Genomic DNA was digested with *Eco* R1, *Sau*3 AI, *Eco*R V and *Hin*D III, and restriction fragments were separated by agarose gel

electrophoresis prior to Southern transfer to a nylon membrane. DNA was bound to the membrane, and then hybridised to a radioactively labelled probe comprising a pool of the 3' regions amplified from the three species with primers Probe1 and Probe2. Following washing, the filters were used to expose X-ray film, which was then developed. In each species, with each enzyme, at least two genomic restriction fragments hybridised to the probe. In each case one band hybridised more strongly than the other(s). Figure 19 shows the results obtained using *Eco*R V and *Hin*D III, which were most clear.

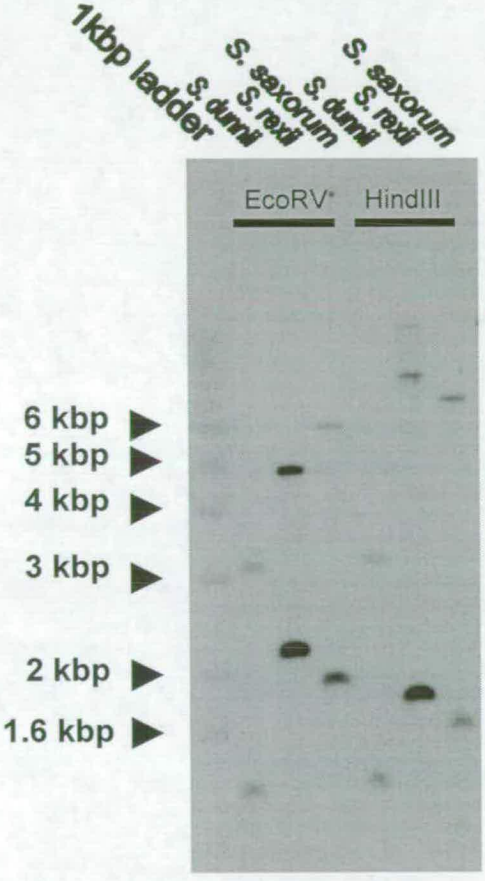


Figure 19: At least two restriction fragments from each species hybridise to a 3' *Sknox1* probe in Southern analysis.

Amplification of two products by PCR and RACE, and detection of two fragments in Southerns, suggested that there were two closely related *STM*-like genes in *Streptocarpus*, as in *Antirrhinum*. However, the 3' region of *knox* genes is relatively highly conserved. Therefore to examine the affinities of *Streptocarpus knox* (*Sknox*)

genes and achieve maximum specificity in expression analysis, I also needed to obtain 5' sequences.

Cloning 5' *Sknox* sequences

Extension in the 5' direction using 5'RACE

The first method I used to attempt this was 5' RACE, which operates on a similar principle to 3' RACE. In 5' RACE, cDNA synthesis is primed using an antisense gene specific primer (GSP1 for *S. saxorum*; GSP4 for acaulescent species) and the anchor is provided by a short homopolymeric sequence added to the 5' end of the cDNA using terminal deoxynucleotidyl transferase. Nested gene specific primers (GSP2 and GSP3 for *S. saxorum*; GSP5 and GSP6 for acaulescent species) were then used in PCR with adaptor primers, AAP and UAP, complementing the new tail sequence (figure 20). Although I did amplify *knox* sequences using 5' RACE, they were either short, or more similar to other *knox* genes than to *STM* (*SSknox*3), and I repeatedly amplified non-specific products (table 14 appendix 2.6). As the specificity of 5' RACE is partly dependent on the specificity of the gene specific primers, and my primer design was still constrained by the short stretch of sequence available to me in the most conserved part of the gene, I felt that another approach was merited.

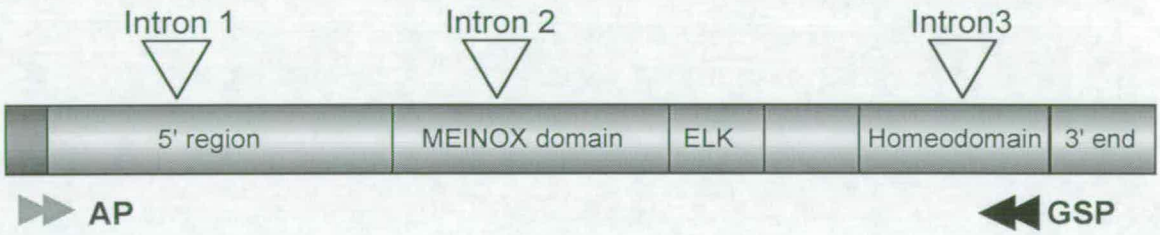


Figure 20. Location of primers for 5' RACE. Sense adaptor primers (grey) are used in conjunction with antisense gene specific primers (black).

Amplification to ELK domain

The next approach I used was inverse (I) PCR (Ochman *et al.* 1988, Triglia *et al.* 1988). In IPCR, genomic DNA is digested with a restriction enzyme that leaves a 4 bp overhang, and is then circularised by ligation. Dilution of the restriction fragments favours intramolecular ligation rather than ligation to other fragments in the pool. I first used *Hin*DIII, which allowed me to predict amplified fragment by comparison to results from Southern blotting; 3.5 or 1 kbp for *S. dunnii* for 1.8, 9, or 12 kbp *S. rexii*, and 1.6 or 7 kbp for *S. saxorum*. A portion of the ligated fragment pool was used in PCR with back-to-back primers, amplifying round the loop and across the re-joined restriction site.

I designed primers against the 3' region already cloned from each species, and used a new primer, Probe1, as the sense primer and GSP3 (*S. saxorum*) or GSP2 (*S. dunnii*, *S. rexii* and *S. wittei*) as the antisense primer. In each case only one product per species was obtained, and these were of the expected sizes; 3.2 kbp, 1.8 kbp, 1.6 kbp and 2 kbp for *S. dunnii*, *S. rexii*, *S. saxorum* and *S. wittei*, respectively.

These products were cloned, and sequenced. As clones were larger than could be sequenced from the initial primers, additional internal primers were also used for sequencing. For *S. saxorum*, two extra primers (UPP1, antisense; saxdown, sense) were designed against the new sequence and used in further sequencing reactions. For the acaulescent species, UPP1 was used in conjunction with a third primer, rexdown. I sequenced complete clones from *S. saxorum* and *S. rexii*, and partial clones from *S. dunnii* and *S. wittei*. For all four species the sequence showed similarity to the ELK domain of other knox genes. The 5' *Hin* DIII restriction site was at the same position in the ELK domain for each species. The 3' restriction site in the *S. dunnii* clone was about 2 kbp downstream of the coding region, and therefore, I did not sequence to this site. Similarly, for *S. wittei*, I only sequenced the region I needed for subsequent experiments.

Amplification to intron 2

Because first round IPCR clones extended only to the ELK domain, a second round of IPCR was carried out. I had detected an *Eco* RI site 3' to the *Hin* DIII site in the ELK domain that would minimise the amount of overlap between fragments. I attempted a second round of amplification with all four species, using primers designed against the sequences of round 1 IPCR products. For *S. saxorum*, using newsaxup (antisense) and newsaxdown (sense), I amplified another 1.6 kbp fragment which I cloned and sequenced. Using newrexup (antisense) and newrexdown (sense) primers, the other species yielded non-specific products that ran as multiple bands or a smear in agarose gel electrophoresis.

Loss of specificity in IPCR

In a third round of IPCR, specificity was also lost for *S. saxorum*, although three new enzyme and primer combinations were used. The first used *Bam* HI, and PCR with Saxfurtherup (antisense) and Saxfurtherdown (sense) primers. I then attempted amplification from *Sac* I restriction fragments using primers SFU (antisense) and SFD (sense). Finally I attempted amplification with *Hin* DIII restriction fragments and

primers SH3U (antisense) and SH3D (sense). These attempts might have been unsuccessful because I had to locate new primer combinations in an intron, which was A-T rich. Therefore primer specificity was likely to be low. The full extent of fragments isolated by IPCR is illustrated in figure 21, below.

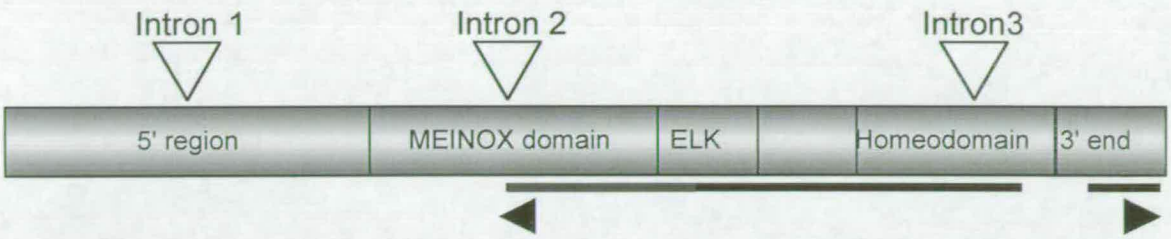


Figure 21: Diagrammatic representation of regions amplified using IPCR. Black bars indicate regions amplified in round 1 (which also extended ~ 2 kbp beyond the 3'end), grey bar indicates region amplified in round 2 (continued into intron 2), Arrows indicate directions of extension.

Increased specificity of amplification using genome walking

The next technique I used to isolate *Sknox1* was genome walking. For genome walking, genomic DNA is digested with a restriction enzyme with a 6 base recognition site producing blunt ends. Digested fragments are ligated to a double stranded adaptor which is blunt at one end, and has a recessed 3' terminus blocked by an amine group at its other end. This prevents extension of the shorter oligo in PCR, which would create a binding site for an amplification primer. This means that priming in the first cycle of PCR can only occur from a gene-specific primer, and then from fully extended PCR products in subsequent rounds. The adaptor sequences are also complementary, so if non-specific products form with adaptor sequence at both ends, they form a pan handle structure that inhibits amplification. Specificity is also promoted by the use of long gene specific primers, and high annealing temperatures in PCR (Tm= ~ 70°C).

Amplification of the 5' end of Ssknox1

I digested genomic DNA with *Eco*RV and ligated the cut fragments to adaptors. Antisense gene specific primers (WALK1 and WALK2) were used in nested amplification with adaptor primers AP1 and AP2 (figure 22). One fragment was amplified, which I cloned and sequenced. Part of the sequence was rich in mono and di-nucleotide repeats, including a TATA repeat, and had stop codons in all three frames. The remaining part consisted of two large open reading frames the first of which started with an ATG translation start sequence. Both showed high similarity to *STM*. The open

reading frames were separated by a region of non-coding sequence which was flanked by consensus intron splice sites, and was at the same position as an intron in other *knox* genes.

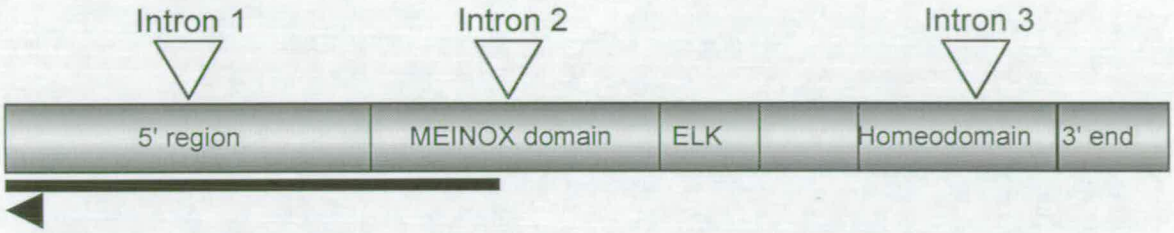


Figure 22: Region amplified using genome walking. Arrow indicates direction of extension. Overlap with sequences from IPCR was in intron 2.

Amplification of 5' coding regions from acaulescent species

The new *Ssknox1* sequence showed high similarity to *Antirrhinum STM*-like genes, even beyond the ATG translation start site. I used this sequence to design a primer (MEGSD) to use in conjunction with primer ELKUP (designed against the ELK domain) to amplify 5' coding sequence from *S. dunnii*, *S. rexii* and *S. saxorum* cDNA in RT-PCR (figure 23).

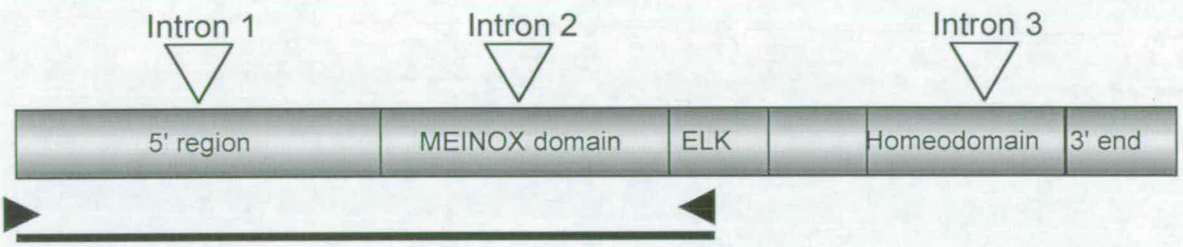


Figure 23: Region amplified using RT-PCR indicated by the bar. Arrows indicate direction of extension.

One product was amplified, from *S. dunnii* and *S. rexii*, and these were cloned and sequenced. Two products were amplified, cloned and sequenced from *S. saxorum*, one of which had a 48 bp deletion at the position of intron 1 in the genomic sequence. All of the products showed high similarity to existing sequence of *Ssknox1* and published *STM*-like genes.

Compilation of contiguous sequences

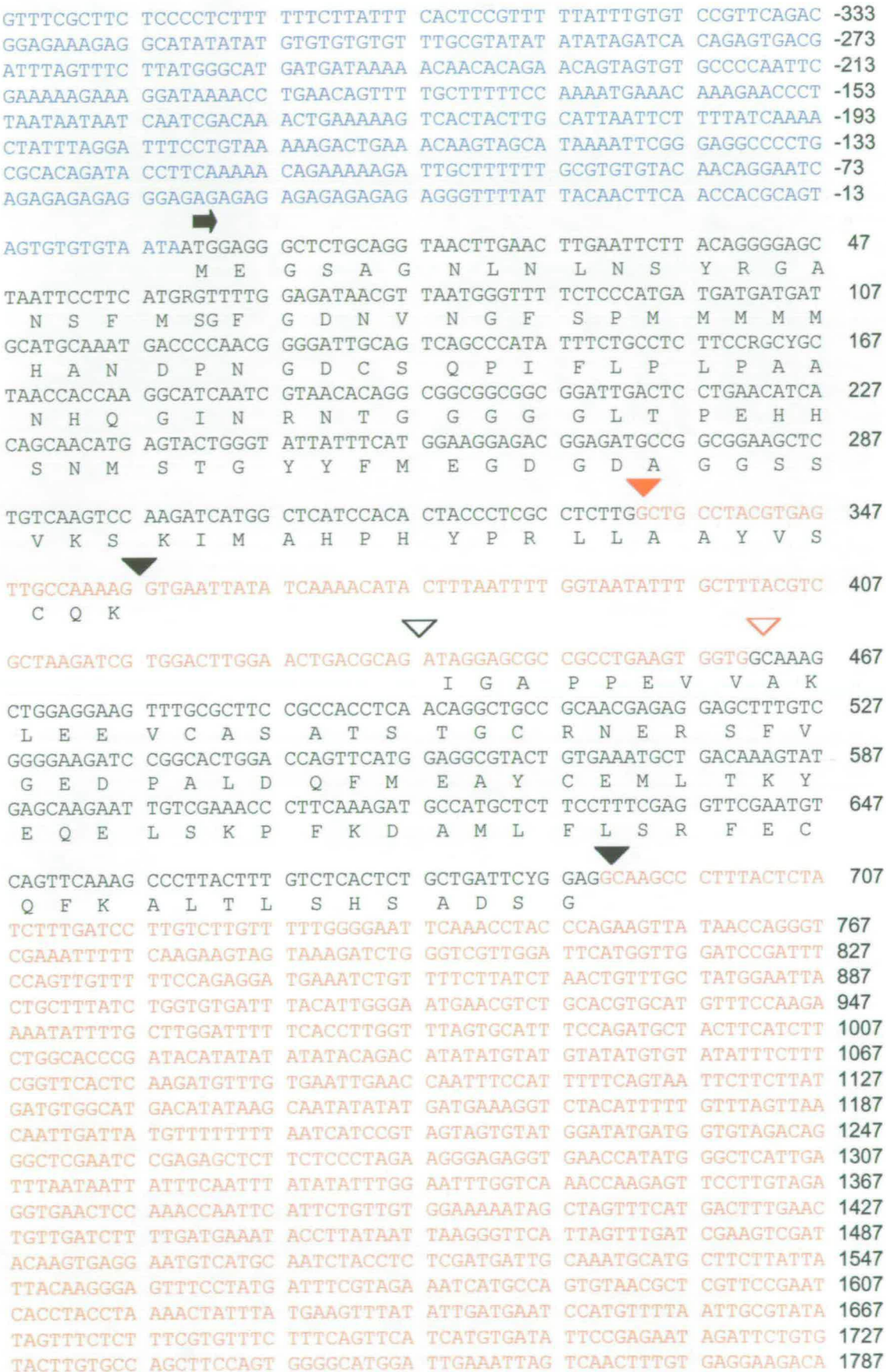
In each of the PCRs above I amplified fairly short stretches of sequence. These were amalgamated to produce a full length sequence from *S. saxorum*, *Ssknox1* (figure 24). To produce the full length *Ssknox1* sequence, I aligned fragments in Sequence Navigator, and checked that the overlapping sequences showed a high degree of similarity. I attributed single base pair substitutions (which were infrequent) to allelism, and where they arose they are recorded (figure 24). Although *S. dunnii*, *S. rexii* and *S. wittei* sequences remained fragmented, almost full length amino sequences could be deduced for *S. dunnii* (*Sdknox1*) and *S. rexii* (*Srknox1*) by amalgamating coding sequences from RT-PCR with those from IPCR and 3'RACE, again checking for sequence similarity (appendices 2.3 and 2.4). The sequence from *S. wittei* (*Swknox1*) remains incomplete.

Identity of isolated genes

BLAST searches of all available databases indicated that I had isolated Class 1 *knox* genes with homology to *STM*. To further investigate affinities of my sequences to other *knox* genes, I compared their structural features.

Sknox1 *genes show conserved features of* knox *genes*

Knox genes are unified by several conserved domains. These are (from N to C terminus) the MEINOX domain, the GSE box, the ELK domain, and the homeodomain (Bürglin 1997, Bürglin 1998), of which the homeodomain is the most conserved. The MEINOX domain has two conserved sub-domains, called *KNOX*1 and *KNOX*2, which are shared between *knox* and MEIS genes, their animal homologues (Bürglin 1997, Bürglin 1998). *knox* genes are also subdivided into two classes, partly by the presence of an intron in the ELK domain of Class 2 *knox* but not Class 1 genes (Kerstetter *et al.* 1994). The genes that I isolated from *Streptocarpus* shared all of the known conserved domains typical of *knox* genes, and the *Sknox1* sequences did not have an intron in the ELK domain (figures 24 and 25, and appendix 2.2), suggesting that they belonged to Class 1.


GTTTCGCTTC TCCCCTCTTT TTTCTTATTT CACTCCGTTT TTATTTGTGT CCGTTCAGAC -333
GGAGAAAGAG GCATATATAT GTGTGTGTGT TTGCGTATAT ATATAGATCA CAGAGTGACG -273
ATTTAGTTTC TTATGGGCAT GATGATAAAA ACAACACAGA ACAGTAGTGT GCCCCAATTC -213
GAAAAAGAAA GGATAAAACC TGAACAGTTT TGCTTTTTCC AAAATGAAAC AAAGAACCCT -153
TAATAATAAT CAATCGACAA ACTGAAAAAG TCACTACTTG CATTAATTCT TTTATCAAAA -193
CTATTTAGGA TTTCCTGTAA AAAGACTGAA ACAAGTAGCA TAAAATTCGG GAGGCCCCTG -133
CGCACAGATA CCTTCAAAAA CAGAAAAAGA TTGCTTTTTT GCGTGTGTAC AACAGGAATC -73
AGAGAGAGAG GGAGAGAGAG AGAGAGAGAG AGGGTTTTAT TACAACTTCA ACCACGCAGT -13
AGTGTGTGTA ATAATGGAGG GCTCTGCAGG TAACTTGAAC TTGAATTCTT ACAGGGGAGC 47
M E G S A G N L N L N S Y R G A
TAATTCCTTC ATGRGTTTTG GAGATAACGT TAATGGGTTT TCTCCCATGA TGATGATGAT 107
N S F M SG F G D N V N G F S P M M M M M
GCATGCAAAT GACCCCAACG GGGATTGCAG TCAGCCCATA TTTCTGCCTC TTCCRGCYGC 167
H A N D P N G D C S Q P I F L P L P A A
TAACCACCAA GGCATCAATC GTAACACAGG CGGCGGCGGC GGATTGACTC CTGAACATCA 227
N H Q G I N R N T G G G G G L T P E H H
CAGCAACATG AGTACTGGGT ATTATTTCAT GGAAGGAGAC GGAGATGCCG GCGGAAGCTC 287
S N M S T G Y Y F M E G D G D A G G S S
TGTCAAGTCC AAGATCATGG CTCATCCACA CTACCCTCGC CTCTTGGCTG CCTACGTGAG 347
V K S K I M A H P H Y P R L L A A Y V S
TTGCCAAAAG GTGAATTATA TCAAAACATA CTTTAATTTT GGTAATATTT GCTTTACGTC 407
C Q K
GCTAAGATCG TGGACTTGGA ACTGACGCAG ATAGGAGCGC CGCCTGAAGT GGTGGCAAAG 467
I G A P P E V V A K
CTGGAGGAAG TTTGCGCTTC CGCCACCTCA ACAGGCTGCC GCAACGAGAG GAGCTTTGTC 527
L E E V C A S A T S T G C R N E R S F V
GGGGAAGATC CGGCACTGGA CCAGTTCATG GAGGCGTACT GTGAAATGCT GACAAAGTAT 587
G E D P A L D Q F M E A Y C E M L T K Y
GAGCAAGAAT TGTCGAAACC CTTCAAAGAT GCCATGCTCT TCCTTTCGAG GTTCGAATGT 647
E Q E L S K P F K D A M L F L S R F E C
CAGTTCAAAG CCCTTACTTT GTCTCACTCT GCTGATTCYG GAGGCAAGCC CTTTACTCTA 707
Q F K A L T L S H S A D S G
TCTTTGATCC TTGTCTTGTT TTTGGGGAAT TCAAACCTAC CCAGAAGTTA TAACCAGGGT 767
CGAAATTTTT CAAGAAGTAG TAAAGATCTG GGTCGTTGGA TTCATGGTTG GATCCGATTT 827
CCAGTTGTTT TTCCAGAGGA TGAAATCTGT TTTCTTATCT AACTGTTTGC TATGGAATTA 887
CTGCTTTATC TGGTGTGATT TACATTGGGA ATGAACGTCT GCACGTGCAT GTTTCCAAGA 947
AAATATTTTG CTTGGATTTT TCACCTTGGT TTAGTGCATT TCCAGATGCT ACTTCATCTT 1007
CTGGCACCCG ATACATATAT ATATACAGAC ATATATGTAT GTATATGTGT ATATTTCTTT 1067
CGGTTCACTC AAGATGTTTG TGAATTGAAC CAATTTCCAT TTTTCAGTAA TTCTTCTTAT 1127
GATGTGGCAT GACATATAAG CAATATATAT GATGAAAGGT CTACATTTTT GTTTAGTTAA 1187
CAATTGATTA TGTTTTTTTT AATCATCCGT AGTAGTGTAT GGATATGATG GTGTAGACAG 1247
GGCTCGAATC CGAGAGCTCT TCTCCCTAGA AGGGAGAGGT GAACCATATG GGCTCATTGA 1307
TTTAATAATT ATTTCAATTT ATATATTTGG AATTTGGTCA AACCAAGAGT TCCTTGTAGA 1367
GGTGAACTCC AAACCAATTC ATTCTGTTGT GGAAAAATAG CTAGTTTCAT GACTTTGAAC 1427
TGTTGATCTT TTGATGAAAT ACCTTATAAT TAAGGGTTCA TTAGTTTGAT CGAAGTCGAT 1487
ACAAGTGAGG AATGTCATGC AATCTACCTC TCGATGATTG CAAATGCATG CTTCTTATTA 1547
TTACAAGGGA GTTTCCTATG ATTTCGTAGA AATCATGCCA GTGTAACGCT CGTTCCGAAT 1607
CACCTACCTA AAACTATTTA TGAAGTTTAT ATTGATGAAT CCATGTTTTA ATTGCGTATA 1667
TAGTTTCTCT TTCGTGTTTC TTTCAGTTCA TCATGTGATA TTCCGAGAAT AGATTCTGTG 1727
TACTTGTGCC AGCTTCCAGT GGGGCATGGA TTGAAATTAG TCAACTTTGT GAGGAAGACA 1787

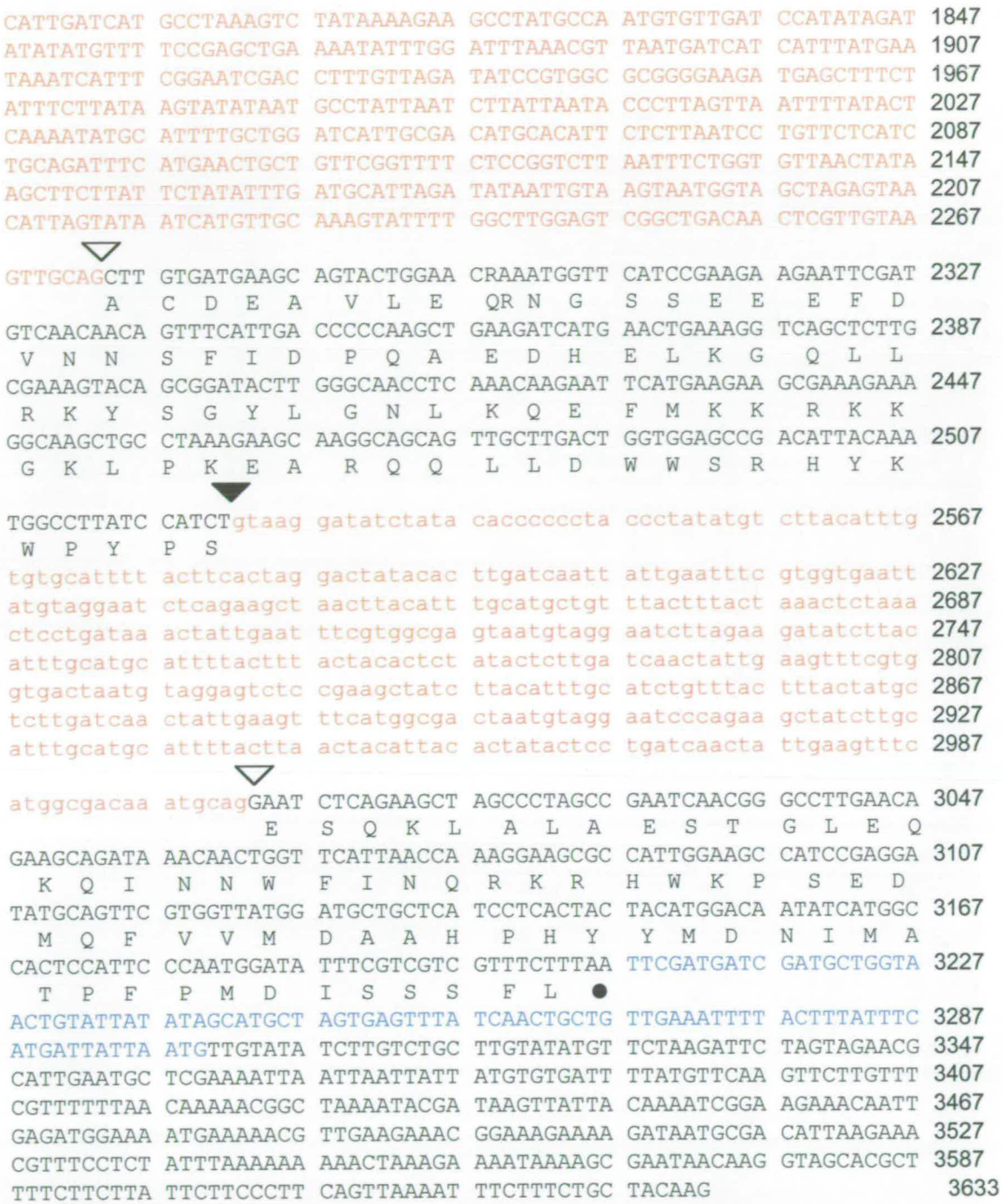


Figure 24: Genomic sequence of a *Streptocarpus saxorum STM* homologue, *Ssknox1*. Coding sequence is shown in **black**, and introns are shown in red. Intron position was indicated by comparison of genomic sequence to cDNA sequences. Splicing alternatives were found for intron 1. Splice donor sites are indicated by a filled triangle, and splice acceptor sites are indicated by an unfilled triangle. Contiguous cDNA sequences are given using a splice donor and acceptor sequences of the same colour. Untranslated regions are shown in blue. The stop codon is indicated with a filled circle.

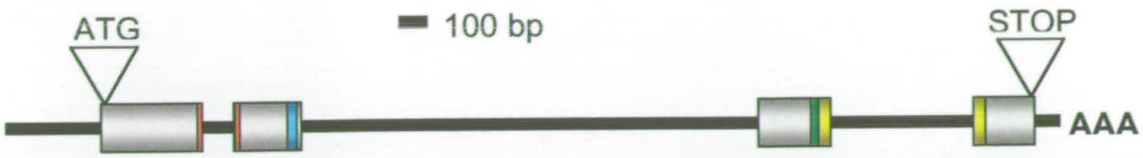


Figure 25: Diagrammatic representation of *Ssknox1* gene structure. Boxes represent exons, and lines represent introns and untranslated regions. Red boxes highlight the position of the *KNOX* 1 domain, blue box represents *KNOX* 2, the green box indicates the position of the ELK domain and yellow boxes highlight the homeodomain.

Similarity between Streptocarpus knox *genes*

The *Sknox1* genes all showed a high degree of similarity to each other, even in unconserved parts of the gene, and it is therefore probable that they are orthologues. Support for this is given by the fact that I amplified more than one *STM*-like product when I was using primers Fc and GR22, designed against the most conserved part of the gene. In nearly all other PCRs I was able to amplify *Sknox1* from all species using standard primers designed against a single species. The second *STM*-like sequence, *Sdknox2*, was amplified with a degenerate primer. This comprised the region 3' of the homeodomain. *Sknox1* sequences in this region showed very high likeness to each other, and to *STM*. The *Sknox2* sequence was also *STM*-like, but was not so similar to *Sknox1* sequences, and it was therefore designated a separate orthologue. Phylogeny using the available *Sdknox2* sequence with other genes, and intron 3 sequences, also supported this classification, and these data are presented later in this chapter, and in chapter 6 respectively.

Similarity of Streptocarpus *genes to other* knox *genes*

To test the affinity of *Ssknox1* to two other Class 1 *knox* genes, *Knotted1*, *STM* and *Ssknox1* amino acid sequences were compared. Sequences were aligned in ClustalW (Thompson *et al.* 1994). Naming and boundaries of conserved domains were defined as in Bürglin (1998) and are not drawn to scale. Pairwise percentage similarity was calculated as (number of bases identical for that region/total number of bases) x 100. This comparison showed a high degree of similarity between the three genes, and that *STM* was more similar to *Ssknox1* (74.5 % similarity) than to *knotted1* (53% similarity). It also showed that the 5' region is far less conserved than other domains (figure 26).

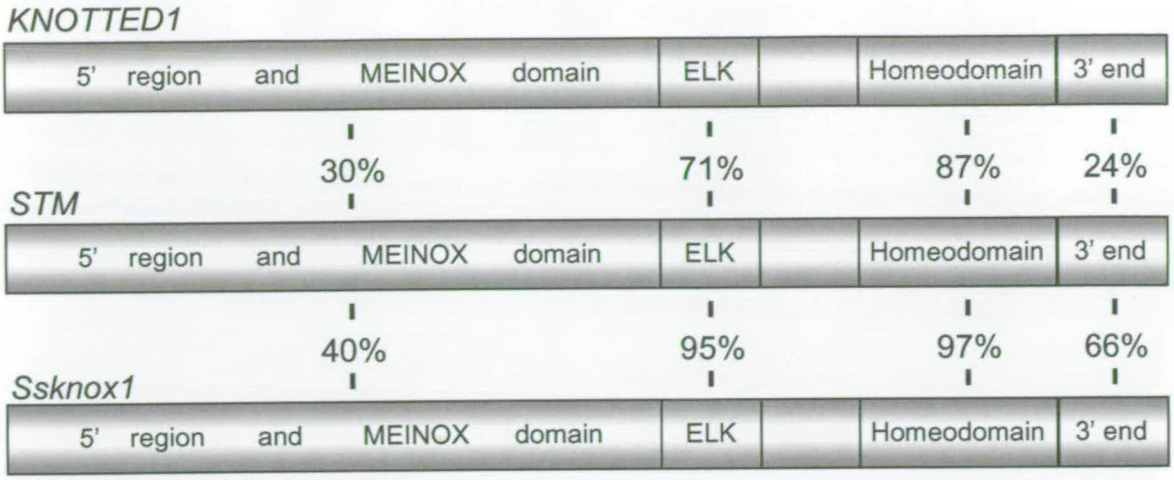


Figure 26: Percentage similarity of *STM* to *knotted1* and *Ssknox1*.

Differences in structure between Sknox1 *sequences, and possible functional implications*

When aligned in CLUSTALX, *Streptocarpus* amino acid sequences showed most similarity to the *Antirrhinum* homologues, *ina* and *hirz*, and more similarity to *ina* than *hirz*. For *S. saxorum*, splice variants were found within the conserved MEINOX domain (figures 24 and 27).

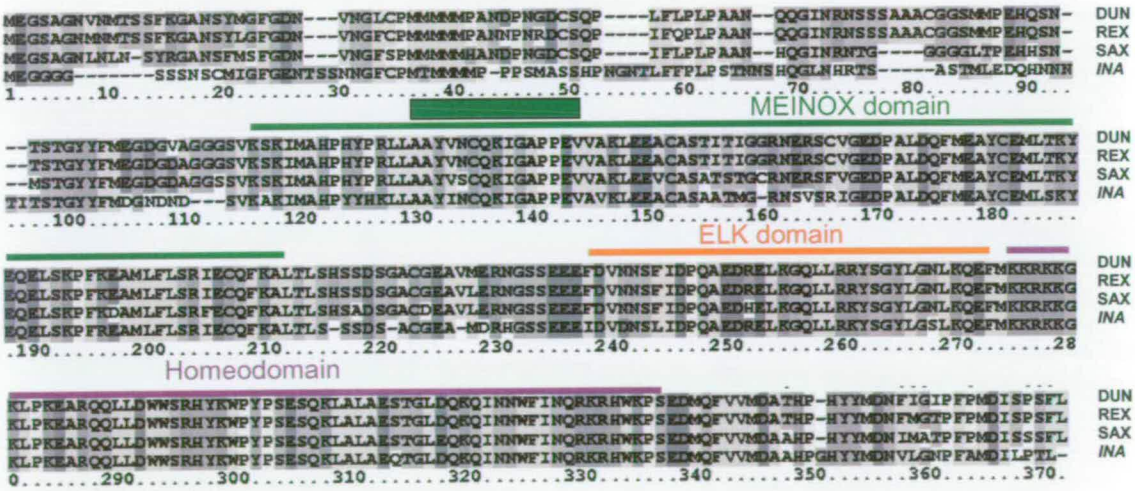


Figure 27: *Knox* amino acid sequence alignment. A 16 amino acid deletion from the MEINOX domain which would result from a splice variant in *S. saxorum* is indicated by the green box.

The different domains of *knox* genes are thought to have different functions, and the splice variant from *S. saxorum* resulted in a deletion of 48bp from the conserved MEINOX domain (figure 27). Structural analysis of the rice *knox* gene *OSH15* has

shown that, when expressed under a constitutive promoter, constructs containing deletions in the MEINOX domain do not effect plant morphology. Therefore the MEINOX domain is important in inducing altered leaf morphology when entire *knox* genes are expressed in the leaf (Nagasaki *et al.* 2001). The MEINOX domain also appears to be important for suppression of target gene expression and protein dimerisation (Bellaoui *et al.* 2001, Nagasaki *et al.* 2001). It would be very interesting to test whether the two transcripts from *S. saxorum* are found in the same or different domains, and therefore if transcripts could have different functions.

Phylogeny

Sequence similarity provides one means of comparing new data to existing data, and by similarity *Sknox1* genes appear to be *STM*-like. However, the *knox* genes comprise a large gene family. To compare the function of genes between species we first need to establish that the genes are truly derived from a shared ancestor (that they are orthologous). In the case of genes within a gene family this issue is complicated, as genes may duplicate within a genome (i.e. they are paralogues) or may be inherited during taxonomic divergence (orthologues). In both cases the genes are homologous, but for comparison of gene function between species it is necessary to establish orthology. Because phylogeny is based on the premise of inheritance of characters through shared common descent, it offers a test of the relationships between genes, and of their evolution, i.e. of their orthology.

Homeodomain nucleotide sequence phylogeny

I used two different data sets to test the relationship of *Streptocarpus knox* genes to other *knox* genes by phylogeny. Firstly I used the nucleotide sequence of the homeodomain from all available plant *knox* sequences on GenBank, and three animal MEIS homeobox genes. In total 47 sequences from animals, algae, mosses, ferns, conifers, monocots and dicots were included. Sequences were aligned unambiguously, and provided a phylogeny directly comparable to work published by Reiser *et al.* (2000). I performed a heuristic analysis in PAUP 3.1.1, with simple addition of sequences and tree bisection-reconnection, and this strategy was used for all subsequent searches. Only one most parsimonious tree was found, which I rooted on the animal taxa. Tree statistics for this and all subsequent analyses are included in appendix 4.3. For the *knox* phylogeny, where available, more than one gene from each taxon was included. By comparison to

published work, knox sequences were expected to fall into at least two classes. This phylogeny divided the *knox* genes as had previous phylogenies into Class 1 and Class 2 (figure 28). The *STM*-like genes formed a monophyletic group, as did the *knotted*-like genes, but the two groups were distant.

If genes used to construct phylogeny are orthologous, species relationships between the taxa sampled should be maintained. With the taxa sampled, Class 1 and Class 2 genes were each expected to be nested within a paraphyletic group containing moss, fern and conifer genes. For neither class of genes was this the case. In the Class 1 clade, *mkn4* and *mkn2* (moss: highlighted by black arrows) form a sister group to the other genes, but *crknox1*, *crknox2* and *Picea* (fern and conifer: highlighted by black arrows) are interspersed throughout this group. In the Class 2 group, Moss and fern (*mkn* Class 2, and *crknox3*: highlighted by black arrows) sequences are terminal taxa. Even within groups of far more closely related taxa, species relationships are not maintained. In the *STM* clade, the *Antirrhinum STM* homologues *hirz* and *ina* (highlighted by pink arrows) fall with genes from the Solanaceae (*Tkn2* and *NTH15*) and not with *Ssknox1* which is taxonomically closer. This implies that the sequence data used are either not orthologous or are homoplastic. As the sequences are taken from the homeodomain which is highly conserved at the amino acid level, and for nucleotide data there are only four character states, it is more likely that the data are homoplastic, and therefore are not informative with regards to finding orthology between genes.

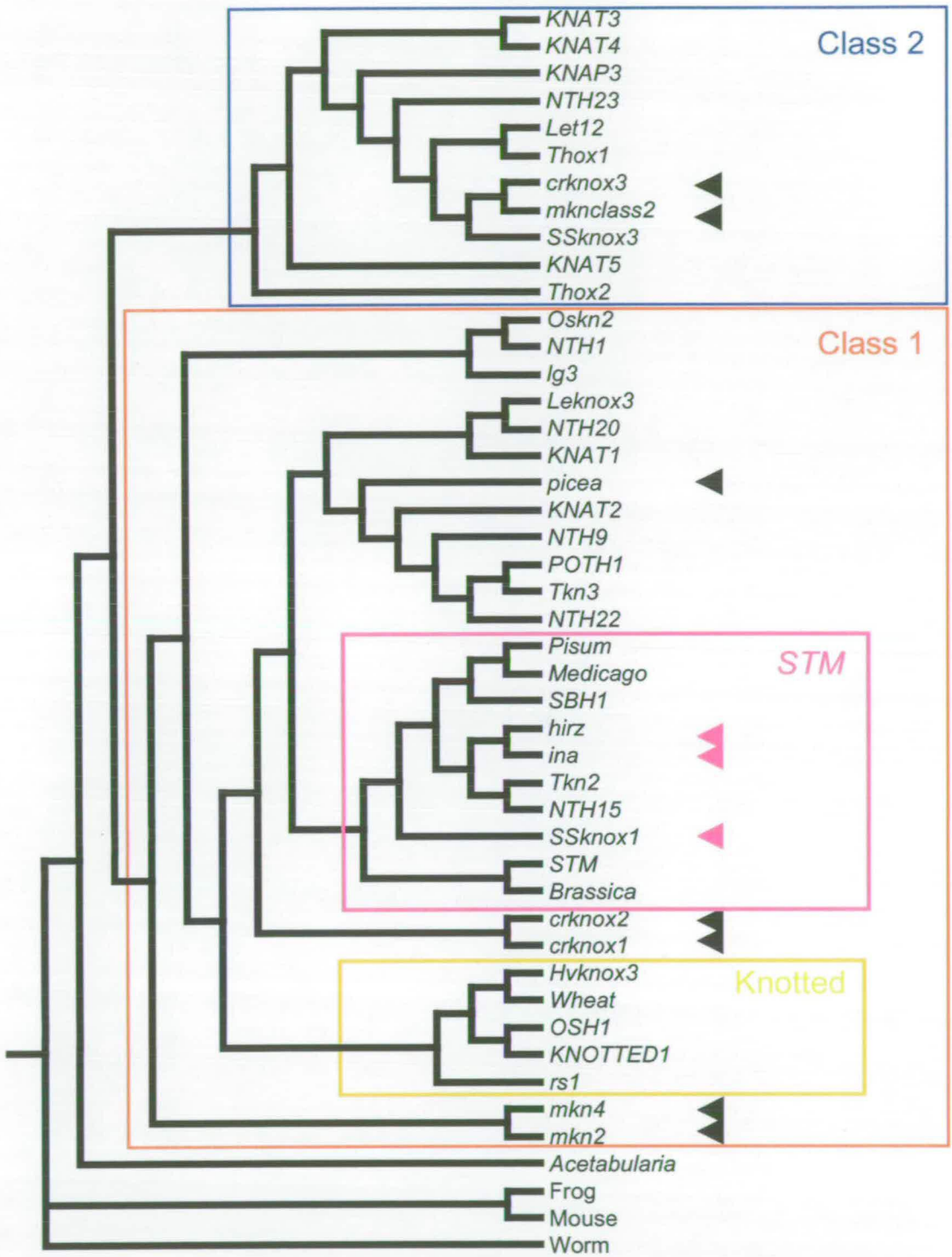


Figure 28: Homeodomain nucleotide sequence phylogeny of *knox* and MEIS genes. Accession numbers for all taxa are included in appendix 6.

Amino acid sequence phylogeny

As the homeodomain sequence did not resolve species relationships, and was probably homoplastic, I decided to use sequences outwith the homeodomain for further analyses. Attempts to align other regions of nucleotide sequence were inhibited by the high divergence between sequences. I therefore switched to using full length deduced amino acid sequences which aligned well in ClustalX, necessitating a smaller data set. I used PAUP (Version 3.1.1) with standard settings to perform analyses. I rooted the phylogenies on the Class 2 *knox* genes, which are a sister group to the Class 1 genes (Bharathan *et al.* 1997, Reiser *et al.* 2000). Two most parsimonious trees were found, which differed only in their relative position of *Antirrhinum* homologues. In one tree *hirz* was sister to the *Streptocarpus* homologues, and in the other, *ina* took its place. In the strict consensus which is shown, all other nodes were preserved.

One way of selecting those characters that are most informative in building a phylogeny is using successive weighting. This selects characters that have high consistency from a data set, and allows them to be used to reconstruct a phylogeny, in essence as a new data set (Kitching *et al.* 1998). Successive weighting can fundamentally change the topology of phylogenies, as well as the number of most parsimonious trees.

When analyses were performed with reweighted characters using a rescaled consistency index (mean value), or with a reduced data set comprising the Class 1 *knox* genes alone, only one most parsimonious tree was found, and *ina* appeared more closely related to *Sknox1* genes than *hirz*.

Trends shown

This phylogeny again separated the Class 1 and Class 2 genes (figure 29). Within the Class 1 clade, a *KNOTTED* group and an *STM* group were formed, consistent with previous phylogenies. The monophyletic *KNOTTED*-like group shown here forms a monocot sister group to the dicot *STM* like genes. However, one dicot gene (*KNAT1*) that generally falls within the *KNOTTED* clade severely disrupted my alignments, so I excluded it from my analyses.

Sknox1 genes are most closely related to *hirz* and *ina*, two *Antirrhinum knox* genes (Golz *et al.* 2002), and fall within a monophyletic group comprising *STM* and *STM* orthologues. This group is paraphyletic with respect to a monocot-only *KNOTTED* clade. The monophyly of the dicot clade indicates that with the exception of duplication in *Antirrhinum*, the *STM* clade is probably a single gene phylogeny of *STM* orthologues.

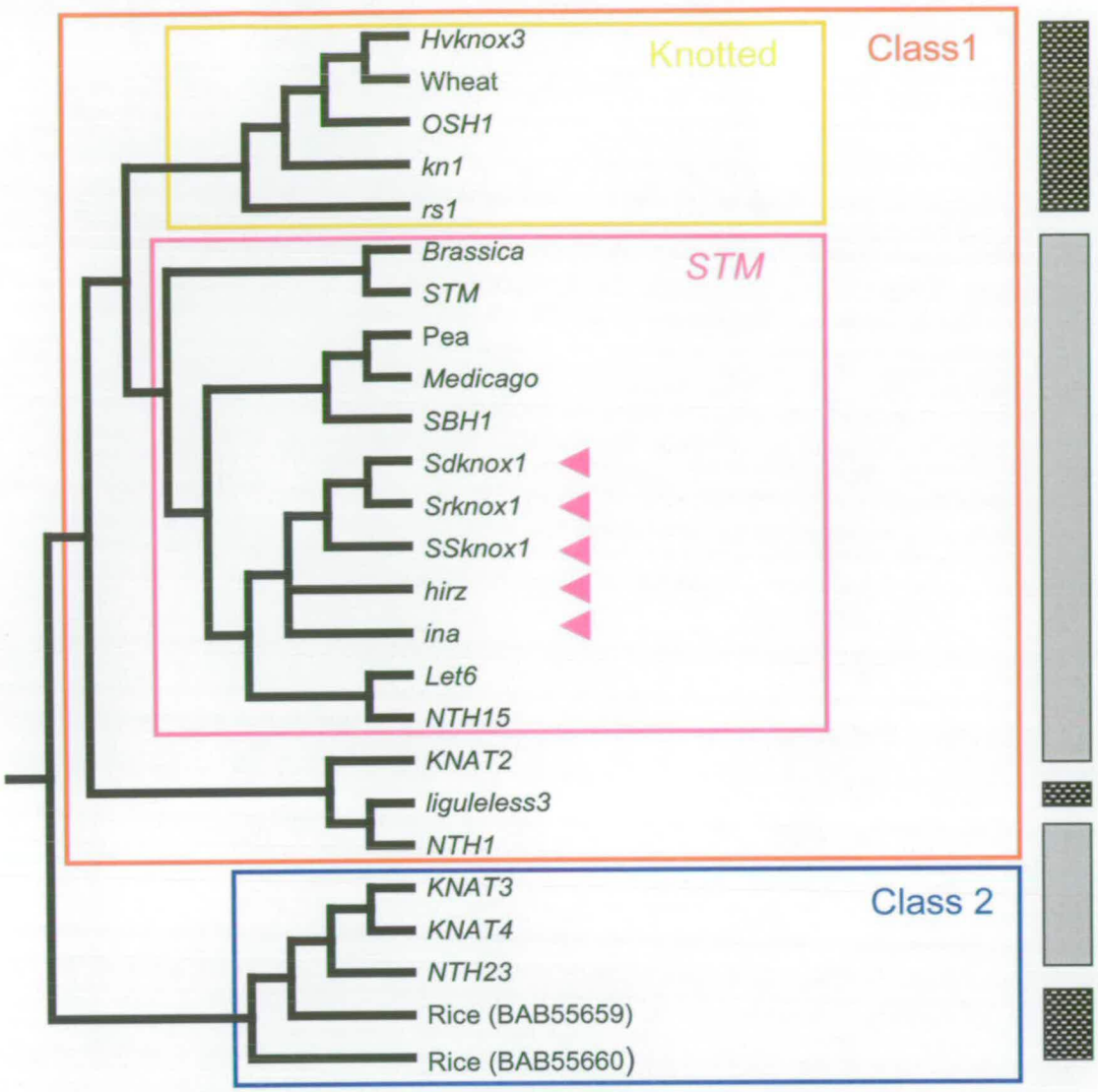


Figure 29: Phylogeny of *KNOX* proteins. *Streptocarpus* homologues fall unequivocally in the dicot *STM* gene clade, and are most closely related to *Antirrhinum* sequences. Class 1 and Class 2 *knox* genes are indicated, as are monocot(▩) and dicot (□)taxa. Accession numbers for sequences used were: *Hvknox3* (*Hordeum vulgare*, AF022390), wheat (*Triticum aestivum* AF224500), *OSH1* (*Oryza sativa*, P46609), *kn1* (*Zea mays*, S14283), *rs1* (*Z. mays*, L44133), Brassica (*Brassica oleracea*, AF193813), *STM* (*Arabidopsis thaliana*, U32344), pea (*Pisum sativum*, AF063307), Medicago (*Medicago trunculata*, AF308454), *SBH1* (*Glycine max*, P46608), *Let6* (*Lycopersicon esculentum*, T04317), *NTH15* (*Nicotiana tabacum*, T01735), *KNAT2* (*A. thaliana*, P46640), *lg3* (*Z. mays* , AAD13611), *NTH1* (*N. tabacum*, BAA76750), *KNAT3* (*A. thaliana*, P48000), *KNAT4* (*A. thaliana*, P48001), *NTH23* (*N. tabacum*, T02220). *hirz* and *ina* are *Antirrhinum* sequences obtained within our lab. (Golz *et al.* 2002).

Sdknox2 *is the result of a duplication that occurred after the divergence of* Streptocarpus *and* Antirrinum

Although it was not possible to align the *Sdknox2* sequence with the other entire *knox* sequences to ascertain its position in phylogeny, the sequence available was used to test when the duplication that gave rise to *Sdknox2* occurred. This encompassed 72 bp of nucleotide sequence between the homeodomain and the stop codon. Solanaceae *knox* sequences, *hirz*, *ina*, *Sknox1* and *Sknox2* sequences were analysed. Phylogenies were complied using exhaustive searches of the nucleotide sequence and its translated amino acid sequence data, and were rooted on Solanaceae sequences, that were shown in other phylogenies of mine to be the closest taxa to the monophyletic group containing *Antirrhinum* and *Streptocarpus*. With the nucleotide phylogeny, six most parsimonious trees were generated. Characters were reweighted using the rescaled consistency index, and the analysis was repeated. Three most parsimonious trees were found, and their majority rule consensus is shown in figure 31 A (overleaf). For the amino acid sequence data only one most parsimonious tree was obtained, shown in figure 31 B.

The most parsimonious hypothesis regarding *Sknox* gene duplication in *Streptocarpus* and *Antirrhinum* is that duplication occurred once in their shared common ancestor. If this were the case, we could predict that each species succeeding that taxonomic divergence should contain both homologues, and that the topology of the gene tree including those duplicates should reflect relationships between species (figure 30 A). If the duplication succeeded taxonomic divergence, we would expect homologues within *Streptocarpus* to be more closely related to each other than to homologues from *Antirrhinum*, as in figure 30 B.

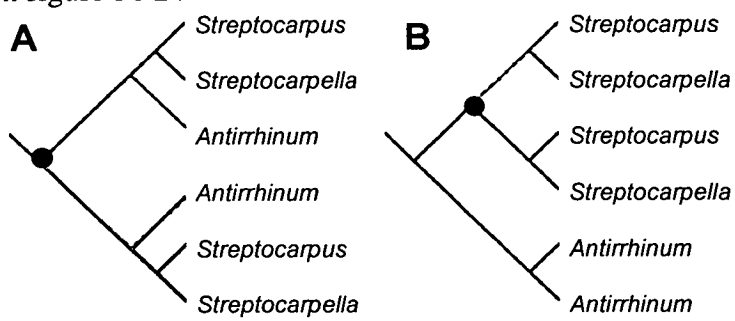


Figure 30: Hypothetical phylogenies to investigate the timing of gene duplication.

Phylogenies presented in figure 31 A and 31 B both show that the appearence of *Sknox* genes is likely to succeed the duplication that gave rise to *hirz* and *ina* in *Antirhinum*,

and therefore must have arisen form an independent duplication. From these analyses it is not clear whether the duplication preceded or succeeded the subgeneric split of *Streptocarpus* into *Streptocarpus* and *Streptocarpella*. As the topology of the nucleotide tree does not reflect species relationships within the *Streptocarpus* clade (the *Srknox1* homologue falls with the *Ssknox1* homologue), it seems more likely that the topology in B reflects the true relationship between *Streptocarpus* homologues, which would offer support for a duplication occurring after the subgeneric split of *Streptocarpus*.

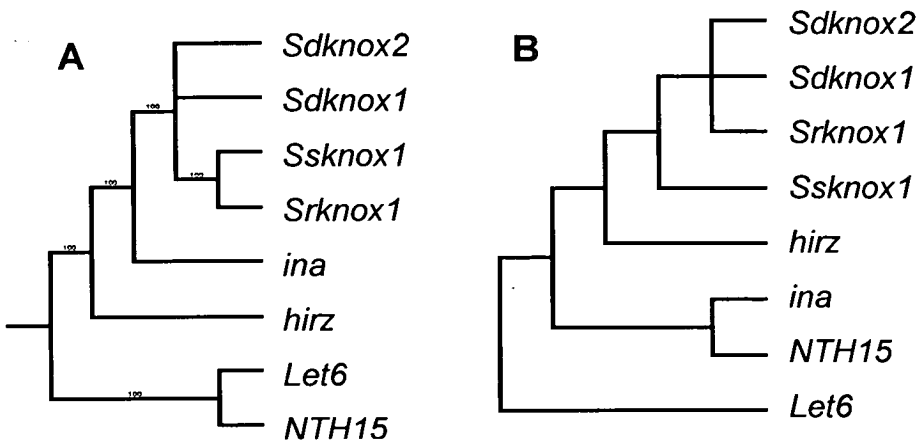


Figure 31: Affinity of *Sdknox2* to other *knox* genes. A: Majority rule consensus phylogeny based on nucleotide sequence data. Branches preserved in all most parsimonious trees are indicated by 100. B: Phylogeny based on amino acid sequence data

Expression of *Streptocarpus STM* homologues

As the *Streptocarpus STM*-like genes share common descent with other *STM*-like genes, which share conserved function, their function is also likely to be conserved. We can therefore suggest that *Streptocarpus* homologues should be expressed in a similar domain to other *STM*-like genes. To test this, I needed to see where they were expressed in each of the different forms of *Streptocarpus*.

Experimental design

Using primers designed against *Streptocarpus* consensus *Sknox1* sequences in RT-PCR (as in figure 23), cDNA fragments were amplified from different tissues of three species representing the three different growth forms. Further primers were designed against the consensus between a *Streptocarpus* 40s ribosomal subunit gene amplified in 5' RACE (ribosomal up and ribosomal down), and *Arabidopsis* 40s ribosomal subunit sequence available on TAIR. RT-PCR with these primers was used to control for the amount of cDNA added to each sample. For experimental and control samples, a genomic DNA

sample was included to control for the possibility that genomic DNA was being amplified rather than cDNA. In both cases the primers used spanned two introns. In the experimental (*knox*) samples, the introns comprised about 2 kbp. A DNA free sample was also included with both primer combinations to control for cDNA contamination. PCR products from experimental samples were transferred to nylon filters by Southern transfer, bonded to the filter. To verify their identity they were probed with a Dig-labelled riboprobe transcribed from the 5' region of *Ssknox1*.

Possible outcomes

By sequence analysis and phylogeny, *Sknox1* genes appear orthologous to *STM*. If *Sknox1* genes are functionally orthologous to STM, we could expect their expression in similar domains to *STM*. If *Sknox1* genes do not have the same function, we could expect their expression in different domains. Within the dicot taxa included in my phylogeny, those genes whose expression patterns are published are *STM*, *SBH1*, *hirz*, *ina*, *Let6* and *NTH15* (Ma *et al.* 1994, Long *et al.* 1996, Chen *et al.* 1997, Tamaoki *et al.* 1997, Golz *et al.* 2002). All of these bar *Let6* are expressed throughout the meristem but are excluded from lateral organs and their primordia. When mapped on to phylogenies, this expression pattern is plesiomorphic for the clade. The exceptional pattern shown by *Let6* is likely to be a derived state, which is associated with leaf compounding in tomato (Janssen *et al.* 1998).

Outcome

The genomic DNA controls amplified as predicted, and in both cases produced bands that were larger than those amplified from cDNA, indicating that cDNA synthesis had been successful. The ribosomal control gene was detected ubiquitously and at similar levels throughout the tissues sampled. Expression of *Sknox1* genes in the experimental samples varied widely within and between species (figure 32).

In the caulescent species, *S. saxorum*, *Sknox1* was expressed strongly in shoot apical meristems and the internode, as is conventional within the *STM* clade. As the caulescent form is plesiomorphic within *Streptocarpus*, this suggested that *Ssknox1* is functionally orthologous to *STM*, and behaves typically for the dicot *STM* clade. However, weak expression was also detected in the petiole. In *Antirrhinum*, which shows conventional expression of *hirz* and *ina* detected by RNA *in situ* hybridisation, low levels of *hirz* and *ina* expression were occasionally detected in leaves and petioles by RT-PCR (John Golz

pers. comm.). The apparent low level expression of *Ssknox1* detected in *S. saxorum* petioles may be similar, representing rare, ectopic transcripts and this could be verified by RNA *in situ* hybridisation to an *Sknox1* probe.

Acaulescent *Streptocarpus* showed radically different expression patterns, which also differed between the two species examined. Both species showed expression in the midrib. As inflorescences arise acropetally along the midrib in acaulescent *Streptocarpus*, it is possible that the expression detected came from presumptive inflorescence meristems. This explanation is unlikely, as expression was detected in both proximal and distal tissue samples, and inflorescences only form proximally.

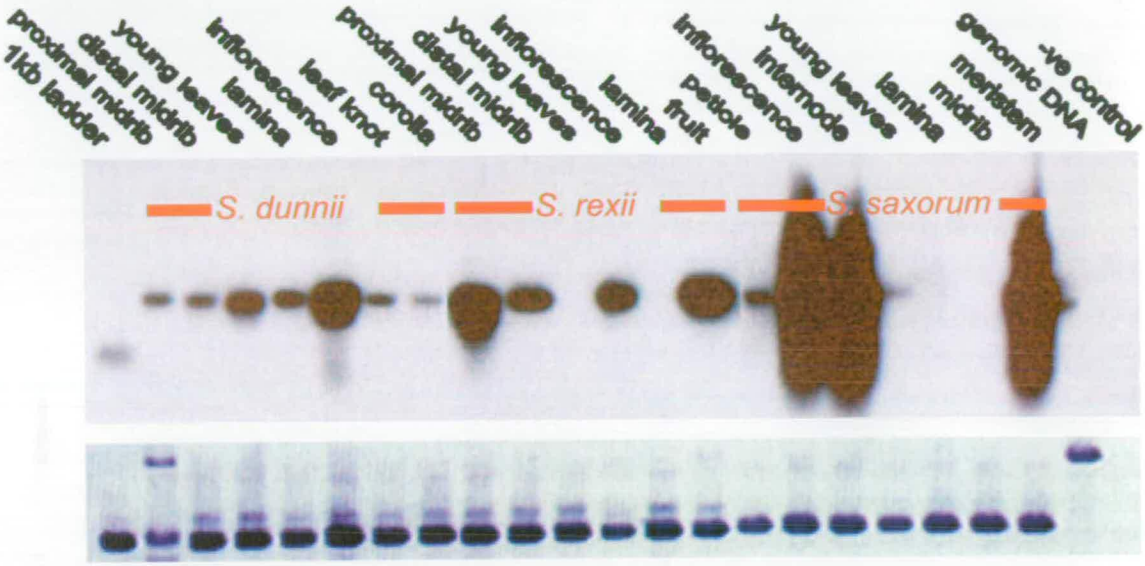


Figure 32: RT-PCR analysis of expression of *Streptocarpus STM* orthologues (top lanes) and 40s ribosomal subunit gene (bottom lanes). Samples in the top gel were transferred by Southern blotting to a nylon filter, bonded to the filter, and hybridized to a *S. saxorum* 5' *STM* probe to confirm the identity of the PCR products.

In the rosulate species analysed, *S. rexii*, expression was confined to the midrib, and was only apparent in older leaves.

In the unifoliate species, *S. dunnii*, expression was found throughout the plants, including the midrib, lamina, and corolla, all of which are lateral organs. Expression was also found in young leaves, contrasting with *S. rexii*, in which expression was confined to older leaves.

Levels of expression between samples varied, but as this experiment was not considered quantitative, it is not clear whether this had biological relevance.

RT-PCR also showed that *Sknox1* was expressed strongly in *S. rexii* fruits. As fruits contain many developing embryos, and in *Arabidopsis STM* is first expressed in the globular stage embryo, it is possible that the expression seen was embryonic. This could be confirmed using RNA in situ hybridisation to an *Sknox1* probe.

Meristematic properties in Streptocarpus *phyllomorphs*

Expression of *knox* genes in leaves of other species leads to promotion of less determinate fates (Lincoln *et al.* 1994, Chuck *et al.* 1996, Nishimura *et al.* 1999, Byrne *et al.* 2000). In phyllomorphic *Streptocarpus*, lateral organs have the capacity for continued growth. Phyllomorphs therefore share two properties normally associated with meristems; they are to an extent indeterminate, and have aberrant *knox* expression in the leaf.

Further meristem-like features of the phyllomorph

Another way of looking at meristems is that they hold pools of dividing cells. A marker for nuclear DNA replication, which precedes cell division, is provided by BrdU incorporation. BrdU is a thymidine analogue, and it is incorporated into DNA as it is synthesised. BrdU can be detected using antibodies, labelled with fluorophores which emit visible light when excited. To control for the possibility that all nuclei in a particular organ are dividing, sections need to be treated with a ubiquitous nuclear stain. I used DAPI, which forms fluorescent complexes with double stranded DNA, particularly in AT rich regions. Fluorescent emission occurs at a different wavelength to BrdU, so nuclei within a single section can be differentiated by the different colours of light emitted. These experiments were supervised by H. Tsukaya, in H. Uchimiya's laboratory at the University of Tokyo, and my visit was sponsored by the British Council and the Japanese Ministry for Science, MONBUSHO.

In *Monophyllaea*, another one-leafed member of the Gesneriaceae, BrdU uptake occurs to the base and sides of the cotyledon (Tsukaya 1997), and it has been suggested that these regions are meristematic in *Streptocarpus* (Jong and Burtt 1975). Incorporation of BrdU by seedlings of *S. rexii* was examined (figure 33). Each half of the figure depicts a section through a seedling, with the upwards pointing cotyledon forming a bulge at the top, and the hypocotyl beneath it.

The DAPI control shown on the left of figure shows an even distribution of fluorescent foci emitting blue light throughout the section, and the brightest of these are nuclei.

is greatly increased when it is bound to DNA. With BrdU staining (on the right) fewer nuclei were illuminated, and these were clustered to the base and sides of the cotyledon. This indicates that cell division, a further property usually associated with meristems, occurs to the base and sides of the cotyledon of *S. rexii* as in *Monophyllaea*.

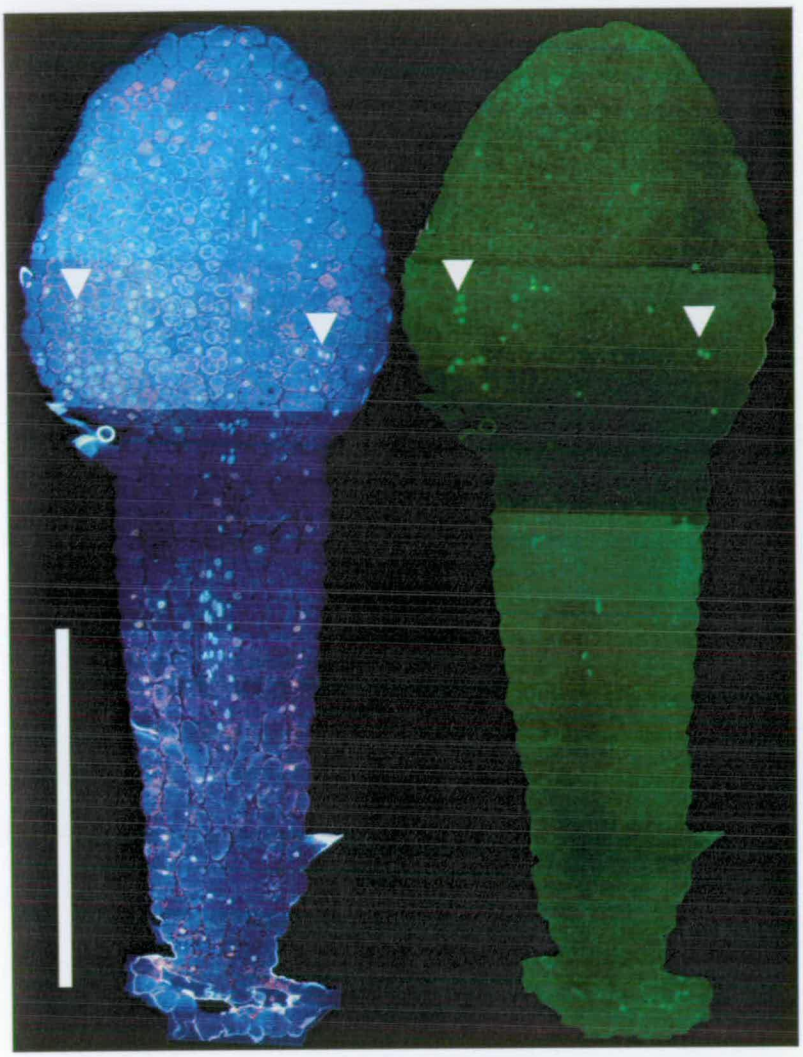


Figure 33: Sagittal section through *S. rexii* seedling stained with DAPI and BrdU. Arrows indicate double stained nuclei. Nuclei replicating DNA are mainly to the base and sides of the cotyledon, whereas other nuclei are more evenly dispersed. Scale bar = 40μm.

Further genes isolated from *Streptocarpus*

As discussed in the introduction, *STM* is one of several genes known to affect meristem function and plant growth in model species. Other genes are also candidates for induction of morphological change in *Streptocarpus*. I therefore cloned *Streptocarpus* homologues of some of these with a view to using them as further tools for understanding *Streptocarpus* form.

CLAVATA-*like genes*

A second set of candidates were *CLAVATA1*-like genes, encoding receptor-like kinases. *CLV1* in *Arabidopsis* is involved in regulating meristem size and is expressed only in the meristem. I was given degenerate primers (clavata forward and clavata reverse; figure 34) to *CLV1*-like genes by Richard Waites (University of York).

Figure 34: Diagrammatic representation of *CLV1*. Grey box represents the extracellular domain, darkest box represents the transmembrane domain, and the intermadiate box represents the kinase domain. Lines represent introns or untranscribed regions. Arrows indicate the position of primers *CLV*F and *CLV*R. Drawn from Clark *et al.* (1997).

Antirrhinum CLV homologues are expressed at very low levels (R. Waites; pers. comm.), so I used genomic DNA in PCR from a caulescent (*S. saxorum*), a unifoliate (*S. dunnii*) and a rosulate (*S. rexii*) species. I obtained PCR products of the predicted size (approx. 275 bp) from all three, and cloned and sequenced them. Of these, two of the products from *S. saxorum*, and one each from *S. dunnii* and *S. rexii* appeared to encode receptor-like kinases.

Relationships between CLV1*-like genes*

Antirrhinum has 3 *CLV1* homologues (AmA, AmB, and AmC), and I aligned the deduced amino acid sequence from *Streptocarpus* homologues with these and sequences available from GenBank using ClustalW (default parameters). The relationships between aligned sequences were analysed using PAUP (Version 3.1.1). Two most parsimonious trees were found in a heuristic search, and the strict consensus (figure 35) was computed

and rooted at its midpoint. This showed that two of the *Streptocarpus* genes are most closely related to the C class of *Antirrhinum CLV1*-like genes, and two are most closely related to the B class of *Antirrhinum* genes.

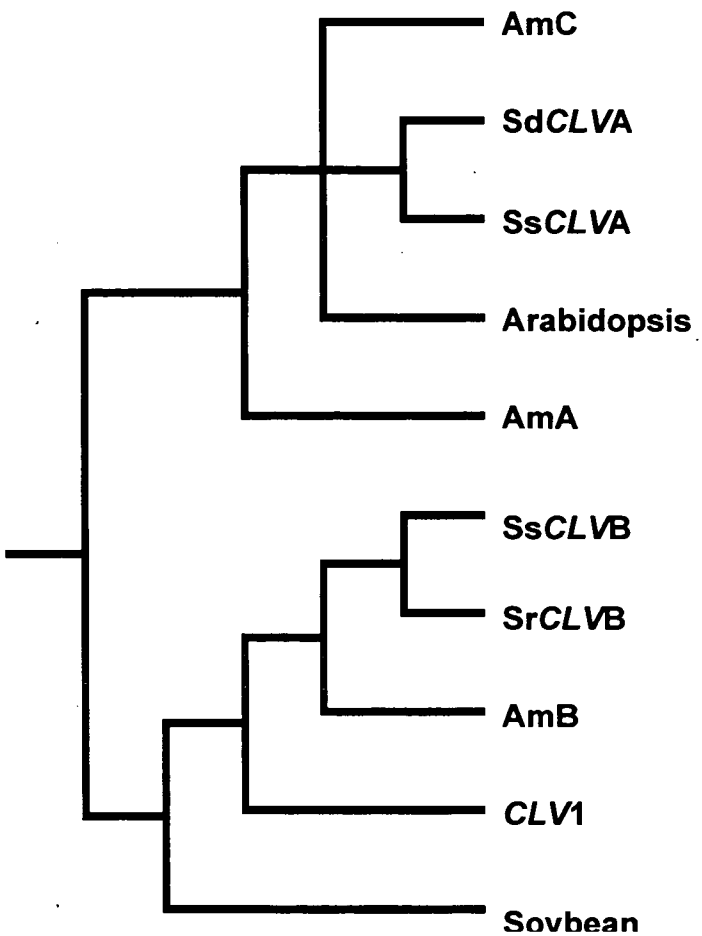


Figure 35: Phylogeny of *CLV1*- like genes. GenBank accession numbers of species used are: Sd*CLVA* (*S. dunnii*; AY061836), Ss*CLVA* (*S. saxorum*; AY061849), Sr*CLVB* (*S. rexii*; AY061838), Ss*CLVB* (S. saxorum; AY061842), *CLV1* (*Arabidopsis thaliana*; u96879), *Arabidopsis* (*A. thaliana*; Lambda clone T46033), Soybean (*Glycine max*; AF 197946), AmA, AmB, AmC (*Antirrhinum majus*; sequences donated by R. Waites).

NAC domain genes

A third set of candidate genes were NAC domain genes. Expression of two NAC domain genes in *Arabidopsis* is required for establishment of *STM* expression in the embryo, and their expression driven by the 35S promoter leads to ectopic meristem formation on the cotyledons (Aida *et al.* 1999, Takada *et al.* 2001). As primary phyllomorphs of *Streptocarpus* arise as a result of continued growth of the cotyledon, and phyllomorphs

show some meristem-like properties including *Ssknox1* expression, NAC homologues would provide a powerful complement to *STM* in expression analysis.

Isolation of a Streptocarpus *NAC gene*

I designed degenerate nested antisense primers (*CUC1* and *CUC2*) against the consensus between conserved domains of *Arabidopsis* and *Antirrhinum* NAC homologues (figure 36). These were used in 3' RACE with *S. dunnii*, *S. rexii*, and *S. saxorum*, the three species previously chosen to represent the three growth forms.

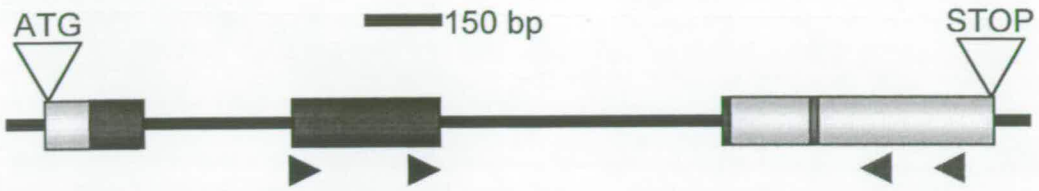


Figure 36: Diagrammatic representation of *NAM* (a NAC domain gene from *Petunia*) redrawn from Souer *et al.* (1996) and Takada *et al.* (2001). Lines represent introns or untranslated regions, boxes represent exons, darkest shading indicates the NAC domain, and paler shading indicates a conserved domain in the third exon. Black arrows indicate the position of degenerate primers used in 3' RACE, grey arrows indicated the position of genome walking primers.

I cloned and sequenced the 300 bp PCR products I obtained, one of which (from *S. dunnii*) showed homology to NAC domain genes in a BLAST search. The presence of a polyadenyl tract, and long open reading frame indicated that this was likely to be a cDNA sequence. I used this sequence to design primers for use in genome walking (*CUC* walk1 and *CUC* walk2), and amplified a 900 bp fragment, which I cloned and sequenced. Part of this sequence was contiguous with the sequence I amplified using RACE, and part of it was new. In the contiguous region, there was a region of non-coding sequence, which contained consensus splice site sequences at each end, and was therefore probably an intron. In the new sequence, there was a similar tract, at the same position of an intron in *NAM* and *CUC1*, NAC domain genes in *Petunia* and *Arabidopsis*; this probably represented a second intron. Alignment of the deduced amino acid sequence with other NAC domain sequences available on GenBank suggested that the most 5' sequence extended almost to the ATG. The full nucleotide sequence, the deduced amino acid sequence and structure of the cloned region, are given in figures 37 and 38 (overleaf), and I named the gene *SdNAC1* (*Sd* for *S. dunnii*).

```
CGCCCGGGCA GGTCTGAATT TGCCACCTGG GTTTCGTTTC TTCCCCACAG ACGAGGAACT   60
  A  R  A    G  L  N    L  P  P  G    F  R  F    F  P  T    D  E  E  L
TCTGGTGCAG TACCTCTGCC GGAAAGTTGC CGGTCACCAT TTCCCCTTGC AGATAATCGG  120
  L  V  Q    Y  L  C    R  K  V  A    G  H  H    F  P  L    Q  I  I  G

AGAAATCGAT CTGTACAAAT TCGACCCATG GGATCTCCCC Tgtaagtaaa aactcgaaag  180
  E  I  D    L  Y  K    F  D  P  W    D  L  P
agtcgaatca aatttctttt gtccaagcta attttgaatt gttttttttt ttggcgcagG  240
                                                                  C
TAAGGCCATG TTTGGAGAGA AGGAGTGGTA TTTCTTCAGC CCAAGAGACA GGAAGTACCC  300
  K  A  M    F  G  E    K  E  W  Y    F  F  S    P  R  D    R  K  Y  P
TAACGGGTCC AGGCCAAACA GGGTCGCCGG AACCGGATAC TGGAAAGCCA CCGGGACCGA  360
  N  G  S    R  P  N    R  V  A  G    T  G  Y    W  K  A    T  G  T  D
TAAGATCATA ACCACAGAAG GAAGAAAAGT CGGTATTAAG AAGTCCCTTG TTTTCTATCA  420
  K  I  I    T  T  E    G  R  K  V    G  I  K    K  S  L    V  F  Y  Q
AGGGAAAGCG CCGAAAGGAT CGAAGACCGA TTGGATTATG CACGAATATC GCCTCTTTGA  480
  G  K  A    P  K  G    S  K  T  D    W  I  M    H  E  Y    R  L  F  E

ACCTTCTAGA AAACATGGCA GCTCAAAGgt aagacgatta agcctaaggt ctccatgtct  540
  P  S  R    K  H  G    S  S  K
tgtatatacc cgtgtgattg cttcatcttt tgtaaaataa tatcgatcat aaataatttt  600

taaaaaatta aaatttctca gaaatttaaa tatttttga tgggttttag TTGGATGATT  660
                                                        L  D  D
GGGTGCTTTG CCGGATTTAT AAGAAGAATT CGAGCAAACG GAGCATCAGC GCTGGTGAAG  720
W  V  L  C    R  I  Y    K  K  N    S  S  K  R    S  I  S    A  G  E
TTGAGAGCAG AGACTACAGT CACGGCTCTT CCCCGTCTTG CTCGTCGCAG TACGACAGCG  780
V  E  S  R    D  Y  S    H  G  S    S  P  S  C    S  S  Q    Y  D  S
TTTTGGAGTC GCTGGAACGG ATCGATGACC AAGGCTACGC CAATTTGTCG AGAATGGGCT  840
V  L  E  S    L  E  R    I  D  D    Q  G  Y  A    N  L  S    R  M  G
CTCTGAAGCT GGACCTTCAG AACTTGGGCT CTGGGAATTT CGACTGGGCC ACGTTAGCTG  900
S  L  K  L    D  L  Q    N  L  G    S  G  N  F    D  W  A    T  L  A
GGCTTAACCC TCTGCCCGAA CTTATGAAGG CCCAACATAA CCAGGCGAAT GGGAAATTCT  960
G  L  N  P    L  P  E    L  M  K    A  Q  H  N    Q  A  N    G  K  F
GATATGACAC TTATGCCCTT GTCTTCCAGC ATGCTGGGCC ATTCGATATG GATGTATTAC 1020
●
TACCTAATAT TAATATAAAA AAAA                                        1044
```

Figure 37: DNA sequence of *SdNAC1*. Coding sequence is shown in black, Intron sequence is shown in red, and untranslated sequence is shown in blue.

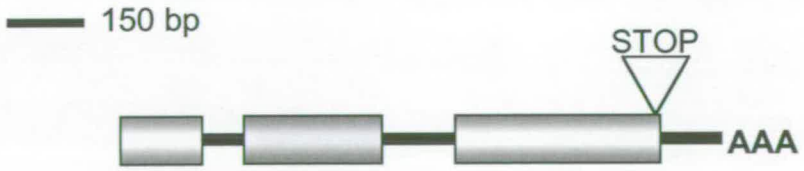


Figure 38: Diagrammatic representation of structure of *SdNAC1*. Boxes represent exons, lines represent introns, and the positions of the stop codon and polyadenylation site in a cDNA transcript are shown.

Affinity of SdNAC1 *to other NAC genes*

The homology of SdNAC1 to other NAC genes was investigated by phylogeny. Amino acid sequences were aligned using ClustalW, and regions of unambiguous alignment were included in a phylogenetic analysis (PAUP Version 3.1.1) using a heuristic search. A single most parsimonious tree was generated, which was rooted at its midpoint (figure 39). This showed that the *Streptocarpus* gene was most closely related to a tomato gene, *LeJA2*. The function of the closest homologues to *SdNAC1* is not yet published, and expression has not yet been examined for *Streptocarpus*.

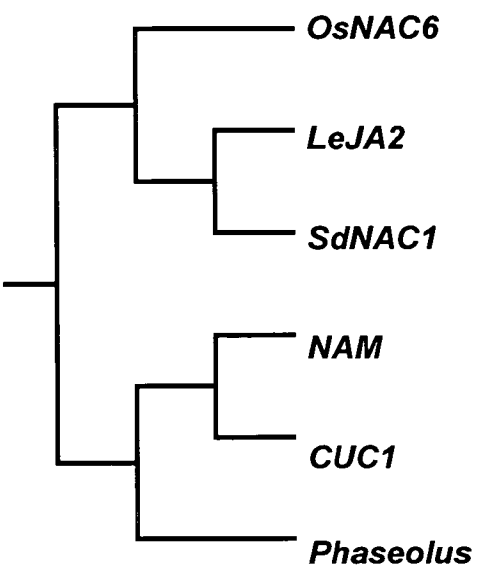


Figure 39: Phylogeny of NAC domain genes. GenBank accession numbers of sequences used are: *OsNAC6* (*Oryza sativa*: Af254558), *LeJA2* (*Lycopersicon esculentum*: AF011555), *NAM* (*Petunia* hybrid; X92205) *CUC1* (*Arabidopsis thalliana*: AB049069) *Phaseolus* (*Phaseolus lunatus* AF402603).

Transposons

A further class of sequences I isolated by chance that could be of particular interest were transposons. One sequence showed homology to the *Mutator* class of transposons in a BLAST search, and the other to a retroelement.

Record of sequence discussed

The likely homology of *Streptocarpus* cDNA and genomic DNA sequences, their GenBank accession numbers, and their potential are given in table 9, below.

Species	GenBank code	Homology by BLAST	Function in other species
S. dunnii	BM34621	NAC domain gene	boundaries between meristem compartments
S. dunnii	AY061835	NAC domain gene	as above
S. dunnii	AY061836	receptor-like kinase	meristem function
S. rexii	AY061838	*CLAVATA 1*-like gene	meristem function
S. saxorum (SSknox3)	BM34634	*KNAT3*-like gene	transcriptional regulation
S. saxorum	BM34635	MULE (*MUtator*-Like Element)	Transposons from the *Mutator* family are involved in DNA rearrangement in maize (Walbot 1992).
S. saxorum	AY061842	*CLAVATA 1*-like gene	as above
S. saxorum	AY061849	receptor-like kinase	as above
S. saxorum	AY061850	retroelement	DNA rearrangement

Table 9: Sequences isolated from *Streptocarpus.*

Record of further sequence not included in this thesis

During the course of my PhD I also isolated many other sequences, which are of potential future value. These are all included in a table in appendix 2.6. These sequences were edited to remove primer and adaptor sequences, and submitted to GenBank (Benson *et al.* 2000) or dbEST (Boguski *et al.* 1993). The likely homology of *Streptocarpus* cDNA and genomic DNA sequences, their GenBank accession numbers and their function in other species are recorded.

Summary of results

STM-like sequences were isolated from *Streptocarpus* species representing the major growth forms found in the genus. By similarity these appeared to fall into two classes that have been named *Sknox1* and *Sknox2*, and both classes are likely orthologues of *STM*. Southern blotting supported the existence of two classes of *STM*-like genes. Sequences showed high affinity to each other and to the *Antirrhinum* genes *hirz* and *ina*. Expression of *Sknox1*, as analysed by RT-PCR, differed between species. *Ssknox1*, isolated from a caulescent species, showed expression that was conventional with respect to genes falling in the *STM* clade. Expression was found in meristems, but barely in lateral organs. *Sdknox1* and *Srknox1*, from unifoliate and rosulate species respectively, showed different patterns of expression to *Ssknox1* and also to each other.

Sknox1 expression was ubiquitous in a unifoliate plant, and in a rosulate plant was found in meristems and the midrib of mature leaves. Caulescent form is plesiomorphic within *Streptocarpus*, and acaulescent forms probably evolved from a caulescent ancestor (Möller and Cronk 2001a). The evolution of novel growth forms in *Streptocarpus* therefore correlates with novel expression patterns of *Sknox1*.

Discussion

What could be causing expression of Sknox1 *in the phyllomorph?*

The case for orthology of *Sknox1* genes to *STM*-like genes in well supported by structural similarity and phylogeny. Caulescent form is plesiomorphic within *Streptocarpus*, and in a caulescent species, *Sknox1* has an expression pattern that is plesiomorphic for genes within the *STM* clade. *Ssknox1* is therefore likely to be a functional orthologue of *STM*. If *Sdknox1* and *Srknox1* genes were also orthologous to *STM* we could expect their expression in similar domains to *STM* and other *STM*-like genes including *Ssknox1*. However, in acaulescent species there are novel patterns of expression that also differ between species. It is intriguing to speculate how these may have arisen.

Changes in gene regulation?

The first possibility is that sequence changes in *cis*-acting regulatory elements of *Sknox1* genes produced novel patterns of *Sknox1* expression in acaulescent species, and therefore caused morphological differences between species. *Cis*-regulatory changes are thought to be important in generating morphological novelty, and therefore speciation (Doebley and Lukens 1998, Stern 1998, Wang *et al.* 1999, Sucena and Stern 2000).
If *cis*-regulatory changes at the *Sknox1* locus do cause morphological differences between species, we can predict that the species' trait of interest (here unifoliate or rosulate growth form) should co-segregate with the *Sknox1* allele(s) associated with that trait in crosses. If *Sknox1* expression is regulated by *cis*-acting elements but these are not responsible for morphological differences between species, each *Sknox1* allele should maintain its native expression pattern in backcross plants, regardless of the form of the plant. In *Sknox1* heterozygotes, the unifoliate *Sknox1* allele would be expressed ubiquitously, but the rosulate allele would maintain a more restricted pattern of expression.

An alternative hypothesis is that *trans*-acting elements affect the expression of *Sknox1* and are responsible for changes in growth form. By analogy to other species, several candidates for regulation of *STM*-like genes are known. If this were the case then the predictions above should be inverted. These possibilities are addressed further in the next chapter.

When and where is STM *first expressed in* Streptocarpus*?*

One explanation for unorthodox morphology in *Streptocarpus* is that meristematic activity is displaced into the cotyledon, which continues to grow. This idea is supported by the expression of *Sknox1* in phyllomorphs and also the presence of dividing cells in the cotyledon of *S. rexii*, an acaulescent species. However, it is not known whether cell division occurs in the cotyledons of caulescent species, although as there is cotyledonary accresence in caulescent species, this is likely. To test whether the meristem is established conventionally in the embryo, and when and how meristematic properties are transferred to the cotyledon, *STM* expression could be used as a marker for meristem activity as detected by *in situ* hybridization to RNA in tissue sections.

Redundancy of gene duplicates

Determining the mechanism by which *Sknox1* is expressed in the phyllomorph in unifoliate and rosulate species would be a major step in understanding how these species evolved. As *STM*-like genes are necessary for establishment and maintenance of meristems in other species, it is difficult to see how the acaulescent species tolerated recruitment of *Sknox1* into another role.

One possible explanation is that gene duplication allowed for recruitment of *Sknox1* into a new role. In *Antirrhinum* there are two homologues of STM, which correspond to the classical mutations *Hirzina* and *Invaginata* (Golz *et al.* 2002). In WT *Antirrhinum hirz* and *ina* are expressed in slightly different domains. Both are expressed in the SAM and provascular cells of internodes, but whereas *hirz* is expressed in the L1, *ina* is excluded from it. Expression of both genes is excluded from leaf initials and primordia (Golz *et al.* 2002). Jointly these two genes are most closely related to the *Sknox1* genes. My phylogeny shows that *Antirrhinum* and *Streptocarpus* are exceptional in having two *STM*-like genes: so far all other dicots have only one. Surprisingly it appears that these arose from independent gene duplications, rather than duplication in a shared common ancestor of *Antirrhinum* and *Streptocarpus*. Although *hirz* and *ina* are functionally

redundant, there may have been functional divergence between the two *Streptocarpus* homologues since they duplicated. If the second homologue is fully isolated in the future, it will be intriguing to see when and where the genes are expressed with a view to addressing this question.

Chapter 5: Effects of inheritance and expression of *Sknox1* on *Streptocarpus* morphology

Work in the previous chapter described the isolation of *STM* orthologues (*Sknox1*) from four *Streptocarpus* species representing three different growth forms. While *Sknox1* is expressed conventionally in the caulescent species, and caulescence is plesiomorphic, unorthodox lateral expression of *Sknox1* was found in representatives of rosulate and unifoliate growth forms. Although the rosulate and unifoliate species shared lateral expression of *Sknox1*, there were differences in expression between the two species surveyed. The rosulate species expressed *Sknox1* laterally in the midrib alone, whereas the unifoliate species showed delocalised ectopic expression.

To test (1) whether differences in *Sknox1* expression between unifoliate and rosulate species were caused by allelic differences at the *Sknox1* locus itself, and (2) whether *Sknox1* could have been directly implicated in the evolution of unifoliate and rosulate forms, I investigated the inheritance of *Sknox1* in the backcross populations discussed in chapter 3.

If *Sknox1* is one of the two hypothetical dominant alleles determining rosulate morphology the rosulate *S. rexii* parent should have the genotype R/R, and the *S. dunnii* or *S. wittei* parent r/r, where R is one of the dominant alleles at the early or late acting loci. The F1 *S. rexii* x *S. dunnii* hybrid should therefore have the genotype R/r, and should be rosulate, and the back cross population should consist of two different genotypes, R/r and r/r in equal proportions.

A CAPS based screen for segregation of *Sknox1* alleles

Results of a screen using intron 3 alone

To investigate segregation of *Sknox1* alleles, I developed a screen for *Sknox1* genotype based on Cleft Amplified Polymorphic Sequence (CAPS) markers. PCR products 200 bp long were initially amplified from DNA from backcross individuals using primers Fc and GR22 spanning intron 3 in the homeodomain of *Sknox1*. Sequencing the fragments revealed a restriction polymorphism between *S. rexii* and *S. wittei* for *Bcl* I. *S. rexii* products were cleft 92 bp within the sequence, whereas *S. wittei* products did not have a restriction site for *Bcl* I. As the *S. wittei* fragments had no restriction site, I needed to control for the possibility that lack of digestion reflected absence of the *Bcl* 1 site. Both

species had an *Sph* I restriction site 69 bp into the sequence, so PCR products were also cleft in separate reactions with *Sph* I.

Cleavage patterns were analysed by gel electrophoresis, and segregated in a 1:2:1 (r/r: r/R: R/R) ratio, typical of an F2 generation. This was inconsistent with the possible outcome of a backcross in which 1:1 segregation of RR:Rr genotypes is expected. A possible cause was that the F1 plants had selfed, and that the backcross plants were in fact F2 plants.

However, when I was initially trying to isolate *Sknox1* using the same primers in PCR, two bands had frequently amplified. Therefore in the screen PCRs it is likely that I was detecting more than one *STM*-like gene, and this was the more probable cause of the discrepancy. In further experiments I used one primer to a less conserved region of the gene.

Screen using coding sequence 5' of intron 3

Restriction polymorphisms further 5', and in coding sequence, were used to differentiate between the *Sknox1* genes of different species. 400 bp fragments were amplified from each of the backcross plants with primers *new rex down* and *Screen up* (figure 40). The aligned sequences of the region amplified including primer sites are shown in figure 41 A, overleaf.

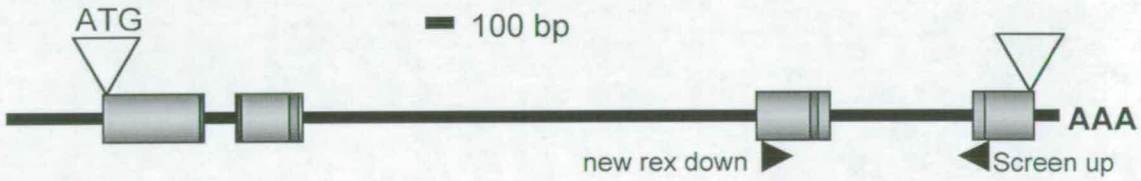


Figure 40: Diagram of *SSknox1*, showing position of primers used to amplify fragments for a CAPS screen in acaulescent species. Domains are demarcated as in figure 25.

Cleavage of PCR products with *Sau*3 AI gave fragments of 210, 165 and 35 bp in *S. dunnii*, 250, 100, 60 and 15 bp in *S. rexii*, and 250, 150 and 15 bp in *S. wittei* (figure41 B, overleaf). These showed up as clear banding pattern differences on a gel (figure 41 C). The approximately 400 bp band at the top (grey arrow) is uncleaved PCR product, and the bands beneath it are cleavage products. When cleavage was incomplete, as here, bands of additive predicted sizes were sometimes seen (white arrow).

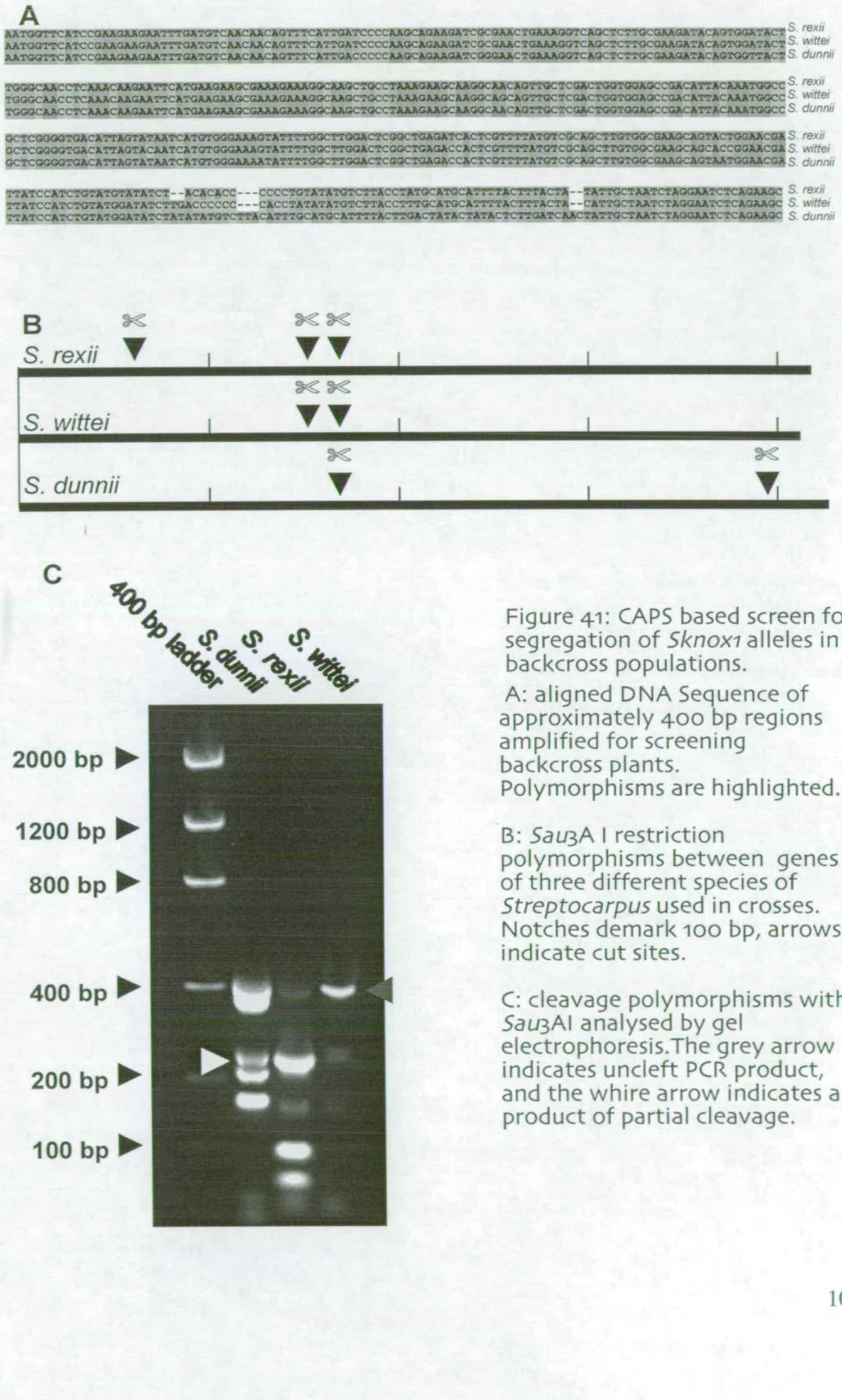


Figure 41: CAPS based screen for segregation of *Sknox1* alleles in backcross populations.

A: aligned DNA Sequence of approximately 400 bp regions amplified for screening backcross plants. Polymorphisms are highlighted.

B: *Sau*3A I restriction polymorphisms between genes of three different species of *Streptocarpus* used in crosses. Notches demark 100 bp, arrows indicate cut sites.

C: cleavage polymorphisms with *Sau*3AI analysed by gel electrophoresis.The grey arrow indicates uncleft PCR product, and the whire arrow indicates a product of partial cleavage.

Inheritance of Sknox1 *alleles was Mendelian*

Each of the 132 plants scored for morphology was scored for *Sknox1* type, and a sample of the results is shown in figure 42, below. Complete results are given in table 15 (appendix 3) and summarised in table 10 overleaf. In total for the *S. dunnii* x (*S. rexii* x *S. dunnii*) backcross there were eight homozygous (r/r) plants and (r/R) plants. In total for the *S. wittei* backcrosses there were 64 *S. wittei* homozygous (r/r) plants and 68 *S. wittei/ S. rexii* heterozygotes (r/R). The ratio of alleles segregating in both backcrosses taken together was not significantly different to the expected 1:1 ratio (*p* value = 0.121, χ^2 = between 2.7 and 3.8)

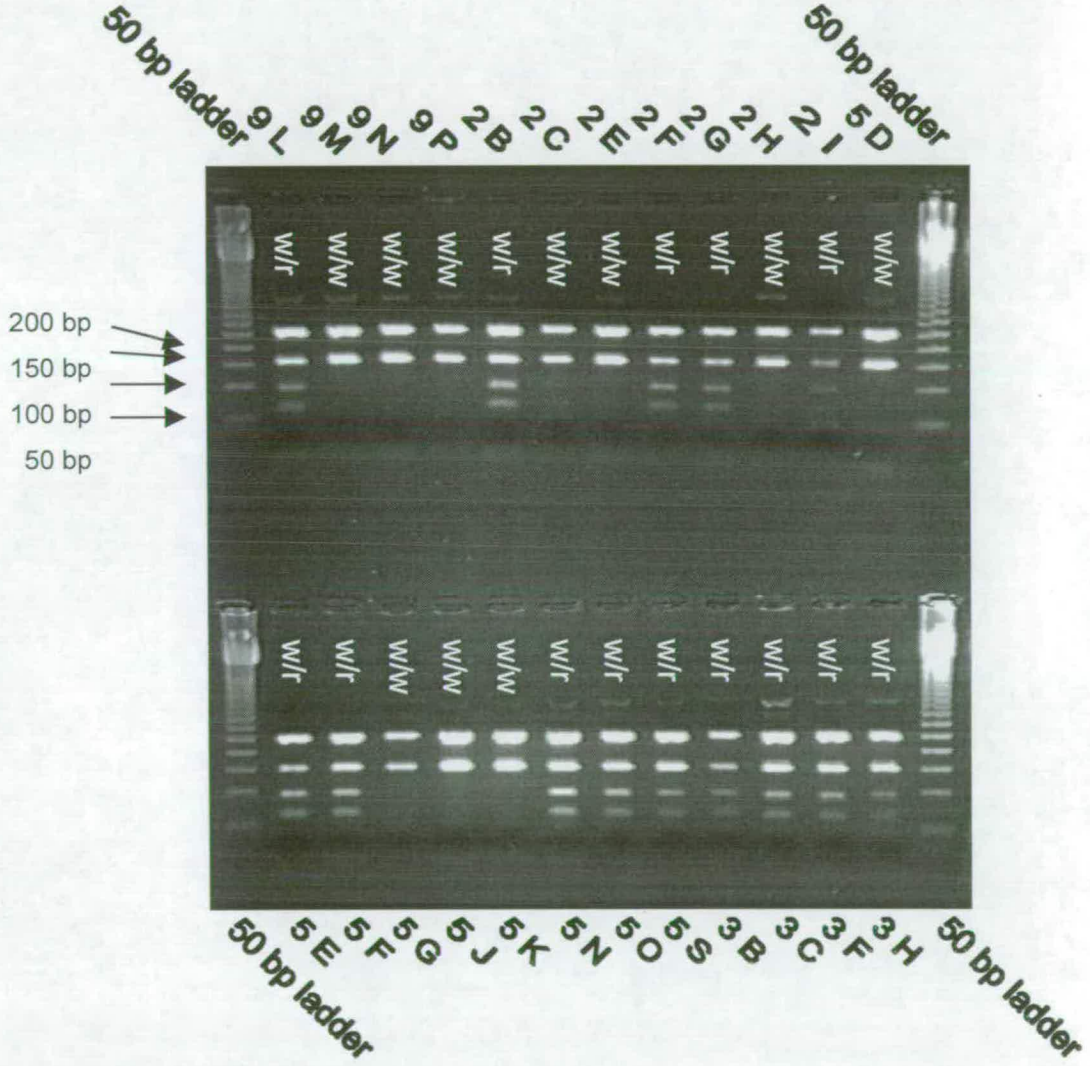


Figure 42: Inheritance of *Sknox1* alleles in *S. wittei* (*S. rexii* x *S. wittei*) backcross populations. For each individual, family number and alphabetical plant qualifier are shown. Products were separated by agarose gel electrophoresis in a 2% gel in 0.5 x TBE at 100 V for 1h30.

Family number	Number of homozygotes	Number of heterozygotes	Total
Family 1	8	8	16
Family 2	4	4	8
Family 3	5	5	10
Family 4	4	7	11
Family 5	5	5	10
Family 6	1	2	3
Family 7	5	3	8
Family 8	5	7	12
Family 9	5	8	13
Family 10	8	10	18
Family 11	7	5	12
Family 12	7	4	11
Totals	64	68	132

Table 10: Number of *Sknox1* homozygotes and heterozygotes in *Streptocarpus* backcrosses.

Correlation to morphology?

Having scored *Sknox1* type and morphological type for each of the plants in the backcrosses, I could then compare the scores to see if they correlated. A tight link between the two would indicate that genetic variation at or near to the *Sknox1* locus causes phenotypic differences between *Streptocarpus* species, and therefore that it could be implicated in their evolution. If *Sknox1* is one of the loci controlling form, we would expect the following premises to hold:

- in unifoliate plants *Sknox1* genotype should always be homozygous unifoliate recessive
- all heterozygotes should have more than one leaf at maturity.
- 50% of plants with unifoliate recessive *Sknox1* type should be unifoliate.

In fact, none of these premises did hold. Of the14 unifoliate plants in the *S. wittei* x (*S. rexii* x *S. wittei*) back crosses, 6 were homozygous (r/r) and 8 were heterozygotes (r/R). Of the 60 hetergozygotes, 7 plants did not have more than one leaf at flowering. Of the 56 *Sknox1* homozygotes, only 6 plants were unifoliate. Taken together these results

show unequivocally that *Sknox1* is not one of the major loci affecting major differences in form between species of *Streptocarpus*.

Could Sknox1 *affect other aspects of form?*

Streptocarpus leaves are generally gnarled, and sometimes heavily so, and this trait segregated in the *S. wittei* x (*S. rexii* x *S. wittei*) backcross populations in a manner that suggested control by two loci, with flat behaving dominantly over gnarled.

I investigated whether this trait could be linked to the *Sknox1* alleles inherited in the *S. wittei* backcrosses. Fewer plants survived at the time of scoring than when gross morphology was scored, so it was first necessary to verify that the subset of plants surviving had an even distribution of *Sknox1* alleles. 38 plants remaining were heterozygotes and 35 were homozygotes, and this distribution approximates to a 1:1 ratio (p value = 0.123, χ^2 between 2.7 and 3.8). As for the inheritance of gross morphology, in the case of *Sknox1* alleles corresponding to the gnarly leaf trait, one would expect that the recessive phenotype is always associated with the recessive alleles (r/r). One could also expect that all flat-leafed plants should carry at least one dominant (R) allele, and that 50% of r/r homozygous recessive plants should be gnarled. As in the case of gross morphology, none of these predictions held. Five plants showed extreme gnarling, and of these only one was r/r, seven plants had very flat leaves, and of these only two were *Sknox1* heterozygotes, and of the r/r *Sknox1* homozygotes, only five were gnarled. This indicates that *Sknox1* is not one of the two loci that cause the leaf gnarling trait.

A CAPS based screen for inheritance of SNAC1

As well as developing and using a screen for *Sknox1* segregation in the backcross populations, I also developed a screen for segregation of *SNAC1* genes. An approximately 230 bp fragment was amplified, cloned and sequenced from *S. dunnii*, *S. rexii* and *S. wittei* using primers CUCP1 and CUCP2 (figure 43).

The sequences showed restriction polymorphisms which could be used to differentiate between species' *SNAC1* alleles. The full sequence of these fragments including primer sequences is shown in figure 44 A, and the position of restriction sites which could be used in a screen is shown in figure 44 B. For *S. dunnii* the expected fragment sizes are 98, 79 and 60 bp, for *S. rexii* they are 83, 61, 42, 36 and 15 bp, and for *S. wittei* they are 83, 78, 61 and 15 bp. Due to time constraints this screen has not yet been tested.

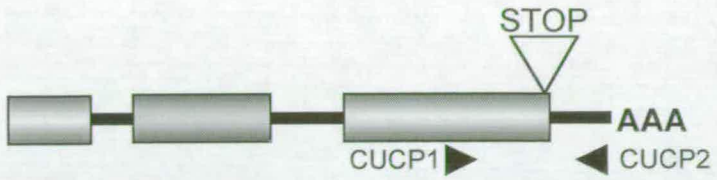


Figure 43: Diagrammatic representation of *SdNAC1* showing position of primers CUCP1 and CUCP2, used to amplify *S. dunnii*, *S. rexii* and *S. wittei* sequences for CAPS screening.

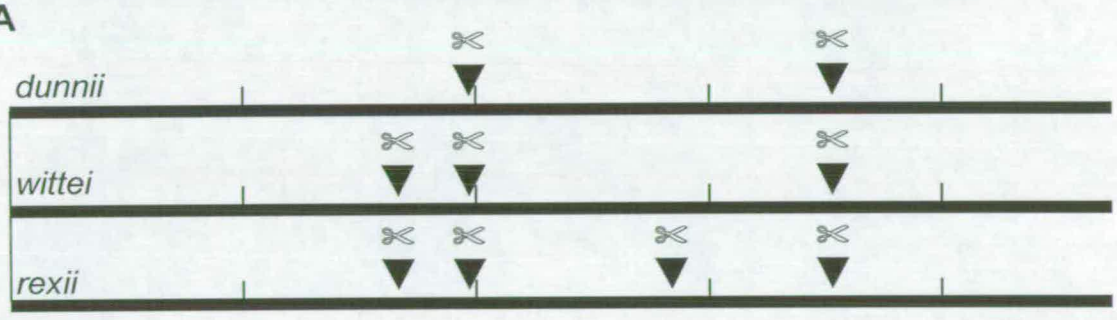
```
GCGTTTTGGAGTCGCTGGAACGGATCGATGACCAAGGCTACGCCAATTTGTCGAGAATGG  S. dunnii
GCGTTTTGGAGTCGCTGGAACAGATCGATGACCAAGGCTACGCCAATTTGTCGAGAATGG  S. wittei
GCGTTTTGGAGTCGCTGGAACGGATCGATGACCAAGGCTTCGCCAATTTGTCGAGAATGG  S. rexii

GCTCTCTGAAGCTGGACCTTCAGAACTTGGGCTCTGGGAATTTCGACTGGGCCACGTTAG  S. dunnii
GCTCTCTGAAGCTGGACCTACAGAATTTGGGCTCTGGAAATTTCGACTGGGCCACGTTAG  S. wittei
GCTCTCTGAAGCTGGACCTACAGAATTTGGGCTCTGGGAATTTCGACTGGGCCACGTTAG  S. rexii

CTGGGCTTAACCCTCTGCCCGAACTTATGAAGGCCCAACATAACCAGGCGAATGGGAAAT  S. dunnii
CTGGGCTTAACCCTCTGCCCGAACTTATGAAGGCCCAACATAACCAGGCGAATGGGAAAT  S. wittei
CTGGGCTTAACCCTCTGCCCGAATTTATGAAGGCCCAAAATAACCAGGCGAATGGGAAAT  S. rexii

TCTGATATGACACTTATGCCCTTGTCTTCCAGCATGCTGGGCCATTCGATATGGATG     S. dunnii
TCTGATATGACACTTATGCCCTTGTCTTCCAGCATGCTGGGCCATTCGATATGGATG     S. wittei
TCTGAAATGACACTTATGCCCTTGTCTTCCAGCATGCTGGGCCATTCGATATGGATG     S. rexii
```

A



B

Figure 44: Basis of a CAPS based screen for the segregation of *SNAC1* alleles in backcross populations. A: sequence of 23 bp *SNAC1* PCR products from *S. dunnii*, *S. rexii* and *S. wittei*, with polymorphisms highlighted.
B: *Apo* I restriction map of *SNAC1* fragments. Notches demark 50 bp intervals.

Other genes could affect morphology

As explained previously, other genes as well as *STM* affect meristem behaviour and plant morphology in other species. Therefore these genes are also candidates for morphological change in *Streptocarpus*. Partial homologues of some of these have been isolated from *Streptocarpus* (chapter 4), and the backcross plants could also be used to test if they are involved in segregation of morphology.

Summary of results

CAPS markers were developed to screen backcross populations for inheritance of *Sknox1* and *SNAC1* alleles. Plants were screened for segregation of *Sknox1* alleles, which was Mendelian. Gross morphology and leaf gnarling, two morphological traits segregating in backcross plants, were not linked to allelic differences at the *Sknox1* locus. This rules out the possibility that *cis*-regulatory changes in *Sknox1* are directly implicated in morphological evolution between the unifoliate and rosulate species studied. However, differences in expression pattern between *Sknox1* alleles were found in parent species, implying that a gene at a *trans*-acting locus might be implicated.

Discussion

Could Sknox1 *genes be regulated by* cis-*acting elements?*

Although *cis*-regulatory changes at the *Sknox1* locus cannot be implicated in morphological change between *S. dunnii* and *S. rexii*, the backcross plants offer a valuable opportunity to test whether *Sknox1* expression is regulated by *cis*-acting elements. In *S. dunnii Sknox1* was expressed ubiquitously, whereas in *S. rexii* expression in the phyllomorph was confined to the midrib. If *Sknox1* gene expression is regulated by *cis*-acting elements we would expect that each allele retain its native expression pattern in heterozygous backcross plants regardless of the form of the plant. The *S. dunnii* allele should still be expressed ubiquitously, and the *S. rexii* allele should still be restricted to the midrib in the phyllomorph. This could be tested by RT-PCR with samples of midrib and laminar tissue for each plant using primers MEGSD and ELKUP as in chapter 4. In Southern blotting, a 5' *Sknox1* probe detected fragments from all species, so to differentiate between alleles, they would have to be screened another way. There is a *Sca* I restriction polymorphism between *S. dunnii* and *S. rexii* in the region that would be amplified. *S. rexii* products should cleave into 386, 273 and 66 bp fragments, and *S. dunnii* products should cleave into 452 and 273 bp fragments. If *Srknox1* expression was found in the lamina, *Sknox1* expression must be regulated by *trans*-acting elements.

Major differences in growth form between species

Changes at the *Sknox2* locus?

This leads to the question if *Sknox1* genes do not cause growth form differences, then what does? As there is probably a second *STM*-like gene in *Streptocarpus*, one possibility is that this rather than the locus that I isolated is a major locus affecting form.

It would be possible to test this following isolation of a complete homologue. I have isolated sufficient sequence of a second potential *STM* orthologue (*Sknox2*) to design primers for this, but my attempts to amplify fuller sequence were frustrated.

Trans-acting elements?

The second major possibility is that *Sknox1* alleles are regulated by a *trans*-acting gene, and therefore this could be a major locus affecting morphology. If the differences in expression pattern between *S. dunnii* and *S. rexii* were brought about by a molecular transition between species, we need to know the direction of the switch before we can say whether it must have involved an up- or down-regulation of *Sknox1*. If ubiquitous expression of *Sknox1* is the basal character, then to restrict its expression the trans-acting gene is likely to be a repressor or loss of activator. If the more restricted pattern is basal, then the *trans*-acting gene is likely to be an activator or a derepressor.

Although the unifoliate character is plesiomorphic in the clade encompassing *S. rexii* (Möller and Cronk 2001a), the unifoliate species I sampled (*S. dunnii*) is in a separate clade to *S. rexii*. This means that the ubiquitous *Sknox1* expression pattern seen in *S. dunnii* may not be the plesiomorphic character state. From the results available we cannot know the direction of the switch.

By analogy to other species, there are some candidates, but these partly depend on whether the phyllomorph is viewed as an embryonic or a post-embryonic structure.

The meristem may not form properly in embryogenesis

In unifoliate species, the phyllomorph is a cotyledon that has continued to grow. Cotyledonary phyllomorphs show several meristem like properties. They have a capacity for continued growth, they have regions of dividing cells, and they sometimes iterate leafy organs along the midrib. They also express *Sknox1* genes. Because the cotyledon is an embryonic structure this suggests incorrect establishment of the SAM in the embryo.

In *Arabidopsis* embryos redundant expression of two genes, *CUC1* and *CUC2*, is required for the establishment of *STM* expression and establishment of the meristem (Aida *et al.* 1999). Expression of *CUC1* driven by the 35S promoter sometimes leads to the formation of meristems above veins in cotyledons, and cotyledons in 35S::*CUC1* plants express *STM* (Takada *et al.* 2001). In *Arabidopsis*, *STM* is not required for meristem formation in absence of *AS1* (Byrne *et al.* 2000). Given a situation where both *AS1* and *STM* homologues are present, but the boundaries between them are not

established correctly, one can envisage that the expression domains of genes expressed in the embryo overlap inappropriately. As *knox* genes repress expression of ARP genes, this would lead to loss of organ-like features, and acquisition of stem-like features in the cotyledon. Failure to establish a meristem, and acquisition of meristem properties in the phyllomorph, in acaulescent species could be therefore be explained by a loss of function of *CUC/ NAM*-like genes.

The transition from embryonic to post-embryonic growth may be perturbed

A second hypothesis is that in acaulescent species the vegetative phase of development, typified by iteration of leaves from a SAM, is skipped. In seedlings of acaulescent *Streptocarpus* the plumule does not develop, and the primary phyllomorphs have an embryonic origin. Inflorescence production is normal, and supplied by the groove meristem, which is large compared to meristems in caulescent *Streptocarpus* (Harrison *et al.* 1999). In unifoliate species where there is only one phyllomorph, this phenotype closely resembles *EMF* mutants of *Arabidopsis*. In *emf* mutants, the SAM initiates as an inflorescence meristem. The meristem is larger then vegetative SAMs from the onset, and it never iterates rosette leaves which characterise vegetative growth in *Arabidopsis*. Interestingly, as in *Streptocarpus*, the cotyledons are sessile in *emf* mutants. In contrast to *Streptocarpus*, floral development is also perturbed in *emf* mutants, and seed set was not observed (Sung *et al.* 1992, Yang *et al.* 1995).

The meristem may not be maintained correctly

As yet there are no known regulators of *STM* expression in post embryonic development in model species, but regulatory mechanisms are known for other *knox* genes, and these involve the ARP genes. In *Arabidopsis*, *STM* negatively regulates *AS1* in stem cells. *STM* expression is down regulated in organs and their primordia, and this is permissive to *AS1* expression. *AS1* is represses expression of *KNAT1* and *KNAT2* in organs and their primordia. In the absence of *AS1* and *STM* in *Arabidopsis*, a relatively normal shoot forms, suggesting that redundant pathways regulate the meristem (Byrne *et al.* 2000). It is possible that *Sknox1* is regulated by an ARP homologue, which behaves differently in different species of *Streptocarpus*, inducing different patterns of *Sknox1* expression.

Differences in degree of leaf gnarling between species

In other species, a gnarled leaf phenotype is associated with mis-expression of *KNAT* and *KNOTTED*-like genes in the leaf. This may be accompanied by reduced apical

dominance, indistinction between proximal and distal identity, and ectopic meristem formation, all of which are properties shown by phyllomorphic *Streptocarpus*. As the switch between rosulate and unifoliate forms correlates with novel patterns of expression of *Sknox1* in the phyllomorph, but inheritance of *Sknox1* does not cause any obvious phenotype, the gnarled leaf phenotype could result from *KNAT*-like gene expression in the leaf. In *Arabidopsis STM* represses *AS1* expression, and *AS1* directly represses *KNAT1* and *KNAT2* expression in lateral organs. If *Sknox1* genes repress *Streptocarpus* homologues of ARP genes, and these also repress expression of *Streptocarpus KNAT*-like genes, it is possible that in *S. dunnii Sknox1* excludes ARP gene expression from the phyllomorph and therefore permits expression of *Streptocarpus KNAT*-like genes which cause gnarling. In *S. rexii Sknox1* expression is restricted to the midvein, and the phyllomorphs are flat. In the model postulated above lack of *Sknox1* expression in the lamina would permit ARP gene expression, which would repress *KNAT*-like gene expression, and decrease gnarling. As allelic differences in *Sknox1* do not link to the leaf gnarling trait, both of these scenarios again implicate a *trans*-acting regulator of *Sknox1* expression in determining gnarling.

In other species, traits associated with *knox* mis-expression in leaves are also strongly enhanced by mutations at other loci, such as *SERRATE* and *PICKLE*. Neither *SERRATE* nor *PICKLE* directly affect *knox* expression, rather they appear to affect competence of cells to respond to knox expression (Tsiantis 2001). Some genes are also known to buffer genotypic variation and plant phenotype and these may also be important in regulating plant form.

Chapter 6: Phylogenetic analysis and evolution of *Sknox* genes

By Southern blotting and PCR, I showed in chapter 4 I that there are probably two *STM* homologues in *Streptocarpus*. My first step in isolating *Sknox1* genes from *Streptocarpus* was to amplify across an intron (intron 3) in the homeodomain using primers Fc and GR22 in PCR (chapter 4 figure 17). I frequently obtained two or three products, and their length was variable within and between species, consistent with the presence of more than one gene. To further investigate the relationships between possible *Sknox* homologues within *Streptocarpus*, and to see whether these could have utility in resolving as yet unresolved nodes in phylogeny, I amplified *Sknox* introns from multiple species of *Streptocarpus*. Part of this work comprised an investigation into repeating refrains in longer introns, some of which was carried out by Nicola Wright, an honours student, under my supervision.

Phylogenetic analysis of intron 3 sequences

Taxon sampling

The first part of the work described in this chapter investigated the phylogeny of intron sequences obtained from *Sknox* genes. A well resolved species phylogeny of *Streptocarpus* based on ITS sequence data was recently published (Möller and Cronk 2001a), and an abridged version of this phylogeny was used for direct comparison to my results (figure 45, overleaf). Taxa were sampled from each of the subgenera of *Streptocarpus*, from the different regions inhabited by each subgenus, and from different growth forms. Other taxa not included in Möller and Cronk's 2001a phylogeny were also sampled, and these were *S. levis*, *S. venosus*, *Saintpaulia tongwensis* and *Sa. velutina*. Previous work has shown that *Saintpaulia* forms an unresolved monophyletic group nested within *Streptocarpus* (Möller and Cronk 1997a, 2001a). *Streptocarpus levis* and *S. venosus* are both Madagascan caulescent species that strongly resemble Madagascan species included in Subgenus *Streptocarpella* (CLADE 1), which are unified by a chromosome number of 2n=30. Their position has yet to be confirmed by karyotyping.

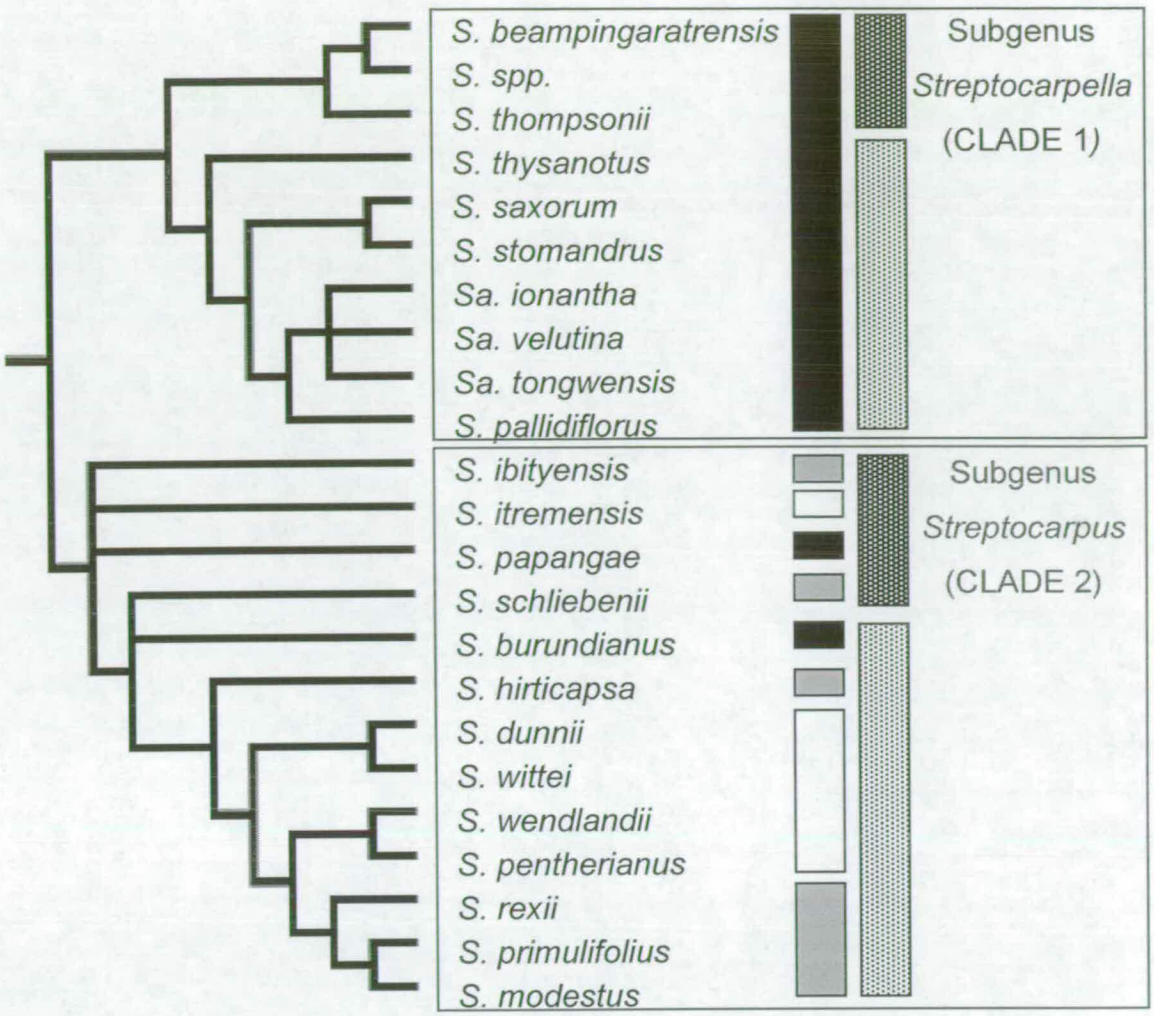


Figure 45: Abridged ITS phylogeny redrawn from Möller and Cronk (1997a, 1997b and 2001a), showing the relationships between species used in this study. Shading in boxes indicates growth from and origin of taxa. Black fill indicates caulescent species, grey fill indicates rosulate species, no fill indicates unifoliate species, white dots on black background indicates Madagascan species, black dots indicate African species.

PCR and sequencing results

Introns are non-coding sequences of genes which interrupt coding sequence and are spliced out of precursor messenger RNA by a protein complex called the spliceosome (Brown *et al.* 1996). Introns have sequences and secondary structure that are necessary for correct splicing, and they therefore show conserved features. In plant introns three of these are particularly important: (1) The introns / exon boundary sequences which in over 90% of plant introns are GT and AG at the start and end respectively, (2) AT richness (3) branchpoint sequences (consensus YTNAN) that are involved in secondary

structure and lariat formation. Although frequently present in plant introns, branchpoint sequences are not essential for splicing (Goodall and Filipowicz 1989, Goodall and Filipowicz 1991, Brown *et al.* 1996). If all intron sequences amplified from *Sknox* genes are functional, it is reasonable to expect that they show at least the first two of these conserved features.

Initially PCR products were amplified using primers Fc and GR22 as described in chapter 4. Most species produced a band of ~200 bp, and in others a second band of ~450 bp was seen (figure 46). Bands were isolated from gels, purified and sequenced directly using Fc and GR22. For several species this gave clear sequence with few, if any, polymorphic sites, and it is therefore likely that these products resulted from amplification of a single sequence.

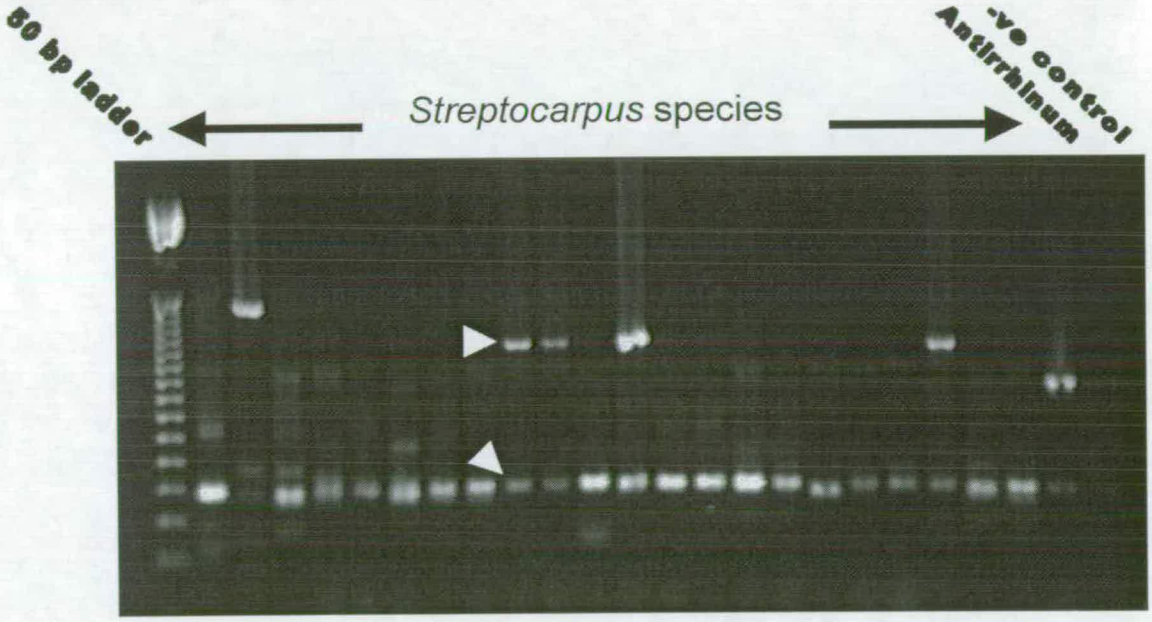


Figure 46: Example of product migration in a minigel to illustrate PCR amplification of multiple products from some species of *Streptocarpus*. Arrows indicate long (top) and short (bottom) products amplified from a single species.

The sequence obtained from the ~ 200 bp product from other species was unreadable due to the presence of multiple peaks at each position. Results presented in chapter 4 showed that *Sknox* genes are probably duplicated in *Streptocarpus*, and it was therefore likely that 200 bp bands contained products of similar length from more than one gene. Products were therefore purified and cloned products into pGEM-T, or pGEM-T EASY vectors. Clones of short products were sequenced in one direction using a vector specific primer, T7, which tended to give the most high quality sequence. Clones of longer

products were sequenced in both directions. Vector and primer sequences were removed and sequences compared and partially aligned using ClustalX in Sequence Navigator. All sequences are given in appendix 4.

Intron 3 sequences from previously isolated *Sknox1* described in chapter 4 were also included in alignments. The length of intron 3 varied from 75- 657 base pairs in different species. Because there was considerable variation in length and composition of the introns, automated alignments gave variable and unsatisfactory outputs. On examining intron sequences, it became apparent that they fell into several clusters, on the basis of conserved motifs within the intron as well as length. I therefore aligned sequences manually in PAUP.

Conserved motifs in intron 3 and sequence alignment

All of the sequences amplified contained canonical intron: exon splice sites (:GT-AG:) except *S. pentherianus* where the 3' splice acceptor site was AC: (figure 49, page 125). AT content is typically above 60% in plant introns, and all of the *Streptocarpus* sequences had high AT content, varying from 61% in *S. papangae* to 70 % in *S. dunnii*. Most of the sequences had two potential branch point sites, indicated by similarity to the canonical consensus branchpoint sequence for plants: YTNAN (figure 49). The first of these had a consensus sequence HTKRW. Although this deviates from the canonical consensus, only two of the sequences with this potential branchpoint did not have a T at branchpoint position 1; one sequence had an A, and the other had a C. At position 3, only one sequence had a T rather than a G, at position 4 one sequence had a G rather than an A, and at position 5 only two sequences had an A rather than a T. The consensus sequence of the second potential branchpoint was YTAWT. In the fourth position about half of the sequences had a U rather than an A. Some sequences had only one of the two potential branchpoints: in *S. wittei*, *S. rexii* and *S. hirticapsa* the first branchpoint sequence was absent, and in *S. ibityensis* the second was absent. All sequences therefore preserved typical characteristics of plant introns that are required for correct splicing, which is consistent with their role as functional genes.

Alignment

As well as these conserved domains that are important for correct intron function in model species, all sequences had a GCAT or a similar modified refrain (consensus KSVW), which was used to anchor sequence alignments. Most sequences also contained

a CATTT or similar motif (consensus CMTHT) that in some species was duplicated, and these were used to further anchor sequence alignments (figure 49). About half of the sequences contained a poly(C) tract at the 5' end of the intron, that was absent in other sequences, and this feature defined two classes of intron and also aided alignment.

Rooting

As *knox* genes comprise a large gene family, and in *Antirrhinum STM*-like genes are recently duplicated, it seemed likely that I had isolated intron sequences from at least two different *Sknox* paralogues in *Streptocarpus*. In phylogenetic studies that use duplicated genes to construct phylogenies, multiple genes are isolated from a single genome. Therefore it is reasonable to assume that, if each copy has remained functional, they have undergone the same selective pressures (Brown and Doolittle 1995, Telford and Holland 1997). In this instance, rooting of trees on a sequence from an out-group taxon is likely to be inappropriate as (1) homology of out-group sequences to in-group sequences may not be clear (2) where the root of the tree is not known, outgroup rooting may mask the true root of the tree (3) sequences within the in-group are most similar each other, and therefore duplicate clades provide the most appropriate root for each other.

Because recent *Streptocarpus* phylogenies raised some debate about the origin of *Streptocarpus*, and the basal relationships between subgenera (Möller and Cronk 2001a, b) and it was likely that I had isolated duplicate genes, I rooted trees at their midpoint.

Phylogenetic analysis

Due to the difficulty in alignment, I edited my data set in different ways including regions that aligned with varying robustness. I analysed two data sets, one with the coding region from all sequences (40 taxa in all: figure 48, page 122) and the second using coding sequence and unequivocally aligned intron positions (figure 50). The second data set was reduced due to the length disparity between some of the sequences which meant that they could not be aligned unambigously. *S. thysanotus*, *S. saxorum* and *S. pallidiflorus* sequences were excluded. Both of these alignments, and the complete data set are included in Nexus format on the CD attached to this thesis. Data were analysed by the PAUP program (Version 3.1.1) using heuristic searches. After initial analysis, data was re-analysed using the 'reweight characters' option. Consensus trees were computed for each analysis and are described here.

Using coding sequence only

In the coding sequence there were eight informative positions (figure 48, page 122). I obtained 1598 most parsimonious trees, and computed their majority rule consensus (figure 49A, page 123). This resolved two main clades, one largely unresolved, and a second nested within it. Three clades were preserved in the strict consensus, and are indicated by arrows in figure 49. One of these was a clade comprising a group of sequences from African acaulescent species that have always been associated in previous phylogenies. The second comprised two sequences (from *S. rexii* and *S. burundianus*) nested within that clade. Finally, two sequences from one species (*S. ibityensis*) clustered, and these were not differentiated in the coding region. It is therefore likely that these were allelic variants. Four further clades were supported, and these are indicated by numbered bars to the right of the trees. The first of these comprised short intron variants from *S. stomandrus*, *S. pallidiflorus* and *S. saxorum*, which are closely related caulescent species that fell together in previous phylogenies. The second and third are comprised of sequences from *Sa. tongwensis*, *S. dunnii*, *S. modestus* and *S. thompsonii*, and from *Sa. tongwensis*, *Sa. ionantha*, *Sa. velutina* and *S. dunnii* respectively. Neither of these groups reflect species relationships, although the latter would if *S. dunnii* were excluded. The fourth group is a couplet including sequences from *S. rexii* and *S. hirticapsa*, which are both African acaulescent species. When characters were re-weighted using the rescaled consistency index (mean value), 292 most parsimonious trees were obtained, further groupings were found in the majority rule consensus (figure 49B). Two main clades formed, but these were separate rather than nested. Each of these encompassed groups consisting of African acaulescent species whose sister taxa were Madagascan caulescent species with 2n=32. These data therefore support previous hypotheses that Madagascan caulescent species are plesiomorphic with respect to subgenus *Streptocarpus*. Sequences from species in both subgenera are included in both clades, which is suggestive of a gene duplication prior to the subgeneric split of *Streptocarpus*. However, groups 2 and 3 are both present within one of the two main clades, which is not congruent with this hypothesis. Monophyletic groups supported by the unweighted analysis were preserved. As there were few informative positions in the coding sequence, support for the clades shown in figure 49 is likely to be weak, so tree statistics were not used.

Using coding sequence and conserved positions within the intron

When data from the introns were included in the data set, there were 29 informative positions (figure 50, page 124). 5200 trees were found using a heuristic search strategy, and these saturated the computer's memory. Strict and majority rule consensus tree were computed, and the majority rule tree is shown in figure 51A (page 125). Characters were re-weighted using the rescaled consistency index, and the data was reanalysed. The majority rule consensus tree is shown in figure 51B.

In both analyses, two main clades were supported, and within each of these two further clades were supported. Clade 1 consisted mainly of sequences from caulescent species corresponding to CLADE 1 (subgenus *Streptocarpella*) from Möller and Cronk's (1997, 2001a) previous phylogenies. As *S. dunnii* and *S. modestus* (underlined) are African acaulescent species, it is odd that they appear in a clade mainly containing African caulescent species, and this placement conflicts with previous phylogenies. One likely explanation for this incongruence was DNA contamination, and this could be verified by replicating the experiment. Sequences from two species (*Sa. tongwensis* and *S. dunnii*) were present in both sub-clades, 1A and 1B, which is suggestive of gene duplication. Like Clade 1, Clade 2 is split into two sublclasses, 2A and 2B. The basal branches of each of these are mainly occupied by CLADE 2 Madagascan species (Möller and Cronk 1997, 2001a). CLADE 2 African acaulescent species are nested within 2A and 2B. Both 2A and 2B therefore share the topological trend shown by ITS phylogeny, and offer further support for the hypothesis that CLADE 2 Madagascan caulescent species are plesiomorphic with respect to subgeneus *Streptocarpus*. Although the *Sknox* introns are mainly short, they are still able to resolve species relationships within *Streptocarpus* to a level comparable to ITS. In taxa with longer introns, such as *Antirrhinum*, they may therefore offer substantial phylogenetic utility.

The structure of groups within Clade 2 is suggestive of gene duplication. Many of the species represented in 2A are also represented in 2B, again supporting the hypothesis of gene duplication. Few taxa are represented more than twice: *S. ibityensis* is represented three times, and in the sequence included in the analysis both sequences are identical, supporting the view that they are alleles. The two sequences are differentiated by a difference in the number of C nucleotides in the poly-C tract near to the 5' end of the intron (figure 56, appendix 4.2). *S. modestus* is also represented three times, and two of the sequences are differentiated by an 8 bp insertion/ deletion, consistent with allelic

variation (figure 56). Representation of *S. dunnii* in triplicate was likely to be erroneous, as two of the sequences were phylogenetically incongruous, and might therefore represent DNA contamination. Results from Southern blotting detected two bands with homology to *Sknox1* at high stringency, and results from 3' RACE experiments also detected two products from *S. dunnii*. Therefore it seems likely that each species of *Streptocarpus* has two *Sknox* homologues. Results presented in chapter 4 suggested that the second *Sdknox2* gene resulted from a duplication that happened after the divergence between *Antirrhinum* and *Streptocarpus*, but it was unclear whether this happened before or after the subgeneric split within *Streptocarpus*. In a duplicate gene phylogeny, where gene duplication preceded taxonomic divergence, we would expect that within each clade species relationships are maintained. If a gene duplication occurred after taxonomic divergence, we would expect the converse. These hypotheses, and predictions as to the outcome for *Streptocarpus* phylogeny are summarised in figure 47.

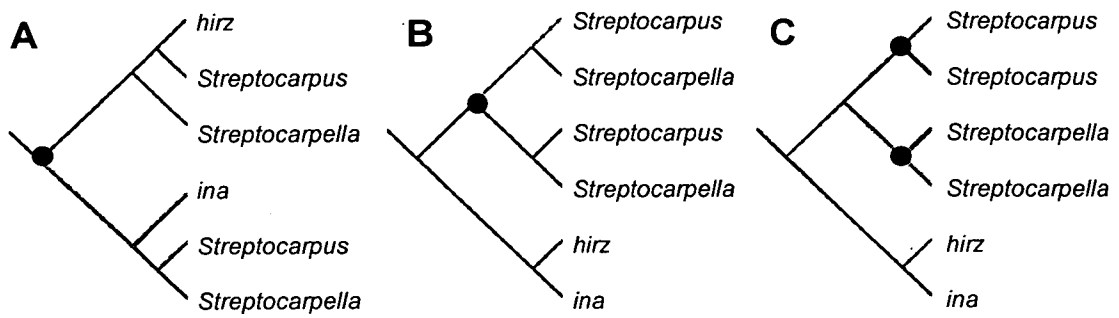


Figure 47: Phylogenetic hypotheses relating to *Sknox* gene duplication. Duplications are indicated by a black dot. A: effect of a gene duplication preceding the split between *Antirrhinum* and *Streptocarpus*. Taxonomic relationships are maintained for each gene duplicate, represented by grey or black lines. B: effect of duplication occurring after the divergence of *Antirrhinum* and *Streptocarpus*, but preceding or coinciding with the subgeneric divergence of *Streptocarpus*. Taxonomic relationships within subgenera of *Streptocarpus* are maintained for each duplicate. C: effect of duplications occurring after the subgeneric split of *Streptocarpus*. Taxonomic relationships within each subgenus should be maintained.

Although species relationships are reflected in the two main weakly supported clades of the coding sequence phylogeny, this is not the case for the phylogenies using intron data. In these, the main split between clades does not duplicate a species tree, but each sub-clade reflects species relationships within the subgenera of *Streptocarpus*. There are two hypotheses as to how this pattern may have arisen. The first is that there have been two separate duplications of *Sknox* genes that happened after the subgeneric split of

Streptocarpus. The second is that there has been concerted evolution of intron 3 sequences that has lead to their homogenisation within each subgenus. With regards to duplication, the most parsimonious hypothesis is that there was one duplication that happened before the subgeneric split of *Streptocarpus*. Some circumstantial evidence that would support this hypothesis is that other developmental genes also appear to have been recently duplicated in *Streptocarpus* (Citerne *et al.* 2000).

Concerted evolution is thought to arise by two mechanisms of recombination between chromatids. The first of these is gene conversion, by which an allele at one locus replaces its sister allele by homologous recombination. The second mechanism is unequal crossing over, which is similar to conventional recombination, but typically occurs in repetitive tracts of DNA, and results in one daughter chromatid acquiring repeats and the other losing them. Whilst this would not provide a mechanism by which sequences carrying the shorther intron could have homogenised, it is a potential mechanism for the evolution of the longer introns in which there are repeats.

It is also worth noting that homogenisation of intron sequences by concerted evolution does not necessarily couple with a similar pattern in coding sequence. This is illustrated by the evolution of the primate eye pigment genes, in which exon sequences from each pigment can be used to resolve a species phylogeny, but intron sequences are homogeneous within pigments (Shyue *et al.* 1994, Zhou and Li 1996).

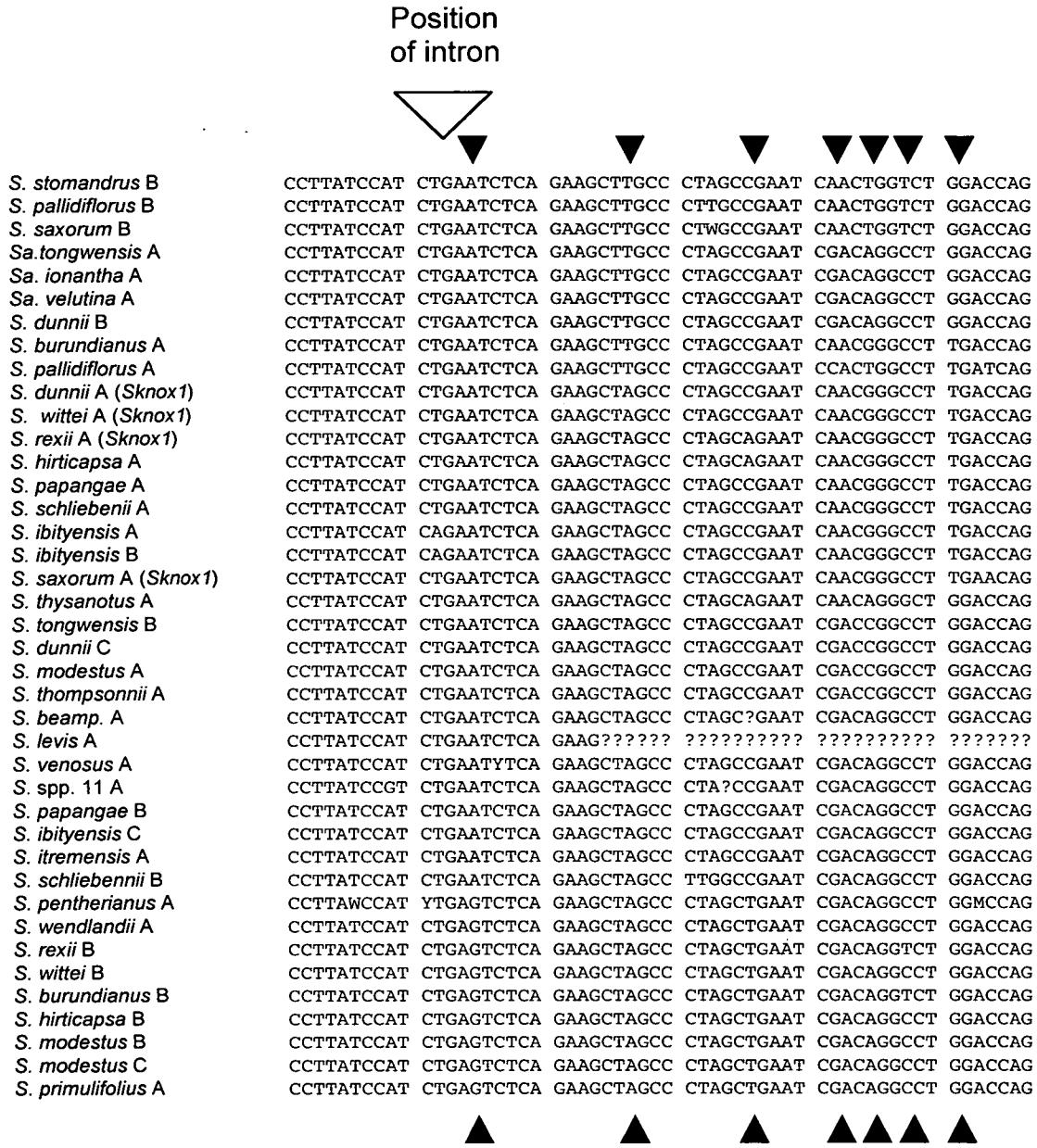


Figure 48: Alignment of coding sequences amplified from *Sknox* genes. Informative positions are indicated by filled arrows.

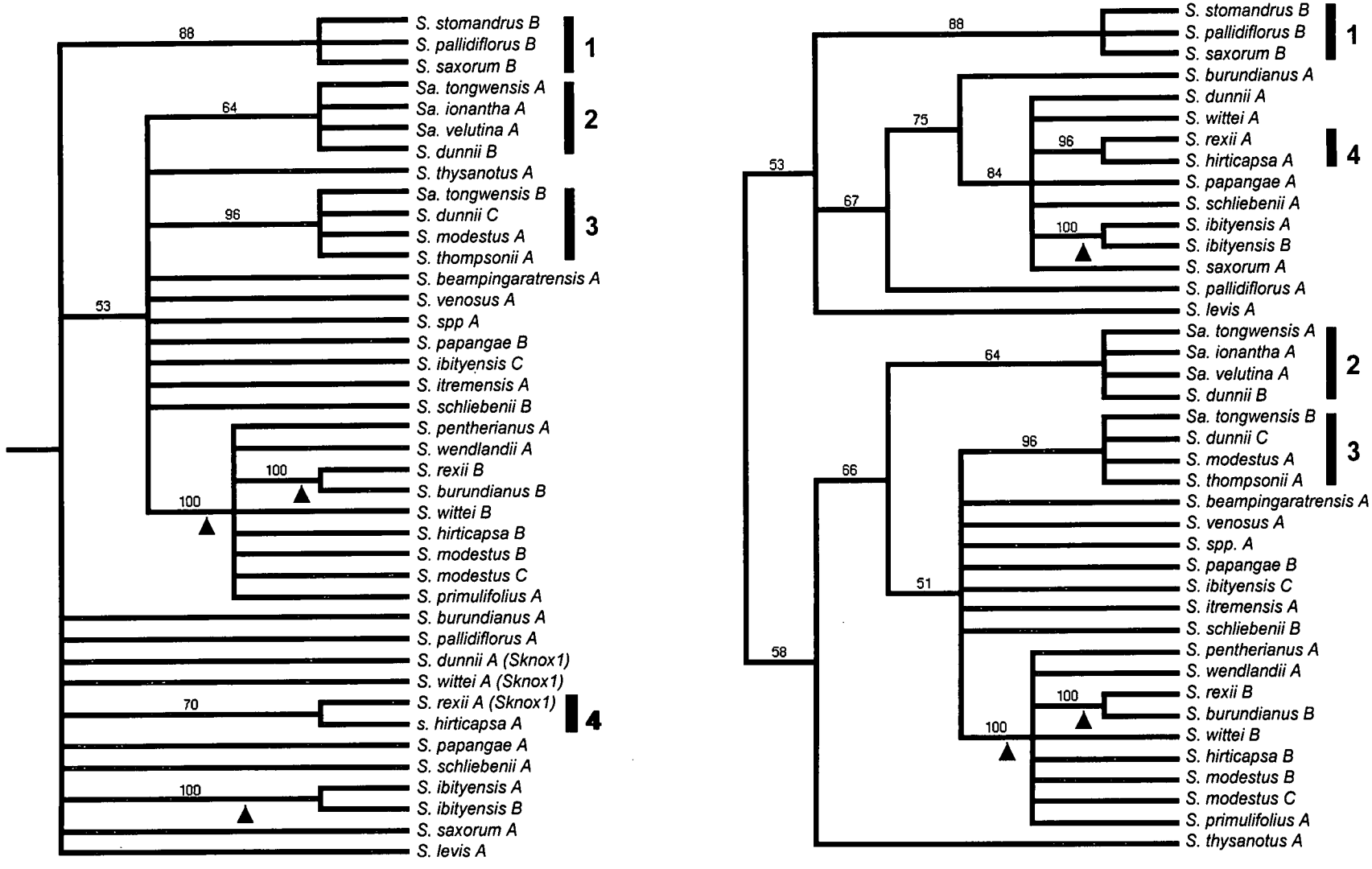


Figure 49: Phylogenetic analyses using data from coding region only. A: majority rule consensus of 1598 most parsimonious trees. B: majority rule consensus of 292 most parsimonious trees when characters were reweighted. Arrows indicate clades preserved in all most parsimonious trees, bars indicate clades found in >50% most parsimonious trees.

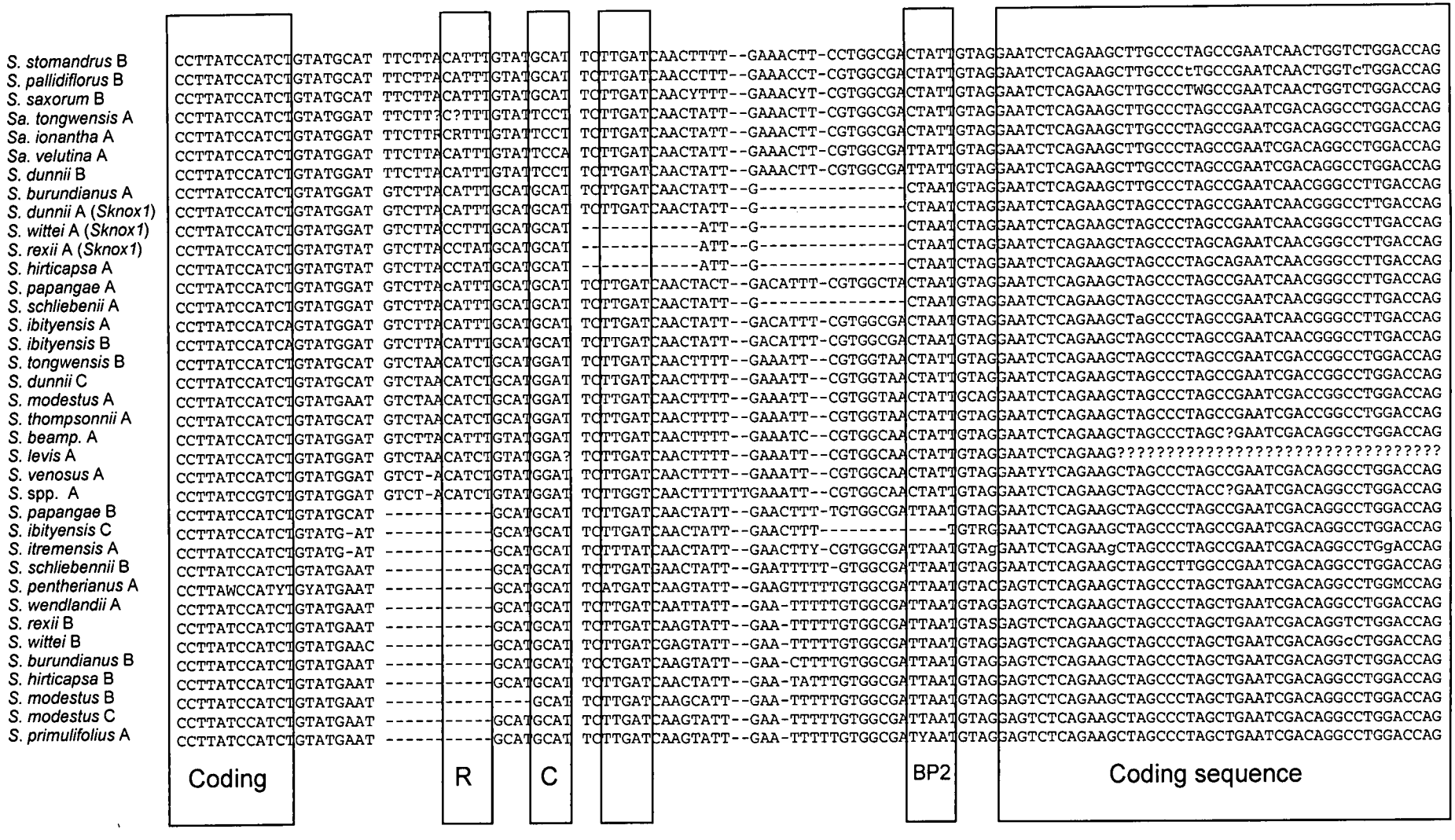


Figure 50: Alignment of coding sequence and conserved positions within intron 3. Positions where alignment was equivocal and data is omitted are indicated by a space. Potential branchpoint positions are indicated BP1 and BP2. A conserved motif is indicated by a C. A further conserved domain, consensus CMTHT, is marked by an R.

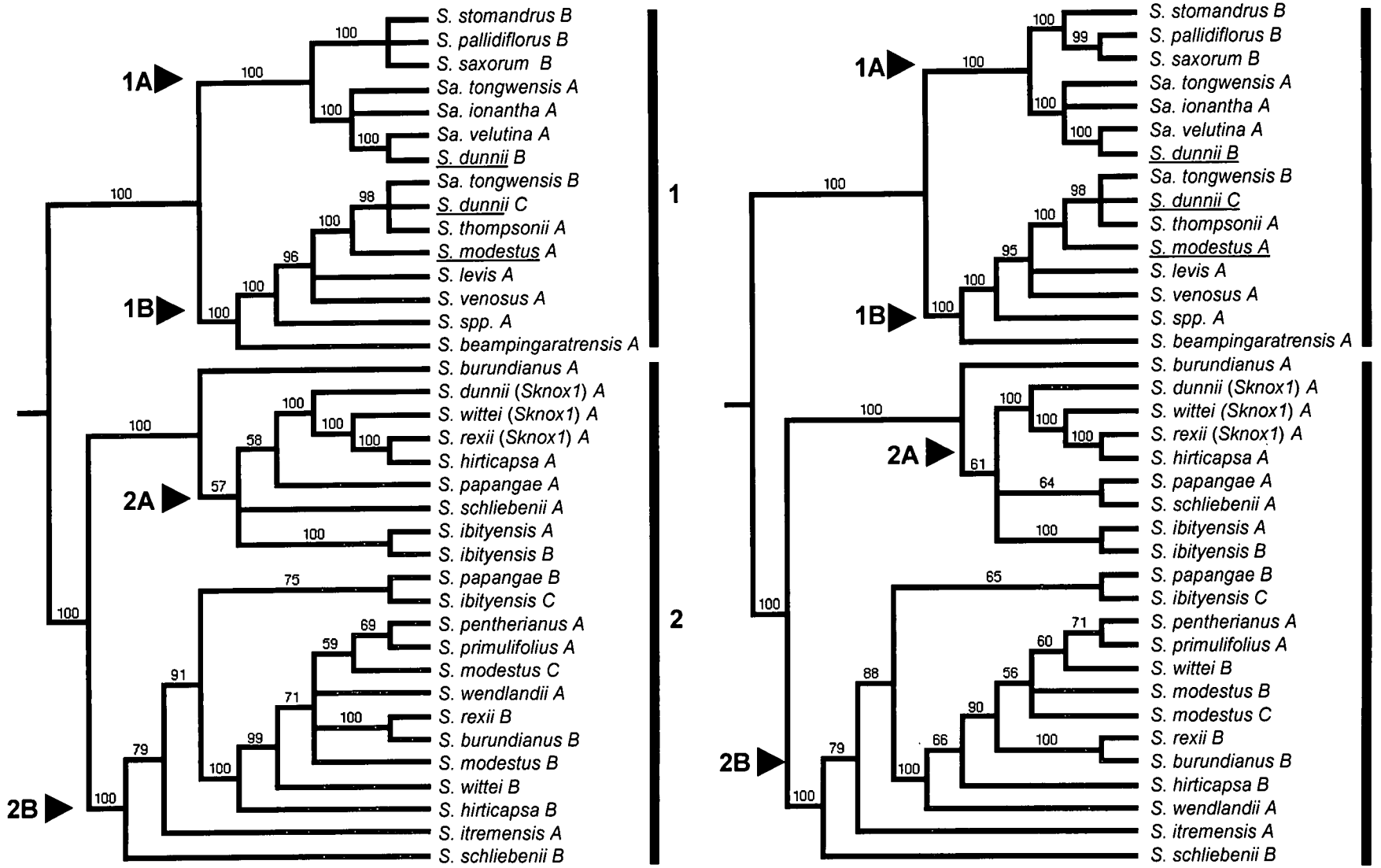


Figure 51: Phylogenetic analysis of coding sequence data and conserved positions in intron data. A: majority rule of 5200 most parsimonious trees. B: majority rule of 5200 most parsimonious trees in re-weighted search. Arrows indicate clades preserved in all most parsimonious trees, bars indicate major clades.

Repetitive elements were found in some species

Some of the sequences isolated from *Sknox* genes could not be aligned with the majority of the sequences due to length differences. In the longer sequences of *S. pallidiflorus*, *S. saxorum* and *S. thysanotus,* which were flanked by coding sequence that showed a high degree of similarity to other *Streptocarpus Sknox1* sequences, the conserved motifs used to aid alignment of shorter intron sequences were present in multiple copies.

DNA repeats are widespread in eukaryotic genomes and are classified according to their size (Richard and Pâques 2000). Microsatellites are repeating elements of <10 bp , and minisatellites are repetitive elements of 10-100bp in length, which are highly unstable, and are commonly GC rich.

Dot plots, which offer a comparison between two sequences by plotting significant matching regions as diagonals, were drawn using the dotmatcher program (JEMBOSS: www.hgmp.mrc.ac.uk) to investigate the length and number of possible repeating units where this was the case (figure 52). Sequences were plotted against themselves, so each plot has one diagonal bisecting the plot. Further diagonals arise due to repeated motifs, and the length of repeats is shown by the distance between diagonals.

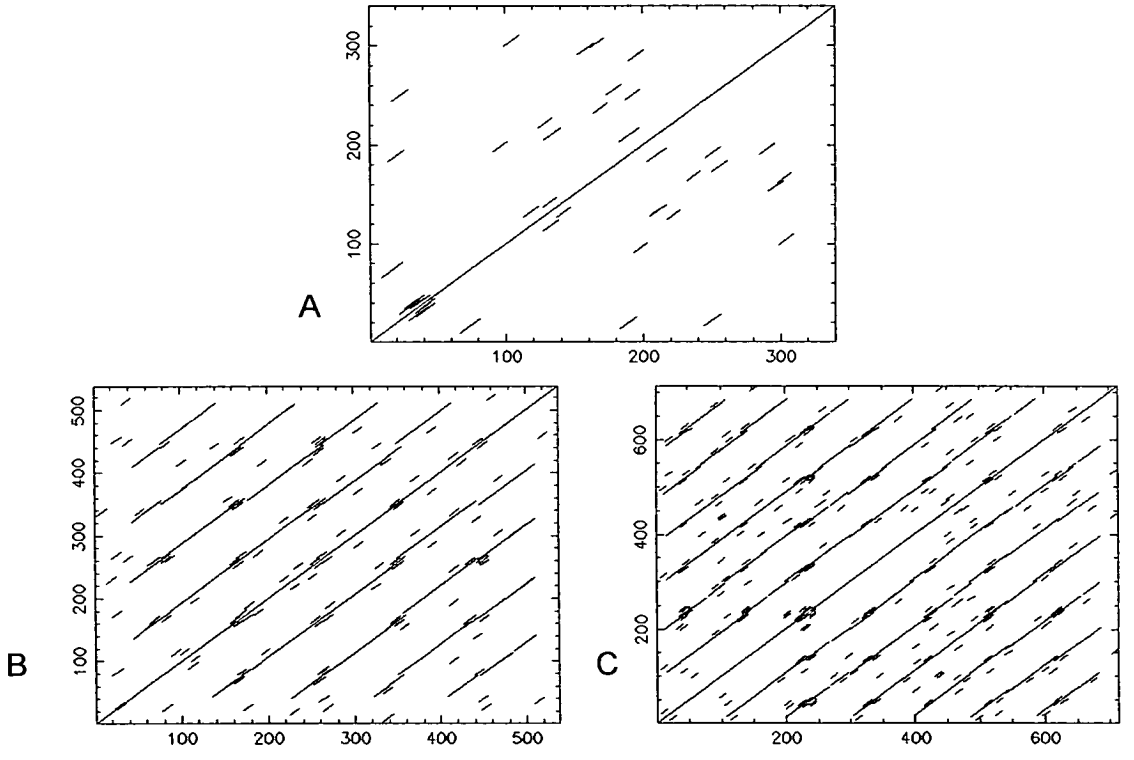


Figure 52: Dot plots showing frequency of repeats in *S. thysanotus* (A), *S. saxorum* (B) and *S. pallidiflorus* (C) introns. X and Y axes are both the number of base pairs from the start of the sequence.

In *S. thysanotus*, several short repeated units of ~ 10-20 bp were detected (figure 52). However, in *S. saxorum* and *S. pallidiflorus* most of the intron was comprised of repeats of about 90 bp, and the differences in length between introns reflected differences in the number of repeats (figure 52).

The high degree of similarity between repeats facilitated their alignment. Each repeated unit showed a high degree of similarity to the entire intron sequence of shorter introns in both species (figure 53). A short stretch of sequence that closely resembled the homeodomain coding sequence 3' of the intron, and a region that reiterated the 3' intron exon boundary sequence were also included in the repeat unit. These sequences were modified between repeats, so that successive repeats shared progressively less similarity to the original intron, intron: exon junction and coding sequence.

As intron: exon boundaries are involved in lariat formation in splicing, it was possible that this reiteration could either disrupt the function of *Sknox* alleles containing repeats, or lead to differential splicing. Neither of these possibilities seemed likely, as in 3' RACE experiments using the same primer that was used to amplify intron sequences, products amplified with ease and appeared correctly spliced. This showed that alleles containing repeats were expressed, and that they were similar to other homeodomain sequences.

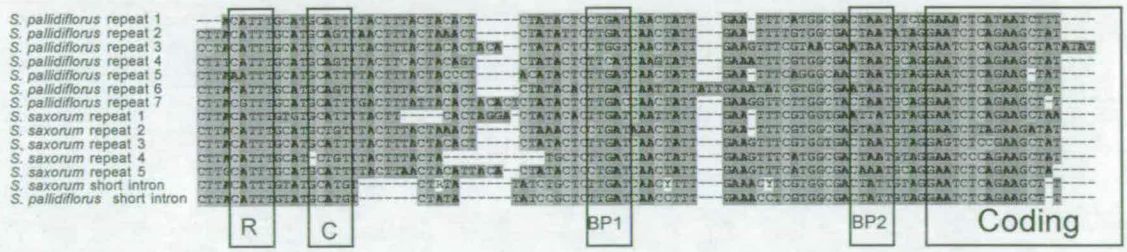


Figure 53: Alignment of repetitive elements in intron 3 of *S. pallidiflorus* and *S. saxorum* with intron 3 of the second putative *Sknox* gene form the same species.

The initiation of repeat units could be identified clearly by the sequence ACATTTG, or a close variant (consensus WMATTTG). Their end was identified by 15 nucleotides that resembled coding sequence and translated to ESQKL, or a variant. Repeats were numbered from the 5' to the 3' end of the intron, so that the first repeat was close to the start of the intron and the last repeat encompassed the end of the intron and the first 15 bp of homeodomain coding sequence. Repeats from both species were aligned with each other, and also with the sequences from the corresponding shorter introns, in ClustalW

(figure 53). The relationships between repeats were analysed, as described for the coding and coding plus intron sequence data sets. Using unweighted characters in a heuristic search, 107 most parsimonious trees were generated, and their majority rule consensus is shown in figure 54 A. An analysis with re-weighted characters was also performed, and this is shown in figure 54 B. Trees were rooted on the gene duplicates from both taxa. In both analyses repeat 2 from *S. pallidiflorus* clustered with repeats 2 and 4 from *S. saxorum.*

This implies that repeats 2 and 4 of *S. saxorum* evolved after speciation of *S. pallidiflorus* and *S. saxorum*, or that their sequences have been homogenised. Also, repeat 1 from *S. pallidiflorus* clustered with repeat 5 from *S. saxorum* and together these clustered with repeat 3 from *S. pallidiflorus*. This grouping could imply that the evolution of repeats 1 and 5 occurred after speciation.

It is interesting that repeats designated odd numbers from each species clustered and that those with even numbers also clustered, as this pattern is suggestive of evolution of repeats in tandem, rather than one by one. However, repeat 1 from *S. saxorum* clustered with repeat 6 from *S. pallidiflorus.*

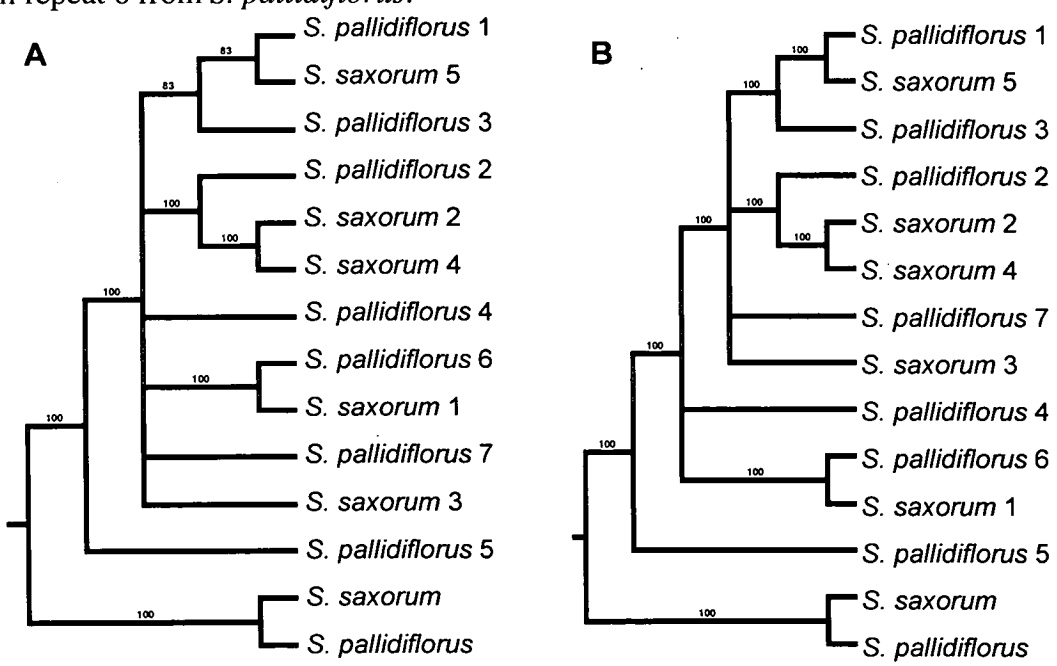


Figure 54: Phylogeny of intron repeats in two species of *Streptocarpus.* A: majority rule of 107 most parsimonious trees using unweighted characters. B: majority rule of most parsimonious trees with characters re-weighted using the re-scaled consistency index. Trees are rooted on *S. saxorum* and *S. pallidiflorus* sequences from the shorter intron. Repeats from the long introns are sequentially numberered, from th 5' to the 3' end.

Summary

Sequences were isolated from *Sknox* introns that share conserved features of functional introns in other plant species. Phylogeny using short stretches of coding sequence that flanked intron 3 of *Sknox* genes supported an African caulescent clade, and this was consistent with previous phylogenies. Weak support was given for the subdivision of sequences into two main clades, that each largely reflected species relationships within the genus, and this pattern was suggestive of an *Sknox* duplication that preceded the subgeneric split of *Streptocarpus*. Phylogeny using the most conserved regions of intron 3 in conjunction with the flanking coding sequence supported the notion of a gene duplication, in that most species were represented by two sequences, and these sequences clustered into four groups each corresponding to a subgeneric group. However, in contrast to analyses using coding sequence alone, the main division of the consensus trees was also at the subgeneric level, and these groups were better supported. This is consistent with either a gene duplication that occurred after the subgeneric division of *Streptocarpus*, or with concerted evolution of intron sequences. Repeating elements were found in some *Sknox* introns, and these incorporated a stretch of sequence corresponding to the shorter intron and 15 nucleotides of the 3' coding sequence. Phylogeny using the repeated units was suggestive of their evolution in tandem.

Discussion

Repeat evolution

There are several hypotheses as to how repeats evolve within genomes, all of which involve DNA replication before cell division.

replication

The first mechanism is slippage, which arises when neighbouring repeats mispair during DNA replication. Stand synthesis subsequently incorporates fewer or extra repeats. Slippage typically leads to loss or gain of small numbers of tandem repeats, and is associated with short contiguous repeats of up to 30 bp (Li 1997). It is therefore unlikely to be the main mechanism by which repeats evolved in *Sknox* genes.

recombination

The second main mechanism by which repeats are thought to evolve is by unequal crossing over between sister chromatids, which would lead to their respective expansion or contraction (figure 55 A). Typically repetition resulting from recombination events is

in large numbers of repeats, which range from a few nucleotides in length to several kbp. It is also often associated with the presence of chi-like sequences. These are GC rich tracts (consensus GCTGGTGG), that are considered involved in recombination in *E. coli* and man (Lam *et al.* 1974, Jeffreys *et al.* 1985). In intron 3 sequences, there were few repeats, and no Chi-like sequences, which suggests that it is unlikely that classical recombination was the mechanism by which repeats evolved in *Streptocarpus*.

Gene conversion

A third means of expansion or contraction involves non reciprocal exchange of DNA between chromatids. This is by gene conversion, in which one allele replaces its sister by homologous recombination following replication of the DNA (figure 55 B). In gene conversion, a double strand break occurs in one chromosome, and the newly synthesised strand melts. This re-anneals to the newly synthesised strand of the other chromosome, which acts as a template to synthesis of new DNA in the first strand in DNA repair. Both strands of the first chromosome acquire an insert from the second, which remain unaffected.

Gene conversion may also be associated with crossover (figure 55 D). As with gene conversion unaccompanied by crossover, a double strand break occurs in one chromosome following DNA replication, and the newly synthesised strands re-anneal to each other. The unbroken chromosome acts as a template to the broken strands in DNA synthesis. DNA exchange has a chance of occurring between the two newly synthesised strands by cross over.

Both of these mechanisms can lead to a change in the number of repeats in a sequences, as annealing of the broken strands can occur at any complementary site on the unbroken sister strands (figure 55 D).

Evolution of repeats in Streptocarpus

The repeats in *S. pallidiflorus* and *S. saxorum* introns are ~100 bp, which is longer than the typical length of repeats that arise by slippage. This supports the hypothesis that they arose by recombination events. *Streptocarpus* repeats do not contain Chi-like sequences, so it seems more likely that they arose by a gene conversion mechanism rather than by classical cross-over. As the type of repeat and pattern of their variation is most consistent with their evolution by gene conversion, and this is one process that has been thought to cause homogenisation of repetitive sequence in concerted evolution,

hypotheses as to the sequence of evolution of repeats in *Sknox* genes are weakly founded.

As the organismal phylogeny of *Streptocarpus* is well characterised, a better way to determine how repeating elements evolved (by loss or gain) in *Streptocarpus* introns would be to map their evolution onto the organismal phylogeny, and examine the relationships between them. This would determine the direction of differences in intron length, and whether the repeat number increased or decreased, and could also determine which repeats evolved first. This would require at least three taxa with repeated elements in their introns, but with the limited sample size available to me, has not yet been possible. As repeats were only found in subgenus *Streptocarpella*, it would be interesting to further investigate members of that clade.

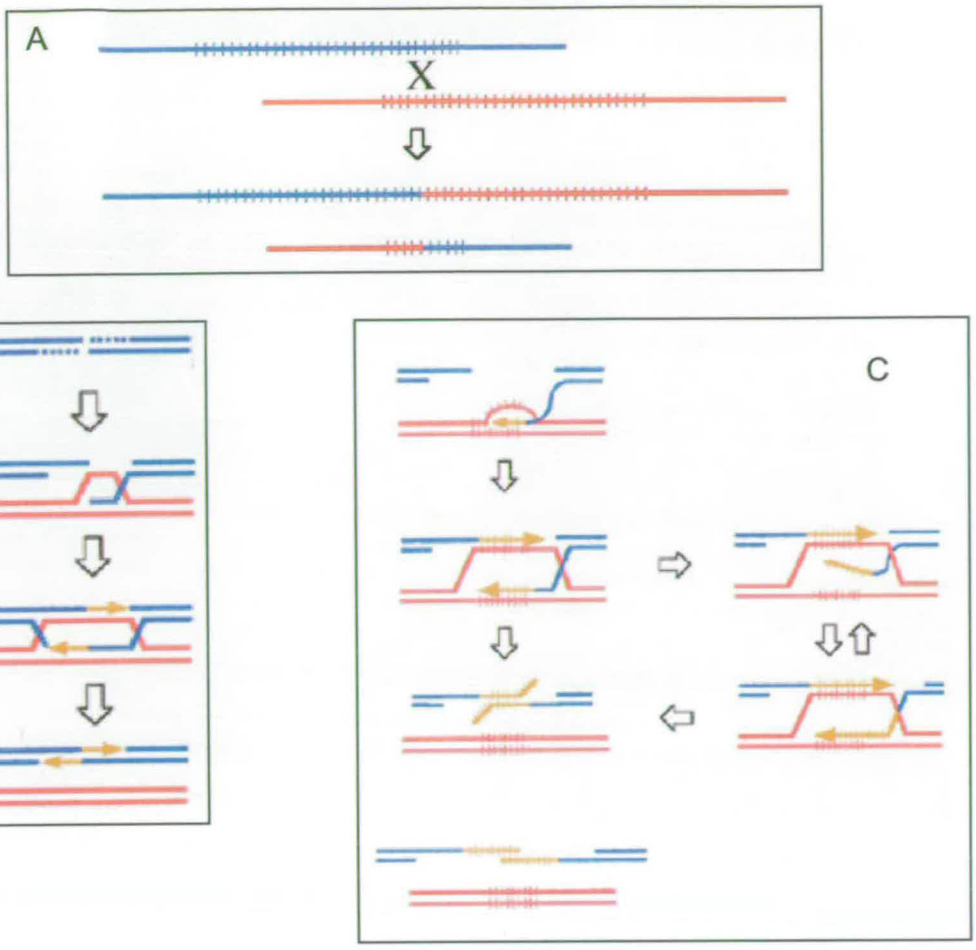


Figure 55: Models of repeat expansion and contraction by recombination, redrawn from Richard and Pâques (2000).
A: alteration in the number of repeats following unequal recombination.
B: illustration of gene conversion.
C: illustration of how gene conversion can lead to a change in the number of repeats in a repetitive sequence.
D: illustration of gene conversion accompanied by crossover.

Chapter 7: Discussion

The main aim of this research was to investigate whether *Streptocarpus* homologues of *STM* could be implicated in morphological innovation and evolution of *Streptocarpus* by altering meristem behaviour through changes in expression. Although differences in *Sknox1* expression were found between species of *Streptocarpus* with different growth forms, these were found to be correlative rather than causal in the two species that could be crossed. How then did *Streptocarpus* evolve such diverse morphology?

Morphological change appears rapid in *Streptocarpus*

Major morphological change, such as change in body plan, has been considered to occur over long periods of evolutionary time. Although arguments for change over shorter periods have been presented, they are not yet widely accepted, partly due to the larger popularity of animal studies than plant (Goldschmidt 1940, Bateman and DiMichele 1994). Phylogeny has shown that the major differences between growth forms of *Streptocarpus* have arisen over an evolutionarily short time scale(probably <18 MYA: Möller and Cronk 2001a). Differences in growth form within *Streptocarpus* are also homoplastic - there appear to have been multiple origins of all three major types, and some of these appear associated with reversions. This implies that the evolution of substantial differences in growth form has been plastic in *Streptocarpus*. Unifoliate and rosulate species hybridise freely, and the large differences in their form appear determined by few genes.

Recent phylogenetic and cytological work (Möller and Cronk 2001a) suggest that some Madagascan caulescent species with 2n=32 are plesiomorphic with respect to subgenus *Streptocarpus*, and these results are supported by my intron sequencing data. As most unifoliate and rosulate species have a chromosome number of 2n=32, this raises the possibility of crosses between caulescent and unifoliate or rosulate species, which could be used to investigate the genetic basis of further differences in growth form within the genus.

Although *Sknox1* is not one of the loci that jointly determines the inheritance of the rosulate character, its novel expression patterns in these species suggest that it may be part of a pathway that is directly implicated in the differences between these forms. It is also possible *Sknox1* genes were involved in the transition between caulescent and acaulescent types, which is suggestive of a loss in capacity to form a shoot. If *Sknox1*

genes were involved in this transition, and had lost their function, we could expect that they would not be expressed in acaulescent species. As *Sknox1* genes are expressed in acaulescent species, but their expression is not confined to the meristem, it is more likely that a regulator of *Sknox1* expression is implicated in the transition. Some candidates for interaction with *Sknox1* genes are available by analogy to model species, and these were discussed in chapters 4 and 5.

Inflorescence meristem development in acaulescent species

It is interesting that although *STM* is implicated in meristem function in other species, and confers meristem-like properties on leaves when mis-expressed, its expression is not sufficient to induce meristem formation. This also appears the case in *Streptocarpus*, where *Sknox1* was expressed in the phyllomorph in two species, but only rarely did further phyllomorphs arise from the primary phyllomorph, and this was usually associated with tearing of the lamina above a vein. However, in *S. dunnii* and *S. rexii* inflorescences conventionally arise from the proximal part of the midrib necessitating the formation of a meristem. Presumptive inflorescence meristems are identifiable as mounds of cells with meristem-like organisation. *Sknox1* was expressed proximally and distally in the midrib of both acaulescent species studied. In *S. rexii* the distance between inflorescence meristems varies (pers. obs.), and it is not known what determines where on the midrib they arise. One possibility is that they arise in response to a proximo-distal signal gradient. It would be intriguing to investigate this, and whether the entire midrib has the capacity to form inflorescence meristems, by transplanting distal midrib to more proximal regions.

Possible means of identification of the loci that determine form

Although *Sknox1* genes may be indirectly linked to the evolution of morphological novelty in *Streptocarpus*, it would be most interesting to discern what genes do actually determine the inheritance of the rosulate character over the unifoliate character.

At present there are two possible ways that this could be achieved. The first is by mapping - if a genetic map of *Streptocarpus* became available it would be possible to find genetic markers that linked to the rosulate or caulescent characters in crosses. These could then be used as anchor points from which to clone the loci implicated in determining these characters by chromosome walking.

The second is to look for differences in gene expression between species with different growth forms. Two means of doing this could be done in *Streptocarpus*. One way is subtractive hybridisation (Diatchenko *et al.* 1999), and the second way is AFLP-cDNA display. Both of these approaches would at present be fairly impractical as they would be labour intensive and expensive, but may become more feasible in the future.

Mechanisms of evolution

As stated in the introduction, morphological variation is a pre-requisite of Darwinian evolution by natural selection. Another aim of this research was to investigate the genetic mechanisms by which morphological variation arises in natural circumstances. In animals, development is predictable in that it leads to formation of an adult with determinate form. Therefore mutations that have large morphological effects are likely to be severely detrimental to an individual's survival and reproductive success. The importance of continuous variation which arises from the accumulated impact of few to many mutations of little effect, has therefore been stressed (Doebley and Lukens 1998), but there is recent evidence of subtle morphological differences between species that are caused by mutations at single developmental loci.

There are many examples of laboratory induced mutations in single genes in Drosophila that have profound phenotypic effects, such as *Antennapedia* and *Ultrabithorax* mentioned in the introduction. Theoretically these provide a genetic basis by which morphological variation in nature could be generated. Evidence of such processes is given by several examples of natural phenotypic variation that correlates with differences in gene expression. An example is butterfly wing eyespots, which are important in determining the level of predation on butterflies as they resemble large eyes comprised of concentric coloured rings. Eyespots are formed around an organizer region called the focus by pigmented scales, and are specified by the action of different transcription factors and signalling molecules in four stages. Colour pattern is specified in the last two stages, coinciding with the formation of scales and in response to signalling from the eyespot focus. Genes involved in determining eyespot formation are *Engrailed/Invected, Spalt* and *Distal-less,* and differences in their expression patterns in different species of the butterfly genus *Bicyclus* are correlated with morphological differences between the eyespots in those species (Brunetti *et al.* 2001).

Evidence of links between genetic and morphological variation

A more direct link between the generation of morphological diversity by sequence differences has been established in two recent studies of the pattern of hair growth in closely related species of *Drosophila*. In one of these(Sucena and Stern 1997), two species of *Drosophila*, one with fine hairs on the dorsal cuticle of first instar larvae, and the other naked, were crossed. In hybrids, all larvae had fine hairs, implying dominance of the hairy character. For an X-linked trait in flies, if the hairy character is determined by an allele H, the female hybrid must have the genotype H/h. Male flies carry one X chromosome. Therefore in backcrosses of hybrid females (genotype H/h) to species males (H/0 and h/0 respectively) a 3:1 and a 1:1 ratio of hairy to naked larvae were obtained, indicating that differences were likely to caused by a single locus on the X chromosome. Flies with partial deletions of the X chromosome carrying the dominant allele were crossed with flies carrying the recessive allele. The naked phenotype was thereby revealed in hybrids when the region associated with the hairy trait was deleted. This indicated the involvement of a chromosomal region previously determined to carry a gene that affected hair pattern, namely *ovo/shavenbaby*. Further crosses and mapping experiments indicated that this was the single gene implicated in determining differences in hair pattern between species. The expression pattern of *svb* correlated with the distribution of hairs in larvae, and where no transcripts were detected there were no hairs, supporting this inference. The precise genetic differences between species at the *ovo/svb* loci have not yet been identified.

In a second study (Stern 1998), the nucleotide changes that cause phenotypic differences were identified. This involved crosses between two species of *Drosophila* that are distinguished by the extent of hair growth on their femurs, one having a large patch of naked cuticle, and the other having a small patch. Hair growth is inhibited by the expression of *Ultrabithorax*, a fly homeobox gene, and patterns of growth in flies corresponded to the inverse of a gradient in *Ubx* expression, and to the dosage of *Ubx* expression. The differences in hair growth between species also correlated with levels and patterns of *Ubx* expression in those species. Null alleles of *Ubx* were generated in inbred lines of two species of *Drosophila*, and heterozygous flies from each species were crossed to parent species so that F 1 flies carried a null *Ubx* allele from one species, and a functional allele from the other species were generated. Flies carrying the allele from the parent species with a small naked patch had a small naked patch, and

those with the allele from the second species had a larger naked patch. As the effect of different *Ubx* alleles was determined in flies with identical genetic backgrounds, and Ubx protein sequences are invariant between the species studied, the only differences between alleles must be in *cis*- regulatory sequences. Therefore morphological differences in hair pattern between species of *Drosophila* must have involved mutation in these sequences.

Differences of plant growth and evolution to animals

In angiosperms, growth is not constrained in the same way as in animals, as the plant body is made up of repeating units of internode, leaf, and axillary meristem (Gray 1879, Galinat 1959), and all of these organs are made by the SAM. How these units reiterate depends on the behaviour of meristems in different parts of the plant. Plants therefore have more plastic growth, and adult form is dependent on environmental conditions to a much higher degree, than in animals. Plants also have the capacity to remain viable and attain reproductive success even if a mutation of substantial morphological effect in one part of their body has occurred. Also, such mutations have the chance to become heritable, as flowers can be formed from different parts of plants. It is therefore possible that variation introduced by mutation is better tolerated by plants than animals, and that phenotypes induced by that variation are more likely to withstand selection. Another reason that plants may better withstand large scale mutations is that they have the widespread capacity to regenerate asexually. Therefore, they stand a greater chance of overcoming a mutation that is deleterious to their reproductive success by further mutation. There are many examples of this occurring in studies of polyploidy and hybridisation, in which changes in chromosome number have led to sterility, which has been overcome by a further change in ploidy (reviewed by Briggs and Walters 1997). In plants, as in animals, there are many laboratory examples of mutations at single loci that have large phenotypic effects, and there are also many examples of natural differences between species that are caused by single loci (Hilu 1983).

Evidence of links between genetic and phenotypic variation in plants

Domestication of maize

Studies on the genetic basis of the domestication of maize have shown that traits selected for by man (particularly increased apical dominance and reduced branching) are associated with several QTLs, one of which is *teosinte branched1* (*tb1*; Doebley *et al.*

1995, Doebley *et al.* 1997, Wang *et al.* 1999). In domestication selection is strong, and it is predicted that nucleotide sequence variation at selected sites should therefore be low. Wang *et al.* (1999) looked for evidence of a reduction in nucleotide sequence diversity at the *tb1* locus in maize to test this. They found that the predicted transcribed unit from different strains of maize had variable nucleotide composition that did not differ significantly from the level of variation predicted for a neutral gene. Variation in the untranscribed region (UTR) was greatly reduced to 3% of the diversity found in the equivalent region of different *Teosinte* alleles. Therefore the UTR of the *tb1* locus was implicated in the domestication of maize.

This study points to the importance of evolution by *cis*- regulatory changes in genes that are transcriptional regulators in plants, as is the case in animals. However, the selective forces acting on crops bred by man may be quite different from those acting on naturally arising variants, necessitating the study of natural variation.

An epigenetic mutation in *Linaria*

One study of a naturally arising morphological plant mutant suggested a different mechanism for the generation of variation that was epigenetic (Cubas *et al.* 1999). This was the naturally arising peloric mutant of *Linaria* that was originally described by Linnaeaus (1749). Crosses between peloric plants and normal plants indicated that asymmetry was dominant to peloria, and was determined by one locus. Restriction fragment length polymorphisms between F2 plants associated with a *Linaria* homologue of *Cycloidea (Lcyc)*, a gene that partly determines floral symmetry in *Antirrhinum* (Luo *et al.* 1996), were detected. In peloric *Linaria*, this homologue was not expressed, indicating that a mutation blocking its function was probably responsible for the mutant phenotype. An attempt to identify the mutation that disrupted *Lcyc* function in peloric plants using further RFLP analysis showed no polymorphisms at sites previously determined to be polymorphic using methylation sensitive enzymes. It was demonstrated by the use of isoschizomers that the difference between normal and peloric *Lcyc* was that the peloric *Lcyc* was methylated, and this relationship was maintained within plants that had sectors with both normal and peloric flowers.

These case studies illustrate possible mechanisms by which morphological differences may arise through small changes in the behaviour of developmental regulatory genes,

but as yet there are no examples of genetic variation at particular loci that causes differences in phenotypic traits between plant species, as in animals (Stern 1998).

Streptocarpus *as an evolutionary and developmental case study*

Streptocarpus provides a very interesting case study of evolution, as it has evolved such a wide variety of growth forms in a short period of evolutionary time, and the novel growth forms within the genus have a clear ecological value, in that they can colonise cliffs and screes, which may be free from competition (Burtt 1970, Burtt 1998). As some species cross freely, it may be possible to directly identify the genetic changes that have been associated with the evolution of these species in the future. It is tantalising that *Sknox1* genes show different expression patterns that correlate with the growth forms of *Streptocarpus*, as to me this suggests the involvement of a regulator of *Sknox1* expression. Some candidates for promotion of *STM* expression have already been discussed in chapters 4 and 5. Although there are known repressors of *knox* gene function in other species, there are as yet no known repressors of *STM*. Identification of the loci that determine the differences between growth forms of *Streptocarpus* in the future could therefore offer substantial insight into both plant development and the genetic basis of phenotypic variation. The differences in patterns of *Sknox1* expression between *Streptocarpus* species lends credence to the idea that changes at developmental loci might be particularly important for generating morphological diversity (Doebley and Lukens 1998).

None of the previous studies that have looked at which genetic differences at particular loci cause morphological differences have extended to investigate how these differences behave in plant or animal populations, and it would be very interesting to investigate how this sort of genetic regulatory changes behave in natural populations.

Phylogenetic utility of *Sknox* gene introns

Another aim of this research was to investigate the utility of *Streptocarpus knox* genes in phylogeny. The data presented from the third intron of the *Sknox* genes shows that intron sequences are sufficiently divergent to reproduce an organismal phylogeny, even though they are short. This implies that they are rapidly evolving, and may there fore be of use ingroups that are unresolved by ITS sequencing.

Technical limitations during this research

Genetics

During my research I encountered many technical hurdles, which were sometimes due to the fact that *Streptocarpus* has not previously been used as a genetic model. Although there was some genetic information about the inheritance of morphology between unifoliate and rosulate *Streptocarpus*, this was limited. The generation time in *Streptocarpus* is at least a year, so crosses to determine the genetic basis of traits are slow.

In natural species, the genetic background is heterogeneous. From a genetic point of view, the ideal would be to generate inbred lines from plants of species designated as models for *Streptocarpus*, giving genetic homogeneity. However, this may lead to inbreeding depression, which could have a drastic effect on plant form. It would also mean that the genet used was less representative of its species.

This heterogeneity of genetic background also generated problems in trying to interpret the data I obtained - I had to distinguish between alleles of a single locus and multiple loci. Testing this was only possible using the backcross plants, where what I thought was one *Sknox1* locus was inherited in a Mendelian way.

Molecular biology

PCR

Several PCR based techniques were used to isolate *Sknox* homologues from *Streptocarpus*. The specificity of these was variable, and during my research I cloned partial sequences of many genes in place of *knox* genes. Different techniques varied widely in the amplification of non-specific products. In my hands 3'RACE worked reliably and consistently. However, I found that with 5'RACE in particular I frequently amplified non-specific products, and that those that were specific were short. This may be because in 5'RACE an adaptor is ligated to the cDNAs, and amplification is dependent on adaptor-specific as well as gene-specific primers. I found that sequences from cloned 5'RACE fragments frequently had adaptor sequence at both ends, so priming must have occurred with the adaptor primers alone. One way to get around this problem could be to use longer gene specific primers, and to perform the PCRs at high annealing temperatures, ensuring greater specificity of priming.

Quality of DNA using standard extraction techniques

Systematics research uses PCR as a standard tool for generating sequence based phylogenies. This is dependent on DNA extraction from plant material, usually leaf discs. However, the quality of DNA sufficient for PCR is fairly low, and a crude method of preparation can be used. A similar standard for molecular biology is the restriction digest, and restriction must precede Southern blotting, as well as IPCR and genome walking. The quantity of DNA needed for restriction is much greater than for PCR: for a typical genomic digest I used 2.5-5.0µg DNA. The quality of DNA also affected the degree of digestion.

Approaches to investigating gene function

My research was also limited by attempts at investigating *Sknox1* gene function. One approach that I tried was in situ hybridisation in tissue sections to more precisely localise where *Sknox1* is expressed in the different species I examined. In particular, it would be interesting to determine if the pattern of *Sknox1* expression is established similarly for species with different growth form in embryos and seedlings. In *Antirrhinum, hirz* and *ina* do not always appear strongly expressed by *in situ* hybridisation, but established protocols do allow detection of their expression (Golz *et al.* 2002). I tried to detect *Sknox1* expression using the same protocol, but was unable to. This could be attributed to differences in the region of the genes used as a probe. John Golz used a short 5' fragment of both genes as a probe, and I initially used the 3' UTR of *Sknox1*, and then a 5' fragment, but have not yet attempted using an equivalent probe to John's. Another possible reason that my experiments did not detect expression is that *Streptocarpus* tissue requires a different degree of fixation than *Antirrhinum*. I tried to address this possibility by using three different fixatives (4% Paraformaldehyde, and FAA for 2 h, or FAA overnight), but again, no signal was detected.

A second tool that would be useful is transformation, that would allow a test of whether a particular gene would be sufficient to complement a recessive phenotype in *Streptocarpus*. *Streptocarpus* and *Saintpaulia* regenerate well from leaf cuttings, and transformation procedures are currently being developed at the University of Nottingham for *Saintpaulia*. As *Saintpaulia* is nested within *Streptocarpus* (Möller and Cronk 1997), these should be transferable to *Streptocarpus*.

It would also be very useful to manufacture gene knock-outs by RNA interference, or homologous recombination to investigate gene function, but again, lack of transformation prevents use of these tools in *Streptocarpus*.

References

Aida M, Ishida T, Tasaka M. 1999. Shoot apical meristem and cotyledon formation during *Arabidopsis* embryogenesis: interaction among the *CUP-SHAPED COTYLEDON* and *SHOOT MERISTEMLESS* genes. *Development* **126**: 1563-1570.

Averoff M, Akam M. 1995. Hox genes and the diversification of insect and crustaceans body plans. *Nature* **37**: 420-423.

Bachem CWB, Oomen RJFJ, Visser RGF. 1998. Transcript imaging with cDNA - AFLP: a step-by-step protocol. *Plant Molecular Biology and Reproduction* **16**: 157-173.

Barton MK, Poethig S. 1993. Formation of the Shoot Apical Meristem in *Arabidopsis thaliana*: an analysis of development in the wild type and the *Shoot Meristemless* Mutant. *Development* **119**: 823-831.

Bateman RM, DiMichele WA. 1994. Saltational evolution of form in vascular plants: a neoGoldscmidtian synthesis. In: Hudson and Ingram eds. *Shape and form in plants and fungi,* 63-92.

Bellaoui M, Pidkowich MS, Samach A, Kushalappa K, Kohalmi SE, Modrusan Z, Crosby WL, Haughn GW. 2001. The Arabidopsis BELL1 and KNOX TALE homeodomain proteins interact through a domain conserved between plants and animals. *Plant Cell* **13**: 2455-2470.

Benson DA, Karsch-Mizrachi I, Lipman DJ, OstellJ, Rapp BA, Wheeler DL. 2000. GenBank. *Nucleic Acids Research* **28**: 15-18.

Bharathan GB, Janssen B, Kellogg EA, Sinha N. 1997. Did homeodomain proteins duplicate before the origin of angiosperms, fungi and metazoa? *Proceedings of the National Academy of Sciences USA.* **94**: 13749-13753.

Bharathan G, Janssen B, Kellogg E, Sinha N. 1999. Phylogenetic relationships and evolution of the KNOTTED class of plant homeodomain proteins. *Molecular Biology and Evolution* **16**: 553-563.

Boguski MS, Lowe TM, Tolstoshev CM. 1993. dbEST- database for "expressed sequence tags". *Nature Genetics* **4**: 332-333.

Bohmert K, Camus I, Ballini C, Bouchez D, Caboche M, Benning C. 1998. *AGO1* defines a novel locus of *Arabidopsis* controlling leaf development. *The EMBO Journal* **17**: 170-180.

Bowman J, Eshed Y. 2000. Formation and maintenance of the shoot apical meristem. *Trends in Plant Science* **5**: 110-115.

Briggs D, Walters SM. 1997. *Plant Variation and Evolution.* Cambridge University Press, Cambridge UK.

Brown JR, Doolittle WF. 1995. Root of the universal tree of life based on ancient amino-acyl-tRNA synthetase gene duplications. *Proceedings of the National Academy of Sciences 92*: 2441-2445.

Brown JWS, Smith P, Simpson CG. 1996. *Arabidopsis* consensus intron sequences. *Plant Molecular Biology* **32**:531-535.

Brunetti CR, Selegue JE, Monteiro A, French V, Brakefield PM, Carroll SB. 2001. The generation and diversification of butterfly eyespot colour patterns. *Current Biology* **11**: 1578-1585.

Bürglin TR. 1997. Analysis of TALE superclass homeobox genes (MEIS, PBC, *KNOX*, Iroquois, TGIF) reveals a novel domain conserved between plants and animals. *Nucleic Acids Research* **25**: 4173-4180.

Bürglin TR. 1998. The PBC domain contains a MEINOX domain: Coevolution of Hox and TALE homeobox genes? *Development Genes and Evolution* **208**:113-116.

Burtt BL. 1970. Studies in the Gesneriaceae of the Old World XXXI: some aspects of functional evolution. *Notes from the Royal Botanic Garden Edinburgh* **30**: 1-10.

Burtt BL. 1994. A commentary on some recurrent forms and changes of form in angiosperms. In: Hudson and Ingram eds. *Shape and form in plants and fungi,* 143-152.

Burtt BL. 1998. Climatic accommodation and phytogeography of the Gesneriaceae of the Old World. In Philip Mathew and M. Sivadasan eds.: *Diversity and Taxonomy of Tropical Flowering Plants*: 1-27. Calicut, India: Mentor Books.

Byrne ME, Barley R, Curtis M, Arroyo JM, Dunham M, Hudson A, Martienssen RA. 2000. Asymmetric leaves1 mediates leaf patterning and stem cell function in *Arabidopsis*. *Nature* **408**: 967-971.

Caspary R. 1858. Über die anisokotylie von *Streptocarpus polyanthus* Hook. und *Streptocarpus rexii* Lindl. *Sber. niederrhein. Ges. Nat.-u. Heilk* **15**: 74-75.

Chen J-J, Janssen B-J, Williams A, Sinha N, 1997. A gene fusion at a homeobox locus: alterations in leaf shape and implications for morphological evolution. *Plant Cell* **9**:1289-1304.

Chuck G, Lincoln C, Hake S. 1996. *KNAT1* induces lobed leaves with ectopic meristems when overexpressed in *Arabidopsis*. *Plant Cell* **8**: 1277-1289.

Citerne HL, Möller M, Cronk QCB. 2000. Diversity of *cycloidea*-like genes in Gesneriaceae in relation to floral symmetry. *Annals of Botany* **86**: 167-176

Clark SE, Running MP, Meyerowitz EM. 1993. *CLAVATA1*, a regulator of meristem and flower development in *Arabidopsis*. *Development* **119**: 397-418.

Clark SE, Running MP, Meyerowitz EM. 1995. *CLAVATA3* is a specific regulator of shoot and floral meristem development affecting the same process as *CLAVATA1*. *Development* **121**: 2057-2067.

Clark SE, Jacobsen SE, Levin JZ, Meyerowitz EM. 1996. The *CLAVATA* and *SHOOT MERISTEMLESS* loci competitively regulate meristem activity in *Arabidopsis*. *Development* **122**: 1567-1575.

Clark SE, Williams RW, Meyerowitz E. 1997. The *CLAVATA1* gene encodes a putative receptor kinase that controls shoot and floral meristem size in *Arabidopsis*. *Cell* **89**: 575-585.

Crocker CW. 1860. Notes on the germination of certain species of Cyrtandroideae. *Botanical Journal of the Linnaean Society* **5**: 74-75.

Cronk Q, Möller M. 1997. Strange morphogenesis - organ determination in *Monophyllaea*. *Trends in Plant Science* **2**: 327-328.

Cubas P, Vincent C, Coen E. 1999. An epigenetic mutation responsible for natural variation in floral symmetry. *Nature* **401**:157-161.

Diatchenko L, Lukyanov S, Lan YF, Siebert PD. 1999. Suppression subtraction hybridisation: a versatile method for identifying differentially expressed genes. *Methods in Enzymology* **303**: 349-380.

Doebley J, Stec A, Gustus C. 1995. *Teosinte branched1* and the origin of maize: evidence for epistasis and the evolution of dominance. *Genetics* **141**: 333-346.

Doebley J, Lukens L. 1998. Transcriptional regulators and the evolution of plant form. *The Plant Cell* **10**: 1075-1082.

Doebley J, Stec A, Hubbard L. 1998. The evolution of apical dominance in maize. *Nature* **386**: 485-488.

Doyle JJ and Doyle JL. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin* **19**: 11-15.

Dubuc-Lebreux M-A. 1978. Modification of the unifoliate habit of *Streptocarpus wendlandii* by some growth regulators. *Phytomorphology* **28**: 224-238.

Eshed Y, Baum SF, Bowman JL. 1999. Distinct mechanisms promote polarity establishment in carpels of *Arabidopsis*. *Cell* **99**: 199-209.

Feinberg AP and Vogelstein B. 1983. A technique for radio-labelling DNA restriction endonuclease fragments to high specific activity. *Annals of Biochemistry* **132**: 6-13.

Fletcher JC, Brand U, Running MP. 1999. Signalling of cell fate decisions by *CLAVATA3* in *Arabidopsis* shoot meristems. *Science* **283**: 1911-1914.

Frohman MH. 1995. Rapid amplification of cDNA ends. In: Dieffenbach CW and Dveksler GS eds. *PCR Primer:* 381-409.

Frohman MA, Dush MK, Martin GR. 1988. Rapid production of full-length cDNAs from rare transcripts by amplification using a single gene-specific primer. *Proceedings of the National Academy of Sciences USA* **85**: 8998-9002.

Frugis G, Giannino D, Mele G, Nicolodi C, Innocenti AM, Chiapetta A, Bitonti MB, Dewitte W, Van Onckelen H, Mariotti D. 1999. Are homeobox *Knotted*-like genes and cytokinins the leaf architects? *Plant Physiology* **119**: 371-373.

Fujie M, Kuroiwa H, Kawano S, Mutoh S, Kuroiwa T. 1994. Behavior of organelles and their nucleoids in the shoot apical meristem during leaf development in *Arabidopsis thalliana*. *Planta* **194**: 395-405.

Galinat WC. 1959. The phytomer in relation to the floral homologies in the American Maydea. *Bot Mus Leafl Harv Univ* **19**: 1-32.

Goldschmidt R. 1940. *The material basis of evolution.* New Haven, USA: Yale University Press.

Golz JF, Keck EJ, Hudson A. 2002. Spontaneous mutations in *knox* genes give rise to a novel floral structure in *Antirrhinum*. *Current Biology* in press.

Goodall GJ, Filipowicz W. 1989. The AU-rich sequences present in the introns of plant nuclear pre-mRNAs are required for splicing. *Cell* **58**: 473-483.

Goodall GJ, Filipowicz W. 1991. Different effects of intron nucleotide composition and secondary structure on pre-mRNA splicing in monocot and dicot plants. *The EMBO Journal* **10**: 2635-2644.

Gould SJ. 1977. Ontogeny and phylogeny. Cambridge MA: Harvard University Press.

Gray A. 1879. *Structural botany or organography on the basis of morphology*. New York: Ivison Blalkeman and Co.

Hake S, Char BR, Chuck G, Foster T, Long J, Jackson D. 1995. Homeobox Genes in the Functioning of Plant Meristems. *Philosophical Transactions of The Royal Society of London* **350**, pp.45-51.

Hareven D, Gutfinger T, Parnis A, Eshed Y, Lifschitz E. 1996. The making of a compound leaf: genetic manipulation of leaf architecture in tomato. *Cell* **84**: 735-744.

Harrison CJ, Möller M, Cronk QCB. 1999. Evolution and development of floral diversity in *Streptocarpus* and *Saintpaulia*. *Annals of Botany* **84**: 49-60.

Hill AW. 1938. The monocotylous seedlings of certain dicotyledons with special reference to the Gesneriaceae. *Annals of Botany* **2**: 127-144.

Hilliard OM, Burtt BL. 1971. *Streptocarpus*: *an African Plant Study*. Pietermaritzburg, South Africa: University of Natal Press.

Hilu KW. 1983. The role of single-gene mutations in the evolution of flowering plants. *Evolutionary Biology* **16**: 97-128.

Imaichi R, Nagumo S, Kato M. 2000. Ontogenetic anatomy of *Streptocarpus grandis* (Gesneriaceae) with implications for evolution of monophylly. *Annals of Botany* **86**: 37-46.

Jackson D, Veit B, Hake S. 1994. Expression of maize *KNOTTED1* related homeobox genes in the shoot apical meristem predicts patterns of morphogenesis in the vegetative shoot. *Development* **120**: 405-413.

Janssen B-J, Lund L, Sinha N. 1998. Overexpression of a homeobox gene, *Let6*, reveals indeterminate features of the tomato compound leaf. *Plant Physiology* **117**: 771-786.

Jeffreys AJ, Wilson V, Thein SL. 1985. Hypervariable "minisatellite" regions in human DNA. *Nature* **315**: 67-73.

Jeong S, Trotochaud AE, Clark SE. 1999. The *Arabidopsis CLAVATA2* gene encodes a receptor-like protein required for the stability of the *CLAVATA1* receptor-like kinase. *Plant Cell* **11**: 393-405.

Jong K. 1970. Developmental aspects of vegetative morphology of *Streptocarpus*. Ph. D. Thesis, Edinburgh.

Jong K, Burtt BL. 1975. The evolution of morphological novelty exemplified in the growth patterns of some Gesneriaceae. *New Phytologist* **75**: 297-311.

Jong K. 1978. Phyllomorphic organization in rosulate *Streptocarpus*. *Notes from the Royal Botanic Garden Edinburgh* **36**:369-396.

Kerstetter R, Vollbrecht E, Lowe B, Veit B, Yamaguchi J, Hake S. 1994. Sequence analysis and expression patterns divide the maize *knotted1*-like homeobox genes into two classes. *The Plant Cell* **6**: 1877-1887.

Kerstetter RA, Laudencia-Chingcuanco D, Smith LG, Hake S. 1997. Loss-of-function mutations in the maize homeobox gene, *knotted1*, are defective in shoot meristem maintenance. *Development* **124**: 3045-3054.

Kerstetter RA, Hake S. 1997. Shoot meristem formation in vegetative development. *The Plant Cell* **9**: 1001-1010.

Kitching IJ, Forey PL, Humphries CJ, Williams DM. 1998. *Cladistics: the theory and practice of parsimony analysis* (second edition). Oxford University Press, Oxford UK.

Lam ST, Stahl MM, McMilin KD, Stahl FW. 1974. Rec-mediated recombinational hot spot activity in bacteriophage lambda II: a mutation which causes hot spot activity. *Genetics* **77**: 425-433.

Lawrence WJC. 1940. The genus *Streptocarpus*. *Journal of the Royal Horticultural Society* **65**:17-22.

Laufs P, Dockx J, Kronenberger J, Traas J. 1998. *MGOUN1* and *MGOUN2*: two genes required for primordium initiation at the shoot apical and floral meristems in Arabidopsis thaliana. *Development* **125**: 1253-1260.

Laux T, Mayer KFX, Berger J, Jürgens G. 1996. The *WUSCHEL* gene is required for shoot and floral meristem integrity in *Arabidopsis*. *Development* **122**: 87-96.

Lewis EB. 1985. Regulation of the genes of the bithorax complex in *Drosophila*. *Cold Spring Harbour Symposia in Quantitative Biology* **50**: 155-164

Li W-H.1997. *Molecular Evolution.* Sinauer Sunderland MA.

Lincoln C, Long J, Yamaguchi J, Serikawa K, Hake S. 1994. A *Knotted*-like homeobox gene in *Arabidopsis* is expressed in the vegetative meristem and dramatically alters leaf morphology when overexpressed in transgenic plants. *Plant Cell* **6**: 1859-1876.

Linnaeus C. 1749. *De Peloria. Amoenites academicae.*

Long JA, Barton MK. 1998. The development of apical embryonic pattern in Arabidopsis. *Development* **125**: 3027-3035.

Long JA, Moan E, Medford JI, Barton MK. 1996. A member of the KNOTTED Class of Homeodomain proteins encoded by the STM gene of *Arabidopsis. Nature* **379**: 66-69.

Lynn K, Fernandez A, Aida, M, Sedbrook J, Tasaka M, Masson P, Barton K. 1999. The *PINHEAD/ZWILLE* gene acts pleiotropically in *Arabidopsis* development and has overlapping functions with the *ARGONAUTE1* gene. *Development* **126**: 469-481.

Luo D, Carpenter R, Vincent C, Copsey L, Coen E. 1996. Origin of floral symmetry in Antirrhinum. *Nature* **383**: 794-799.

Ma H, McMullen MD, Finer JJ. 1994. Identification of a homeobox- containing gene with enhanced expression during soybean (*Glycine max* L.) somatic embryo development. *Plant Molecular Biology* **24**: 465-473.

Maniatis T, Fritsch EF, Sambrook J. 1982. *Molecular cloning: a laboratory manual.* Cold Spring Harbour Laboratory, Cold Spring Harbour, USA.

Martin C, Northcote DH. 1981. Qualitative and quantitative changes in mRNA of castor beans during the initial stages of germination. *Planta* **151**:189-197.

Mayer FX, Schoof H, Haeker A, Lenhard M, Jürgens G, Laux T. 1998. The role of *WUSCHEL* in regulating stem cell fate in the *Arabidopsis* shoot meristem. *Cell* **95:**805-815.

McCall RA. 1997. Implications of recent geological investigations of the Mozambique channel for the mammalian colonisation of Madagascar. *Proceedings of the Royal Society of London B.* **264**: 663-665.

McConnell JR, Barton MK. 1995. Effects of mutations in the *PINHEAD* gene of *Arabidopsis* on the formation of shoot apical meristems. *Developmental Genetics* **16**: 358-366.

Möller M, Cronk QCB. 1997a. Origin and relationships of *Saintpaulia* (Gesneriaceae) based on ribosomal DNA internal transcribed spacer (ITS) Sequences. *American Journal of Botany* **84**: 956-965.

Möller M, Cronk QCB. 1997b. Phylogeny and disjunct distribution: evolution of *Saintpaulia*. *Proceedings of the Royal Society of London* **264**: 1827-1836.

Möller M, Cronk QCB. 2001a. Evolution of morphological novelty: a phylogenetic analysis of growth patterns in *Streptocarpus* (Gesneriaceae). *Evolution* **55**:918-929.

Möller M, Cronk QCB. 2001b. Phylogenetic studies in *Streptocarpus* (Gesneriaceae): reconstruction of biogeographic history and distribution patterns. *Systematics and Geography of Plants* **71**: 545-555.

Moussian B, Schoof H, Haeker A, Jürgens G, Laux T. 1998. Role of the *ZWILLE* gene in regulation of central shoot meristem cell fate during *Arabidopsis* embryogenesis. *The EMBO Journal* **17**: 1799-1809.

Muehlbauer GJ, Fowler JE, Girard L, Tyers R, Harper L. 1999. Ectopic expression of the maize homeobox gene *liguleless3* alters cell fate in the leaf. *Plant Physiology* **119**: 651-662.

Mullis KB, Faloona FA. 1987. Specific synthesis of DNA *in vitro* via a polymerase catalysed chain reaction. *Methods in Enzymology* **155**: 335-350.

Nagasaki H, Sakamoto T, Sato Y, Matsuoka M. 2001. Functional analysis of the conserved domains of a rice homeodomain protein, *OSH15*. *Plant Cell* **13**: 2085-2098.

Nishimura A, Tamaoki M, Sato Y, Matsuoka M. 1999. The expression of tobacco *knotted1*-type class 1 homeobox genes correspond to regions predicted by the cytohistological zonation model. *Plant Journal* **18**: 337-347.

Nishimura A, Tamaoki M, Sakamoto T, Matsuoka M. 2000. Over-expression of tobacco *Knotted-1* type Class 1 homeobox genes alters various leaf morphology. *Plant Cell Physiology* **41**: 583-590.

Ochman H, Gerber AS, Hartl DS. 1988. Genetic applications of an inverse polymer chain reaction. *Genetics* **120**: 621-625.

Oehlkers F. 1938. Bastardierungsversuche in der gattung *Streptocarpus* Lindl. 1. plasmatische vererbung und die geschlechtbestimmung von zwitterpflanzen. *Zeitschrift Für Botanik* **32**:305-393.

Oehlkers F.1940. Bastardierungverusche in der gattung *Streptocarpus* Lindley. III. Neue ergebnisse über die genetik von wuchsgestalt und geschlechtsbestimmung. *Berlin Deutschen Botanischen Gesellschaft* **38**:76-91.

Oehlkers F. 1941. Faktorenanalytische ergebnisse an artbastarden. *Allegmeine Biologie* 280-289.

Oehlkers F. 1956. Veränderungen in der blühbereitschaft vernalisierter cotyledon von *Streptocarpus*, kenntlich gemacht durch blattstecklinge. *Zeitschrift für Naturforschung* **10b**: 158-160.

Oehlkers F. 1964. Cytoplasmic inheritance in the genus *Streptocarpus*. *Advances in Genetics* **12**: 329-370.

Ohki S. 1994. Scanning electron microscopy of shoot differentiation *in vitro* from leaf explants of the African Violet. *Plant Cell, Tissue, and Organ Culture* **36:** 157-162.

Ori N, Eshed Y, Chuch G, Bowman JL, Hake S. 2000. Mechanisms that control *knox* gene expression in the *Arabidopsis* shoot. *Development* **127**: 5523-5532.

Passner JM, Ryoo HD, Shen L, Mann RS, Aggarwal AK. 1999. Structure of a DNA-bound Ultrabithorax-Extradenicle homeodomain complex. *Nature* **397**: 714-719.

Pautot V, Dockx J, Hamant P, Kronenberger J, Grandjean O, Jublot D, Traas J. 2001. *KNAT2*: evidence for a link between Knotted-like genes and carpel development. The *Plant Cell* **13**: 1719-1734.

Piper DE, Batchelor AH, Chang CP, Cleary ML, Wolberger C. 1999. Structure of a HoxB1-Pbx heterodimer bound to DNA: role of the hexapeptide and a fourth homeodomain helix in complex formation. *Cell* **96**: 587-597.

Ponniah WD. 1971. Recherches sur le fonctionnement floral de *Streptocarpus nobilis* C. B. Cl. University of Paris XI: PhD thesis.

Reiser L, Sánchez-Baracaldo P, Hake S. 2000. Knots in the family tree: evolutionary relationships and functions of *knox* homeobox genes. *Plant Molecular Biology* **42**: 151-166.

Richard G-F, Pâques F. 2000. Mini- and microsatellite expansions: the recombination connection. *EMBO Reports* **1**: 122-126.

Ronshaugen M, McGinnic N, McGinnis W. 2002. Hox protein mutation and macroevolution of the insect body plan. *Nature* advance on line publication 6 February 2002.

Rosenblum IM, Basile DV. 1984. Hormonal regulation of morphogenesis in *Streptocarpus* and its relevance to evolutionary history of the Gesneriaceae. *American Journal of Botany* **7**: 52-64.

Sakamoto T, Kamiya N, Ueguchi-Tanaka M, Iwahori S, Matsuoka M. 2001. KNOX homeodomain protein directly suppresses the expression of a gibberellin biosynthetic gene in the tobacco shoot apical meristem. *Genes and Development* **15**: 581-590.

Sanger F, Coulson AR. 1975. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *Journal of Molecular Biology* **95**: 444-448.

Saw a S, Watanabe K, Goto K, Kanaya E, Morita H, Okada K. 1999. Molecular characterisaition of a meristem and organ identity gene of *Arabidopsis*, FILAMENTOUS FLOWER, encoding a zinc finger and a HMG- related domain. *Genes and Development* **13**: 1079- 1088.

Schoof H, Lenhard M, Haeker A, Mayer KF, Jürgens G, Laux T. 2000. The stem cell population of Arabidopsis shoot meristems is maintained by a regulatory loop between the *CLAVATA* and *WUSCHEL* genes. *Cell* **100**: 635-644.

Schneuwly S, Kuroiwa A, Gehring WJ. 1987. Molecular analysis of the dominant homeotic Antennapedia phenotype. *EMBO Journal* **6**: 201-206.

Schulz SR, Jensen WA. 1968a. *Capsella* embryogenesis: the egg, zygote and young embryo. *American Journal of Botany* **55**: 807-819.

Schulz SR, Jensen WA. 1968b. *Capsella* embryogenesis: the early embryo. *Journal of Ultrastructural Research* **22**: 376-392.

Scott MP. 1999. Hox proteins reach out around DNA. *Nature* **397**: 649.

Sentoku N, Sato Y, Matsuoka M. 2000. Overexpression of rice OSH genes induces ectopic shoots on leaf sheaths of transgenic rice plants. *Developmental Biology* **220**: 358-364.

Sharp PA, Sugden B, Sambrook J. 1973. Detection of two restriction endonuclease activities in Haemophilus parainfluenzae using analytical agarose. *Biochemistry* **12**: 3055.

Shyue SK, Li L, Chang BH-J, L WH. 1994. Intronic gene conversion in the evolution of human X-linked colour vision genes in higher primates. *Science* **269**: 1265-1267.

Siebert PD, Chenchik A, Kellogg DE, Lukyanov KA, Lukyanov SA. 1995. An improved PCR method for walking in uncloned genomic DNA. *Nucleic acid research* **23**: 1087-1088.

Smith L, Greene B, Veit B, Hake S. 1992. A dominant mutation in the maize homeobox gene, *knotted-1*, causes its ectopic expression in leaf cells with altered fates. *Development* **116**: 21-30.

Souer E, van Houwelingen A, Kloos D, Mol J, Koes R. 1996. The *No Apical Meristem* gene of *Petunia* is required for pattern formation in embryos and flowers and is expressed at meristem and primordia boundaries. *Cell* **85**: 159-170.

Southern EM. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *Journal of Molecular Biology* **98**: 503-517.

Stern DL. 1998. A role of *Ultrabithorax* in morphological differences between *Drosophila* species. *Nature* **396**: 463-466.

Sucena E, Stern DL. 2000. Divergence of larval morphology between *Drosophila sechellia* and its sibling species caused by cis- regulatory evolution of *ovo/shaven-baby*. *Proceedings of the National Academy of Sciences, USA* **97**: 4530-4534.

Sung ZR, Belachew A, Shunong B, Bertrand-Garcia R. 1992. *EMF*, an *Arabidopsis* gene required for vegetative shoot development. *Science* **258**: 1645-1647.

Swofford DL. 1993. PAUP: phylogenetic analysis using parsimony, version 3.1.1. Illinois Natural History Survey, Champaign, IL.

Swofford DL. 1998. PAUP*: phylogenetic analysis using parsimony (*and other methods). Version 4. Sinauer Associates, Sunderland, MA.

Takada S, Hibara K, Ishida T, Tasaka M. 2001. The *CUP-SHAPED COTYLEDON1* gene of *Arabidopsis* regulates shoot apical meristem formation. *Development* **128**: 1127-1135.

Tamaoki M, Kusuba S, Kano-Murakami Y, Matsuoka M. 1997. Ectopic expression of a tobacco homeobox gene, *NTH15*, dramatically alters leaf morphology and hormone levels in transgenic tobacco. *Plant Cell Physiology* **38**: 917-927.

Telford MJ, Holland PWH. 1997. Evolution of 28S ribosomal DNA in chaetognaths:duplicate genes and molecular phylogeny. *Journal of Molecular Evolution* **44**: 135-144.

Thompson JD, Higgins DG, Gibson TJ. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequences alignment through sequence weighting, position specific gap penalties and weight matrix choice. *Nucleic Acids Research* **22**: 4673-4680.

Thompson JD, Gibson TJ, Jeanmougin F, Higgins DG. 1997. The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research* **24**: 4876-4882.

Timmermans MCP, Hudson H, Becraft PW, Nelson, T. 1999. ROUGH SHEATH 2: a Myb protein that represses knox homeobox genes in maize lateral organ primordia. *Science* **284**: 151-153.

Triglia T, Peterson MG, Kemp DJ. 1988. A procedure for *in vitro* amplification of DNA fragments that lie outside the boundaries of known sequences. *Nucleic Acids Research* **16**: 8186.

Tsiantis M, Schneeberger R, Golz JF, Freeling M, Langdale JA. 1999. The maize rough sheath2 gene and leaf development programs in monocot and dicot plants. *Science* **284**: 154-156.

Tsiantis M. 2001. Control of shoot cell fate: beyond homeoboxes. *Plant Cell* **13**:733-738.

Tsukaya H, Naito S, Rédei GP, Komeda Y. 1993. A new class of mutations in *Arabidopsis thalliana*, *acaulis1*, affecting the development of both inflorescences and leaves. *Development* **118**: 751-764

Tsukaya H. 1997. Determination of the unequal fate of cotyledons of a one leaf plant, *Monophyllaea*. *Development* **124**: 1275-1280.

Viémont J-D. 1980. Néoformations des phyllomorphs cotylédonnaires chez le *Streptocarpus rexii* Lindl.: Ontogenèse et histogenèse. *Bulletin de la Société botanique de France*. **127**: 459-469.

Vollbrecht E, Veit B, Sinha N, Hake S. 1991. The developmental gene *knotted-1* is a member of a maize homeobox gene family. *Nature* **350**: 241-243.

Vollbrecht E, Reiser L, Hake S. 2000. Shoot meristem size is dependent on inbred background and presences of the maize homeobox gene, *knotted1*. *Development* **127**: 3161-3172.

Waites R, Selvadurai H, Oliver IR, Hudson A. 1998. The PHANTASTICA gene encodes a MYB transcription factor involved in growth and dorsoventrality of lateral organs in *Antirrhinum*. *Cell* **93**: 779-789.

Walbot V. 1992. Reactivation of Mutator transposable elements of maize by ultraviolet light. *Molecular and General Genetics* **234**: 353-360.

Wang R, Stec A, Hey J, Lukens L, Doebley J. 1999. The limits of selection during maize domestication. *Nature* **398**: 236-239.

Yang C-H, Chen L-J, Sung ZR. 1995. genetic regulation of shoot development in *Arabidopsis*: role of the *EMF* genes. *Developmental Biology* **169**: 421-435.

Yu LP, Simon EJ, Trotochaud AE, Clark SE. 2000. *POLTERGEIST* functions to regulate meristem development downstream of the CLAVATA loci. *Development* **127**: 1661-1670.

Zhou YH, Li WH. 1996. Gene conversion and natural selection in the evolution of X-linked colour vision genes in higher primates. *Molecular Biology and Evolution* **13**: 780-783.

Acknowledgements

Firstly I would like to thank the people who have supervised my PhD, Andrew Hudson for guidance, interest and insight throughout my PhD, Quentin Cronk and H. Tsukaya.
Thanks to Nicola Wright for help with understanding intron structure in *Sknox* genes.
Thanks to everybody who has passed through the Hudson and Goodrich labs during my time there for help and humour: Tony Bishopp, Amanda Borking, Karine Coenen, Peter Etchells, John and Sarah Golz, Thomas Guebitz, Gwyneth Ingram, Graeme Kirk, Peter Newton, Lucia Primavesi, Sophie Smout, Christine Stock, Carol Wilson, in particular to John, Andrew, Pete N. and Lucia for technical advice and training.
Thanks to M. Hollingsworth (RBGE), H. Atkinson, L. Evenden, J. Lovell for technical support.
Thanks to S. Scott for managing the *Streptocarpus* collection so well.
Thanks to Tim Hewson and Jenny Harrison for proof reading.
Thanks to the Gatsby Charitable Foundation for generous financial support and training throughout my PhD, and The British Council for sponsoring my visit to Japan.
Thanks to N. O'Looney and A. Borking for many coffee and moaning breaks.
Thanks to C. H., G. H., J. H., and T. H. for patience love and support.

Appendices

Appendix 1: Morphological data from backcross plants

Appendix 1.1: Backcross family 1 morphology

Qualifier	Morphological score	Qualifier	Morphological score
A	4	L	1
B	4	M	3
C	3	N	4
D	3	O	1
E	3	P	2
F	3	Q	3
G	3	R	3
H	1.5	S	3
I	2	T	3
J	3	U	1
K	1	V	1.5

Table 11: Scores allocated to family 1 backcross plants.

Appendix 1.2: Morphology of other backcross plants

Backcross family number	Qualifier	Number of leaves when potted on	Number of Leaves at	Number of leaves at 6.6.01	Presence of extra leaves from groove	Number of leaves at flowering
Family 1	A	ns	ns	ns	ns	ns
	C	ns	ns	ns	ns	ns
	E	ns	ns	ns	ns	ns
	F	ns	ns	ns	ns	ns
	G	ns	ns	ns	ns	ns
	H	ns	ns	ns	ns	ns
	J	ns	ns	ns	ns	ns
	K	ns	ns	ns	ns	ns
	M	ns	ns	ns	ns	ns
	P	ns	ns	ns	ns	ns
	Q	ns	ns	ns	ns	ns
	R	ns	ns	ns	ns	ns
	S	ns	ns	ns	ns	ns
	T	ns	ns	ns	ns	ns
	U	ns	ns	ns	ns	ns
	V	ns	ns	ns	ns	ns
Family 2	B	2	3	3	absent	3
	C	1	2	3	present	3
	D	1	2	2	present	2
	E	1	2	2	present	2
	F	1	3	4	absent	4
	G	1	3	3	absent	3
	H	2	4	4	present	4
	I	2	3	6	absent	6
Family 3	B	1	3	3	present	3
	C	2	4	4	present	4
	D	2	5	10	absent	5
	F	1	3	3	present	3
	H	1	ns	6 (not	not flowered	not

				flowered)		flowered
	I	2	4	4	present	4
	J	1	2	2	present	2
	K	1	2	2	present	2
	L	1	ns	2	present	2
	M	1	3	3	present	3
Family 4	A	1	ns	4	not flowered	4
	B	1	2	2	present	2
	C	2	3	dead	ns	3
	D	1	2	3	present	3
	E	2	2	3	absent	3
	F	1+1	4	4	present	4
	G	1	1	2	absent	2
	H	2	3	4	present	4
	I	1	1	1	present	1
	J	1	1	1	present	1
	K	1	3	3	present	3
Family 5	B	1	ns	dead	dead	1
	D	2	3	3	present	3
	E	1	3	17(not flowered)	not flowered	not flowered
	F	1	2	2	present	2
	G	1	2	not flowered	not flowered	not flowered
	J	2	2	5	present	5
	K	2	3	3	present	3
	N	1+1	3	3	present	3
	O	2	2	3	absent	3
	S	2	4	4	present	4
Family 6	B	2	2	2	present	2
	C	1	2	2	present	2
	D	1	2	2	present	2
Family 7	A	1	6	6	present	6
	C	2	3	2	present	2
	D	2	2	2	present	2
	E	1	1	1	present	1
	F	1	1	1	present	1
	H	1+1	4	2	present	2

	J	2	1	dead	dead	1
	K	1	3	2	present	2
Family 8	A	2+2	6	7	present	7
	C	2	3	3	present	3
	D	2	2	4	absent	4
	E	3	4	4	absent	4
	F	1	2	2	present	2
	G	1	2	3	present	3
	H	4	ns	4	present	4
	J	1	2	2	present	2
	M	1	1	1	present	1
	N	1	2	2	present	2
	O	2	2	2	present	2
	Q	2	3	3	present	3
Family 9	A	1	1	1	present	1
	B	2	3	3	present	3
	C	1	2	2	present	2
	D	3	13	not flowered	not flowered	not flowered
	F	1+1	4	4	present	4
	G	1	2	2	present	2
	H	1	ns	1	absent	1
	I	2	ns	3	not flowered	3
	J	1	2	2	absent	2
	L	2	3	7	absent	7
	M	2	2	3	absent	3
	N	2	1	3	absent	3
	P	2	2	3	present	3
Family 10	A	1	3	15	present	15
	B	1	2	3	present	3
	C	2	3	3	present	3
	D	1	1	dead	dead	not flowered
	E	1	3	dead	dead	not flowered
	F	2	3	3	present	3
	H	1	3	dead	dead	3
	I	1	ns	2	present	2

	J	1	2	dead	dead	2
	K	2	ns	not flowered	not flowered	2
	L	2	3	3	present	3
	M	2	1	1	absent	1
	N	2	4	4	present	4
	O	1	2	2	absent	2
	P	1	2	2	present	2
	Q	1	2	2	present	2
	R	4	3	3	present	3
	S	1	2	2	present	2
Family 11	A	1	2	2	present	2
	B	1	3	not flowered	not flowered	3
	C	1	2	2	present	2
	D	2	3	3	absent	3
	E	1	2	4	present	4
	F	1	3	5	absent	5
	G	2	3	3	present	3
	H	1	2	2	present	2
	I	1	2	2	present	2
	J	2	2	2	present	2
	K	1	2	2	present	2
	L	2	2	5	absent	5
Family 12	A	1	1	2	present	2
	B	1	2	not flowered	not flowered	2
	C	1	1	1	present	1
	D	2	4	4	present	4
	E	1	3	3	present	3
	G	1	1	dead	dead	1
	I	1+1	3	3	present	3
	J	2	3	dead	dead	3
	K	1	2	2	present	2
	L	1	2	2	present	2
	M	2	2	3	present	3

Table 12: Number of phyllomorphs of backcross plants scored at different times

Appendix 1.3: Leaf gnarling phenotypes in backcross plants

Backcross family number	Qualifier	Genotype	Degree of leaf gnarling	Backcross family number	Qualifier	Genotype	Degree of leaf gnarling
Family 2	B	w/r	2		K	w/w	1
	C	w/w	2	Family 8	A	w/r	2
	D	w/w	1		C	w/w	2
	E	w/w	2		D	w/w	3
	F	w/r	3		E	w/w	2
	H	w/w	2		F	w/w	1
	I	w/r	3		G	w/r	2
Family 3	B	w/r	1		J	w/r	2
	D	w/w	2		M	w/w	2
	F	w/r	2		N	w/r	2
	H	w/r	4		O	w/r	2
	I	w/w	2		Q	w/r	2
	J	w/r	4	Family 9	A	w/r	3
	K	w/w	2		B	w/r	3
	L	w/w	3		C	w/r	3
	M	w/w	2		F	w/r	3
Family 4	A	w/w	3		I	w/w	2
	B	w/w	2		L	w/r	3
	D	w/w	2		P	w/w	2
	E	w/r	4	Family 10	C	w/r	2
	G	w/r	3		L	w/r	2
	H	w/r	1		N	w/w	2
	J	w/r	2		O	w/w	2
	K	w/r	2		R	w/w	2
Family 5	D	w/w	2	Family 11	A	w/w	1
	E	w/r	3		E	w/r	3
	F	w/r	2		F	w/r	3
	J	w/w	4		G	w/w	1
	K	w/w	2		I	w/w	2
	N	w/r	2		K	w/w	2
	O	w/r	4		L	w/r	2
	S	w/r	2	Family 12	A	w/r	2
Family 6	B	w/w	2		C	w/w	2
	C	w/r	2		E	w/w	2
	D	w/r	2		I	w/w	2
Family 7	A	w/r	2		K	w/w	3
	F	w/r	2				

Table 13: leaf gnarling phenotypes in backcross plants.

Appendix 2: *Sknox* sequences

Appendix 2.1: cDNA sequences of *Streptocarpus STM*-like genes obtained by 3'RACE

Appendix 2.1A: Ssknox1 *3' region*

CAGAAGCAGA TAAACAACTG GTTCATTAAC CAAAGGAAGC GCCATTGGAA GCCATCCGAG
GATATGCAGT TCGTGGTTAT GGATGCTGCT CATCCTCACT ACTACATGGA CAATATCATG
GCCACTCCAT TCCCAATGGA TATTTCGTCG TCGTTTCTTT AATTCGATGA TCGATGCTGG
TAACTGTATT ATATAGCATG CTAGTGAGTT TATCAACTGC TGTTGAAATT TTACTTTATT
TCATGATTAT TAATGCAAAA AAAAAAAA

*Appendix 2.1B:*Srknox1 *3' region*

TCTCAGAAGC TAGCCCTAGC AGAATCAACG GGCCTTGACC AGAAGCAAAT AAATAACTGG
TTTATCAACC AAAGGAAGCG CCATTGGAAG CCGTCCGAGG ATATGCAGTT CGTGGTTATG
GATGCCACTC ATCCTCACTA CTACATGGAC AATTTCATGG GCACTCCATT CCCAATGGAT
ATTTCGCCCT CGTTTCTTTG ATTCGATCAT CGAGTCTGGT AACTTTTATA TATATAGCAT
TCATGGGAGT GAATCTATCA ACTGCTACTG AAATTTTACT TATTCCATGA CAATTAATTT
GTCTGCCTGT AACTGTTCTA AGATTCTAGT ACAACATATT GAACGCTCGA AAATTAATTA
ATTATTATGT GTGATTTTAT TTTCAAAAAA AA

Appendix 2.1C: Sdknox1 *3' region (a genomic contaminant from 3' RACE)*

GAATCTCAGA AGCTAGCCCT AGCCGAATCA ACGGGCCTTG ACCAGAAGCA GATAAATAAC
TGGTTTATTA ACCAAAGGAA GCGCCATTGG AAGCCGTCCG AGGATATGCA GTTCGTGGTT
ATGGATGCCA CTCATCCTCA CTACTACATG GACAACTTCA TTGGCATTCC ATTCCCAATG
GATATCTCGC CGTCGTTTCT TTAATTCGAT TATCGATTCT GGTAACTTTG ATATATAGCA
TTCATGATAG GTGAATCTAT CAATTGCTAC TGAAATTTTA CTTAATTTCA TGAGAATTAA
TTTGTCTGCC TGTAACTGTT CTAAGATTCT AGTACAACAT ATTGAACGCT TCGATAATTA
ACTA

*Appendix 2.1D:*Sdknox2 *3'region*

CCTCACTACT ACATGGACAA CTTCATGGGC ACTCCATTCC CAATGGATAT CTCGCCGTCG
TTTCTTTAAT TCGATTATCG ATTCTGGTAG TGAATCTATC AACTGCTACT GAAATTTTAC
TTATTTCATG ATAATTAATT TGTCTGCCTG TAACTGTTCT AAGATTCTAG TACAACATAT
TGAACGCTCG AAAATTAATT AATTATTATG TGTGATTTTA TTTTCAATTT CTTGTTTAGT
TCCTTGTAAT GCATTAACAT CGAAACCTAA ATATTAATGT TTATGTCTGA AAAAAAAAAA
AAAA

Appendix 2.2:Partial genomic products obtained by inverse PCR

Appendix 2.2A: Sdknox1 *sequences*

i. Sequence of *Sdknox1* from the end of intron 2 to middle of the homeodomain. Coding sequence is indicated in upper case, intron sequence is indicated in lower case

agcttcttat tctatatttg atgcattaaa tataattcca agtaatgata gctcggggtg
acattagtat aatcatgtgg gaaaatattt tggcttggac tcggctgaga ccactcgttt
tatgtcgcag CTTGTGGCGA AGCAGTAATG GAACGAAATG GTTCATCCGA AGAAGAATTT
GATGTCAACA ACAGTTTCAT TGACCCCCAA GCAGAAGATC GGGAACTGAA AGGTCAGCTC
TTGCGAAGAT ACAGTGGTTA CTTGGGCAAC CTCAAACAAG AATTCATGAA GAAGCGAAAG
AAAGGCAAGC TGCCTAAAGA AGCAAGGCAA CAGTTGCTCG ACTGGTGGAG CCGACATTAC
AAATGGCCTT ATCCATCTgt atggatatct atatatgtct tacatttgca tgcattttac
ttgactatac tatactcttg atcaactatt gctaatctag GAATCTCAGA AGCTAGCCCT
AGCCGAATCA ACGGGCCTTG AC

ii. *Sdknox1* 3' coding and untranslated sequence. Coding sequence is indicated in upper case, Untranslated sequence is indicated in lower case

GCAGTTCGTG GTTATGGATG CCACTCATCC TCACTACTAC ATGGACAACT TCATTGGCAC
TCCATTCCCA ATGGATATCT CGCCGTCGTT TCTTTAAttc gattatcgat tctggtaact
ttgatatata gcattcatga tagtgaatct atcaactgct actgaaattt tacttaattt
catgataatt aatttgtctg cctgtaactg ttctaagatt ctagtacaac gtattgaacg
ctcgaaaatt aattaattat tatgtgtgat tttattttca atttcttgtt tagttccttg
taatgcatta acatcgaaac ctaaatatta atgtttatgt ctcaattcgt catgtttgat
tctgataaac tataatatat taatttgttg aatctatatg atttaagaat caaacattac
ttgtaaatcc acaatgttgt atttttttct cctccaaaat atgatttttt aattga

iii. *Sdknox1* further 3' Untranslated sequence

caataaatat gaaaattgct tcacaaaatc ttaataattt tggataatta tctgaataac
ttgacaaaaa attaatagaa gaggacctaa caaaccaaga aacagcaata acaatacaaa
agatccaagg ataatgtcaa atttcaaaaa taaaaaacat taaacataaa aatattatat
aagaacattt tataatcaat attaagtgaa gaaaatatat tttctaaaaa atcaatgaaa
aaagatataa ctatttttag tattcagaat atgataaaaa gaaagccatt tgccttttgc
cattaacttc aaattatatt tgtcactttt aaattagtta ttcacaaaaa ataaaatata
taatgcatat tagttttggc aataagtaat tgtgaaaata atatttggat cgatttaaaa
atgctcaaat tgattgaaaa aatatatatg tttaaataaa aaaattggaa aatattcaga
gtaatcaatc agatgaaaat gttacaaaac tgtggattac cctcatctca ctacaa

Appendix 2.2B: Srknox1 *sequences*

i. Sequence of *Srknox1* from the end of intron 2 to middle of the homeodomain. Coding sequence is indicated in upper case, intron sequence is indicated in lower case

```
agcttcttat tctatatttg atgcattgaa tataattctg agtaatgata gctcggggtg
acattagtat aatcatgtgg gaaagtattt tggcttggac tcggctgaga tcactcgttt
tatgtcgcag CTTGTGGCGA AGCAGTACTG GAACGAAATG GTTCATCCGA AGAAGAATTT
GATGTCAACA ACAGTTTCAT TGATCCCCAA GCAGAAGATC GCGAACTGAA AGGTCAGCTC
TTGCGAAGAT ACAGTGGATA CTTGGGCAAC CTCAAACAAG AATTCATGAA GAAGCGAAAG
AAAGGCAAGC TGCCTAAAGA AGCAAGGCAA CAGTTGCTCG ACTGGTGGAG CCGACATTAC
AAATGGCCTT ATCCATCTgt atgtatatct acacaccccc ctgtatatgt cttacctatg
catgcatttt actttactat attgctaatc tagGAATCTC AGAAGCTAGC CCTAGCAGAA
TCAACGGGCC TTGAC
```

ii. *Srknox1* 3' coding and untranslated sequence. Coding sequence is indicated in upper case, Untranslated sequence is indicated in lower case

```
GCAGTTCGTG GTTATGGATG CCACTCATCC TCACTACTAC ATGGACAATT TCATGGGCAC
TCCATTCCCA ATGGATATTT CGCCCTCGTT TCTTTGATTC GATCATCGAG TCTGGTAAct
tttatatata tagcattcat gggagtgaat ctatcaactg ctactgaaat tttacttatt
ccatgacaat taatttgtct gcctgtaact gttctaagat tctagtacaa catattgaac
gctcgaaaat taattaatta ttatgtgtga ttttatttтc aatttcttgt ttagttcctg
ttcatgcatt aacattgaga cctaaatgtt aatgtttatg tctcaattcc tgtagttttt
catgataaac tataatatat taatttgttg aatctatatg ttttaacaga atcacacatt
actttgtaaa tccacaatgg agtttattta ctaaacaaac cctttatcca tatttagtgg
gagatacaaa agtttctaga aggctctgat tttacaaaaa tatgcaacta tcacccтttt
ttaattaatt ttattaaata tttatttgat aggttatcca ataaaaatga aaattgctca
caagatttaa ttagctaaat aaatattttt ggataactat ccgaagaact tgacaaaaat
taatttagga ttaagttaaa agaagaggac ctcacaaaac aagaaacaat aataacaata
caaaagaata aggataatgt caaatttтaa aattttctca actaaaaaac attaaaaata
aaaatattat ataagaaaca gatcattact aaaatttata atcaatatta agtgtagaaa
acataactat ttttagtatc agaatatgat aaaaaaaagc catttacctt ttgccattaa
cttttcaaat tatatttgtc acttttaaat tagttattca caaaaaaata aaatatataa
tgcatatgtt ggcctcatat agcagagtag ttttggcaat aagtaattgt gaataatatt
ttgataaaat gctcaaattg agagatgaaa ccttttatat gaaatcttca taataatgtt
aacaattgaa aaaatatata tttttaaata taaaattgga aaatattcag agtaatcaat
caaatgaaaa tgttacaaac tgtggattac cctcatctca ctacaa
```

Appendix 2.2C: Swknox1 *sequences*

i. Sequence of *Swknox1* from the end of intron 2 to middle of the homeodomain. Coding sequence is indicated in upper case, intron sequence is indicated in lower case

```
taaatatact tctatgtaat gatagctcgg ggtgacatta gtacaatcat gtgggaaagt
attttggctt ggactcggct gagaccactc gttttatgtc gcagCTTGTG GCGAAGCAGC
ACCGGAACGA AATGGTTCAT CCGAAGAAGA ATTTGATGTC AACAACAGTT TCATTGATCC
CCAAGCAGAA GATCGCGAAC TGAAAGGTCA GCTCTTGCGA AGATACAGTG GATACTTGGG
CAACCTCAAA CAAGAATTCA TGAAGAAGCG AAAGAAAGGC AAGCTGCCTA AAGAAGCAAG
GCAGCAGTTG CTCGACTGGT GGAGCCGACA TTACAAATGG CCTTATCCAT CTgtatggat
atcttgaccc ccccacctat atatgtctta cctttgcatg cattttactt tactacattg
ctaatctagG AATCTCAGAA GCTAGCCCTA GCCGAATCAA CGGGCCTTGA CCAGAAGCAA
ATAAATAACT G
```

ii. *Swknox1* 3' coding and untranslated sequence. Coding sequence is indicated in upper case, Untranslated sequence is indicated in lower case

GCAGTTCGTG GTTATGGATG CCACTCATCC TCACTACTAC ATGGACAATT TCATGGGCAC
TCCATTCCCA ATGGATGTTT CGCCGTCGTT TCTTTAATTC GATCATCGAG TCCGGTAAct
tttatatata gcattcatgg tagtgaatct atcaactgct actgaaattt tacttattct
gtaactgttc taagattcta gtacaacata ttgaacgctc gaaaattaat taattattat
gtgtgatttt attttcaatt tcttgttaat gcatgaacat cgaaacctaa atattaatgt
ttatgtctca attcctctag tttttcatgc gtgcagcttg ataaactgta atatattaat
ttgttgaatc tatatgtttt aacagaatca aacattactt gtaaatccac aattttgtat
ttttttccct ctaaaatatg tatttttaat cgatttataa atatcctcta agaattaaag
tgcaacataa aaaaatgtaa attgttttat aatctccttt ttaattcata tttcttagat
ttcaattatt ttttctgagt tgaagttgat ttactaaa

Appendix 2.3: Partial cDNA products obtained by RT-PCR

Sdknox1

GGTTCTGCAG GCAACGTGAA CATGACTTCT TCTTTCAAGG GAGCTAATTC CTACATGGGT
TTTGGAGATA ACGTTAATGG GCTTTGTCCC ATGATGATGA TGATGCCTGC AAATGACCCT
AATGGGGACT GCAGCCAGCC CCTATTTCTG CCTCTTCCAG CCGCTAACCA GCAAGGAATC
AATCGCAACA GCAGCTCCGC CGCCGCCTGC GGCGGTTCTA TGATGCCTGA ACATCAGAGT
AACACGAGTA CTGGGTATTA TTTCATGGAA GGAGACGGAG TTGCCGGCGG CGGCTCTGTC
AAGTCGAAGA TCATGGCTCA TCCGCACTAC CCTCGCCTCT TGGCTGCCTA TGTTAATTGC
CAAAAGATAG GAGCGCCGCC TGAAGTGGTG GCAAAGCTGG AGGAAGCTTG CGCGTCCACC
ATCACAATAG GCGGCCGCAA CGAAAGGAGC TGTGTCGGGG AAGATCCGGC ACTGGACCAG
TTCATGGAGG CATACTGTGA AATGCTGACA AAGTATGAGC AAGAATTGTC GAAACCCTTC
AAAGAAGCTA TGCTCTTCCT TTCGAGGATT GAGTGCCAGT TCAAAGCCCT TACATTGTCT
CACTCTTCTG ATTCTGGAGC TTGTGGCGAA GCAGTAATGG AACGAAATGG TTCATCCGAA
GAAGAATTTG ATGTCAACA

Srknox1

GGTTCTGCAG GCAACATGAA CATGACTTCT TCTTTCAAGG GAGCTAATTC CTACTTGGGT
TTTGGAGATA ACGTTAATGG GTTTTGTCCC ATGATGATGA TGATGCCTGC AAATAACCCT
AATAGGGACT GCAGCCAGCC CATATTTCAG CCTCTTCCAG CCGCTAACCA GCAAGGCATC
AATCGCAACA GCAGCTCCGC CGCCGCCTGC GGCGGTTCTA TGATGCCTGA ACATCAGAGT
AACACGAGTA CTGGGTATTA TTTCATGGAA GGAGACGGAG ATGCCGGCGG CGGCTCTGTC
AAGTCGAAGA TCATGGCTCA TCCGCACTAC CCTCGCCTCT TGGCTGCCTA TGTTAATTGC
CAAAAGATAG GAGCGCCGCC TGAAGTTGTG GCAAAGCTGG AGGAAGCTTG CGCGTCCACC
ATCACAATAG GCGGCCGCAA CGAAAGGAGC TGTGTCGGGG AAGATCCGGC ACTGGACCAG
TTCATGGAGG CATACTGTGA AATGCTGACA AAGTATGAGC AAGAATTGTC GAAACCCTTC
AAAGAAGCTA TGCTCTTCCT TTCGAGGATT GAGTGCCAGT TCAAAGCCCT TACATTGTCT
CACTCTTCTG ATTCTGGAGC TTGTGGCGAA GCAGTACTGG AACGAAATGG TTCATCCGAA
GAAGAATTTG ATGTCAACA

Appendix 2.4: Deduced amino acid sequences of *Streptocarpus* homologues of *STM*

Appendix 2.4A: Sdknox1

EGSAGNVNMTSSFKGANSYMGFGDNVNGLCPMMMMMPANDPNGDCSQPLFLPLPAANQQGIN
RNSSSAAACGGSMMPEHQSNTSTGYYFMEGDGVAGGGSVKSKIMAHPHYPRLLAAYVNCQKIG
APPEVVAKLEEACASTITIGGRNERSCVGEDPALDQFMEAYCEMLTKYEQELSKPFKEAMLFLSRI
ECQFKALTLSHSSDSGACGEAVMERNGSSEEEFDVNNSFIDPQAEDRELKGQLLRRYSGYLGNLK
QEFMKKRKKGKLPKEARQQLLDWWSRHYKWPYPSESQKLALAESTGLDQKQINNWFINQRKR
HWKPSEDMQFVVMDATHPHYYMDNFIGIPFPMDISPSFL

Appendix 2.4B: Srknox1

EGSAGNMNMTSSFKGANSYLGFGDNVNGFCPMMMMMPANNPNRDCSQPIFQPLPAANQQGINR
NSSSAAACGGSMMPEHQSNTSTGYYFMEGDGDAGGGSVKSKIMAHPHYPRLLAAYVNCQKIGA
PPEVVAKLEEACASTITIGGRNERSCVGEDPALDQFMEAYCEMLTKYEQELSKPFKEAMLFLSRIE
CQFKALTLSHSSDSGACGEAVLERNGSSEEEFDVNNSFIDPQAEDRELKGQLLRRYSGYLGNLKQ
EFMKKRKKGKLPKEARQQLLDWWSRHYKWPYPSESQKLALAESTGLDQKQINNWFINQRKRH
WKPSEDMQFVVMDATHPHYYMDNFMGTPFPMDISPSFL

Appendix 2.4C: SSknox1

MEGSAGNLNLNSYRGANSFMSFGDNVNGFSPMMMMMMHANDPNGDCSQPIFLPLPAANHQGINR
NTGGGGGLTPEHHSNMSTGYYFMEGDGDAGGSSVKSKIMAHPHYPRLLAAYVSCQKIGAPPEV
VAKLEEVCASATSTGCRNERSFVGEDPALDQFMEAYCEMLTKYEQELSKPFKDAMLFLSRFECQ
FKALTLSHSADSGACDEAVLERNGSSEEEFDVNNSFIDPQAEDHELKGQLLRKYSGYLGNLKQEF
MKKRKKGKLPKEARQQLLDWWSRHYKWPYPSESQKLALAESTGLEQKQINNWFINQRKRHWK
PSEDMQFVVMDAAHPHYYMDNIMATPFPMDISSSFL

Appendix 2.5:Ribosomal cDNA sequences amplified by 5' RACE

Appendix 2.5A: Streptocarpus rexii *5.8s ribosomal subunit cDNA*

GACCGTANCC GGGTGCGATC GGCACCAAGG AACAATGAAC TCGGAAGCNG AGGGCCCCTN
GACGTGCNCG GGGGGAGCCC AATGCGTCGG AGACTCCTCG AAATCAAATG ACTNTTCGGC
AATGGATATC TCGGCTCTTG CATCGATGAA GAACGTAGTG AAATGCGATA CTTGGTGTGA
ATTGCAGAAT CTCGTGAACC ATTGAGTCTT TGAACGCAAG TTGTGCCCGA GGCCTTGTGG
CCGAGGGCAC GCCTGCTTGG GCGTCATGGC ATCGTCGCTT TCGCTCCACG CATTGTTGGT
GGCGAGCGGA AATTGGCCCC GTGTGTCCTC GGGCACAGTC GGTCGAAGAG TGGGTAGTCG
GCAGTCGTCG GGCACGATGG GTGTTGGTCG CCGCGAGCGG GAACAGAACG TCGTCCCCGT
CGTCTCGGGA TGAGTCCTCA AGAGACCCTG TGCGAATGCG GCGTCGGCCG AAAGNGCCGC
GCCCATCAAA TTGTGGCCCC AAGTCAGGCG AGGCCACCC

Appendix 2.5B: Streptocarpus rexii *40s ribosomal subunit cDNA*

CTGGTCGACA CACCCCTGGT ACTTTCACCA ATCAGCTTCA GACTTCATTC AGTGAACCTA
GGCTTTTGAT TCTTACCGAC CCAAGAACTG ATCATCAGCC TATCAAAGAA GCTGCTCTCT
ATAACATCCC CACCATAGCT TTCTGCGACA CTGACTCTCC AATGCGTCAT GTCGATATTG
GTATCCCTGC CAATAACAAG GGAAAGCACA GCATTGGCTG TCTTTTTTGG CTCTTGGCTA
GGATGGTCCT CCAGATGCGT GGTGCAGTTG CACCTGGTCA CAAATGGGAT GTCATGGTTG
ATCTATTTTC TACCGGGAGC CTGAGGAATC TAAGGAACAA CAGGAGGAGG AAGCTCCAGC
TCTCCCTGAT TATGCTGATT ATTCAGCTGC TGCAGCTGGA GCTGATTGGT CCAGTGGCCA
GATTCCAGAT GCCCAATGGA CTCCTGATAT GGTTGGTACC GTACCACCAA GTTGCAGGTG
GTTGGAATCC CGATGCAGGT GCAGGTGATG GACGGGATGC AGCTGCTGCC CCGCCCATGA
CAGCTCCTCC AGCTGTCGAC GTAATGCCTG GTGTTGTTCC CAGTGGCTGG GAGTAGAAAC
AGATCATGTA ACAACAGCG

Appendix 2.6: Record of sequences amplified from *Streptocarpus*

Species	GenBank code	Homology by BLAST
S. dunnii	AY061852	maturase k
S. dunnii	BM34618	none
S. dunnii	BM34619	none
S. dunnii	BM34620	extensin
S. rexii	AY061837	none
S. rexii	AY061839	subtilisin
S. rexii	AY061840	transmembrane protein
S. rexii	BM34622	none
S. rexii	BM34623	cytochrome oxidase
S. rexii	BM34624	DNA directed RNA polymerase
S. rexii	not accepted by	5.8s ribosomal subunit
S. rexii	not accepted by	40s ribosomal subunit
S. rexii	BM34625	sulphated glycoprotein
S. saxorum	AY061841	cellobiosidase
S. saxorum	AY061843	folate binding protein
S. saxorum	AY061844	HMG domain gene
S. saxorum	AY061845	HMG domain gene (contiguous with
S. saxorum	AY061846	hypothetical protein 1
S. saxorum	AY061847	hypothetical protein 2
S. saxorum	AY061848	hypothetical protein 3
S. saxorum	BM34636	transcription activator
S. saxorum	BM34627	none
S. saxorum	BM34628	none
S. saxorum	BM34629	none
S. saxorum	BM34630	none
S. saxorum	BM34631	none
S. saxorum	BM34632	none
S. saxorum	BM34633	none
S. saxorum	BM34626	annexin
S. wittei	AY061851	NAD+ synthetase

Table 14: Accession numbers of fragments amplified from different *Streptocarpus* species.

Appendix 3: *Sknox1* genotypes for backcross plants

Backcross family number	Qualifier	Genotype	Backcross family number	Qualifier	Genotype
Family 1	A	d/d	Family 8	A	w/r
	C	d/r		C	w/w
	E	d/d		D	w/w
	F	d/r		E	w/w
	G	d/d		F	w/w
	H	d/d		G	w/r
	J	d/r		H	w/r
	K	d/r		J	w/r
	M	d/r		M	w/w
	P	d/r		N	w/r
	Q	d/d		O	w/r
	R	d/r		Q	w/r
	S	d/d	Family 9	A	w/r
	T	d/r		B	w/r
	U	d/d		C	w/r
	V	d/d		D	w/r
Family 2	B	w/r		F	w/r
	C	w/w		G	w/w
	D	w/w		H	w/r
	E	w/w		I	w/w
	F	w/r		J	w/r
	G	w/r		L	w/r
	H	w/w		M	w/w
	I	w/r		N	w/w
Family 3	B	w/r		P	w/w
	C	w/r	Family 10	A	w/w
	D	w/w		B	w/r
	F	w/r		C	w/r
	H	w/r		D	w/r
	I	w/w		E	w/r
	J	w/r		F	w/w
	K	w/w		H	w/w

	L	w/w		I	w/r
	M	w/w		J	w/r
Family 4	A	w/w		K	w/w
	B	w/w		L	w/r
	C	w/r		M	w/r
	D	w/w		N	w/w
	E	w/r		O	w/w
	F	w/r		P	w/r
	G	w/r		Q	w/r
	H	w/r		R	w/w
	I	w/w		S	w/w
	J	w/r	Family 11	A	w/w
	K	w/r		B	w/w
Family 5	B	w/w		C	w/r
	D	w/w		D	w/r
	E	w/r		E	w/r
	F	w/r		F	w/r
	G	w/w		G	w/w
	J	w/w		H	w/w
	K	w/w		I	w/w
	N	w/r		J	w/w
	O	w/r		K	w/w
	S	w/r		L	w/r
Family 6	B	w/w	Family 12	A	w/r
	C	w/r		B	w/r
	D	w/r		C	w/w
Family 7	A	w/r		D	w/w
	C	w/w		E	w/w
	D	w/w		G	w/w
	E	w/r		I	w/w
	F	w/r		J	w/r
	H	w/w		K	w/w
	J	w/w		L	w/w
	K	w/w		M	w/r

Table 15: Genetic scores for backcross plants.

Appendix 4: Intron 3 sequence data

Appendix 4.1: *Sknox* gene intron 3 sequences

S. beampingaratrensis A (L=96 bp, %UT=63)

```
CCTTATCCAT CTgtatggat atct?gacat cccctagata cgtcttacat
ttgtatggat atctatcctc tcttgatcaa cttttgaaat ccgtggcaac
tattgtagGA ATCTCAGAAG CTAGCCCTAG C?GAATCGAC AGGCCTGGAC
CAG
```

S. burundianus A (L=97 bp, %UT=67)

```
CCTTATCCAT CTgtatggat atctatacac ccccccctata tatgtcttac
atttgcatgc attttacttt actatacact atactcttga tcaactattg
ctaatgtagG AATCTCAGAA GCTTGCCCTA GCCGAATCAA CGGGCCTTGA
CCAG
```

S. burundianus B (L=90 bp %UT=66)

```
CCTTATCCAT CTgtatgaat atctatacac cccttgtgta tatatgcatg
cattttacta cactcctgat caagtattga acttttgtgg cgattaatgt
agGAGTCTCA GAAGCTAGCC CTAGCTGAAT CGACAGGTCT GGACCAG
```

S. dunnii A (L=82 bp, %UT=70)

```
CCTTATCCAT CTgtatggat atctatatat gtcttacatt tgcatgcatt
ttacttgact atactatact cttgatcaac tattgctaat ctagGAATCT
CAGAAGCTAG CCCTAGCCGA ATCAACGGGC CTTGACCAG
```

S. dunnii B (L=83bp, %UT=63)

```
CCTTATCCAT CTgtatggat atctatactt cttacatttg tattccttct
attctctctt gatcaactat tgaaacttcg tggcgattat tgtagGAATC
TCAGAAGCTT GCCCTAGCCG AATCGACAGG CCTGGACCAG
```

S. dunnii C (L=96 bp, %UT=63)

```
CCTTATCCAT CTgtatgcat atctatacac cccctagata cgtctaacat
ctgcatggat agctattctc tcttgatcaa cttttgaaat tcgtggtaac
tattgtagGA ATCTCAGAAG CTAGCCCTAG CCGAATCGAC CGGCCTGGAC
CAG
```

S. hirticapsa B (L=88 bp, %UT=65)

```
CCTTATCCAT CTgtatgaat atctatacat cccttgtgta tatgcatgca
ttttactaca ctcttgatca actattgaat atttgtggcg attaatgtag
GAGTCTCAGA AGCTAGCCCT AGCTGAATCG ACAGGCCTGG ACCAG
```

S. hirticapsa A (L=75 bp, %UT=65)

CCTTATCCAT CTgtatgtat atctacacac ccccctgtat atgtcttacc
tatgcatgca ttttacttta ctatattgct aatctagGAA TCTCAGAAGC
TAGCCCTAGC AGAATCAACG GGCCTTGACC AG

S. ibityensis A (L=110 bp, %UT=63)

CCTTATCCAT CAgtatggat atctatacac ccccccatat gtgtcttaca
tttgcatgca ttttactttt ctatactaca cgatcttgat caactattga
catttcgtgg cgactaatgt agGAATCTCA GAAGCTAGCC CTAGCCGAAT
CAACGGGCCT TGACCAG

S. ibityensis B (L=109 bp %UT=65)

CCTTATCCAT CAgtatggat atctatacac cccccatata tgtcttacat
ttgcatgcat tttacttttc tatactatac gatcttgatc aactattgac
atttcgtggc gactaatgta gGAATCTCAG AAGCTAGCCC TAGCCGAATC
AACGGGCCTT GACCAG

S. ibityensis C (L=79 bp %UT=68)

CCTTATCCAT CTgtatgata tctatacacc cccttatgtt atatgcatgc
attttactag acactcttga tcaactattg aacttttgtr gGAATCTCAG
AAGCTAGCCC TAGCCGAATC GACAGGCCTG GACCAG

Sa.ionantha A (L=83 bp %UT=66)

CCTTATCCAT CTgtatggat atctatactt cttrcrtttg tattccttct
attctctctt gatcaactat tgaaacttcg tggcgactat tgtagGAATC
TCAGAAGCTT GCCCTAGCCG AATCGACAGG CCTGGACCAG

S. itremensis A (L=87 bp %UT=67)

CCTTATCCAT CTgtatgata tctatacacc ccttatgtat atgcatgcat
tttactccac tctttatcaa ctattgaact tycgtggcga ttaatgtagG
AATCTCAGAA GCTAGCCCTA GCCGAATCGA CAGGCCTGGA CCAG

S.levis A (L=94 bp %UT=63)

CCTTATCCAT CTgtatggat atct??acac cccctagata cgtctaacat
ctgtatgga? atcttctctc ttgatcaact tttgaaattc gtggcaacta
ttgtagGAAT CTCAGAAG?? ?????????? ?????????? ??????????
?

S. modestus A (L=96 bp %UT=62)

CCTTATCCAT Ctgtatgaat atctatacac cccctagata cgtctaacat
ctgcatggat agctattctc tcttgatcaa cttttgaaat tcgtggtaac
tattgcagGA ATCTCAGAAG CTAGCCCTAG CCGAATCGAC CGGCCTGGAC
CAG

S. modestus B (L=82 bp %UT=66)

CCTTATCCAT ctgtatgaat atctatccac cccttgtgtg cattttacta
catactcttg atcaagcatt gaatttttgt ggcgattaat gtagGAGTCT
CAGAAGCTAG CCCTAGCTGA ATCGACAGGC CTGGACCAG

S. modestus C (L=90 bp %UT=66)

CCTTATCCAT CTgtatgaat atctatccac cccttgtgta tatgcatgca
ttttactaca cactcttgat caagtattga atttttgtgg cgattaatgt
agGAGTCTCA GAAGCTAGCC CTAGCTGAAT CGACAGGCCT GGACCAG

S. pallidiflorus B (L=94 bp %UT=67)

CCTTATCCAT CTgtatgcat ggatatgtct atacttctta catttgtatg
catgtctata tatccgctct tgatcaacct ttgaaacctc gtggcgacta
ttgtagGAAT CTCAGAAGCT TGCCCTTGCC GAATCAACTG GTCTGGACCA
G

S. pallidiflorus A (L=657 bp %UT=67)

CCTTATCCAT CTgtatgata catttgcatg cattctactt tactacactc
tatactcctg atcaactatt gaatttcatg gcgactaatg tcggaaactc
ataatctttc ttacatttgc atgcagttaa ctttactaaa ctctatattc
ttgatcaact attgaatttt gtggcgacta atataggaat ctcagaagct
atcctacatt tgcattcatt ttactttact acactacact atactcctgg
tcaactattg aagtttcgta acgaataatg taggaatctc agaagctata
tatctttcat ttgcatgcag tttacttcac tacagtctat actcttcatc
aagtattgaa atttcgtggc gactaatgca ggaatctcag aagctatctt
aaatttgcat gcattttact ttactaccct acatactctt gatcaactat
tgaatttcag ggcaactaat gtaggaatct cagaagtatc ttacatttgc
atgcagttta ctttactaca ctctatacac ttgatcaatt attattgaaa
tatcgtggcg aataatgtag gaatctcaga agctatctta cgtttgcatg
catttgactt tattacacta cactctatac tcttgaccaa ctattgaagg
ttcttggcta ctaatgcagG AATCTCAGAA GCTTGCCCTA GCCGAATCCA
CTGGCCTTGA TCAG

S. papangae B (L=88 bp %UT=66)

CCTTATCCAT CTgtatgcat atctatacac cccttatgta tatgcatgca
ttttactaca ctcttgatca actattgaac ttttgtggcg attaatgtag
GAATCTCAGA AGCTAGCCCT AGCCGAATCG ACAGGCCTGG ACCAG

S. papangae A (L=128 bp %UT=61)

CCTTATCCAT CTgtatggat atctatacac tccccccccc ccccccccta
aaatatatat gtcttacatt tgcatgcatt ttacttttct atactatacg
atcttgatca actactgaca tttcgtggct actaatgtag GAATCTCAGA
AGCTAGCCCT AGCCGAATCA ACGGGCCTTG ACCAG

S. pentherianus A (L=95 bp %UT=66)

CCTTAWCCAT YTgyatgaat atctatctat acacccccttg tgtatatgca
tgcattttac aacacactca tgatcaagta ttgaagtttt tgtggcgatt
aatgtacGAG TCTCAGAAGC TAGCCCTAGC TGAATCGACA GGCCTGGMCC
AG

S. primulfolius A (L=91 bp %UT=66)

CCTTATCCAT CTgtatgaat atctatccac ccctttgtgt atatgcatgc
attttactac acactcttga tcaagtattg aatttttgtg gcgatyaatg
tagGAGTCTC AGAAGCTAGC CCTAGCTGAA TCGACAGGCC TGGACCAG

S. rexii B (L=94 bp %UT=66)

CCTTATCCAT CTgtatgaat atctatctat acacccccctg tgtatatgca
tgcattttgc tacactctct tgatcaagta ttgaattttt gtggcgatta
atgtasGAGT CTCAGAAGCT AGCCCTAGCT GAATCGACAG GTCTGGACCA
G

S. rexii A (L=75 bp %UT=65)

CCTTATCCAT CTgtatgtat atctacacac ccccctgtat atgtcttacc
tatgcatgca ttttacttta ctatattgct aatctagGAA TCTCAGAAGC
TAGCCCTAGC AGAATCAACG GGCCTTGACC AG

S. saxorum B (L=94 bp %UT=65)

CCTTATCCAT CTgtatgcat ggatatgtct atacttctta catttgtatg
catgtctrta tatctgctct tgatcaacyt ttgaaacytc gtggcgacta
ttgtagGAAT CTCAGAAGCT TGCCCTWGCC GAATCAACTG GTCTGGACCA
G

S. saxorum A (L=481 bp %UT=65)

CCTTATCCAT CTgtaaggat atctatacac ccccctaccc tatatgtctt
acatttgtgt gcattttact tcactaggac tatacacttg atcaattatt
gaatttcgtg gtgaattatg taggaatctc agaagctaac ttacatttgc
atgctgttta ctttactaaa ctctaaactc ctgataaact attgaatttc
gtggcgagta atgtaggaat cttagaagat atcttacatt tgcatgcatt
ttactttact acactctata ctcttgatca actattgaag tttcgtggtg
actaatgtag gagtctccga agctatctta catttgcatc tgtttacttt
actatgctct tgatcaacta ttgaagtttc atggcgacta atgtaggaat
cccagaagct atcttgcatt tgcatgcatt ttacttaact acattacact
atactcctga tcaactattg aagtttcatg gcgacaaatg cagGAATCTC
AGAAGCTAGC CCTAGCCGAA TCAACGGGCC TTGAACAG

S. schliebenii A (L=93 bp %UT=69)

CCTTATCCAT CTgtatggat atcaatacac cccctatata tgtcttacat
ttgcatgcat tttactttac tatactatac tcttgatcaa ctattgctaa
tgtagGAATC TCAGAAGCTA GCCCTAGCCG AATCAACGGG CCTTGACCAG

S. schliebenii B (L=90 bp %UT=68)

CCTTATCCAT CTgtatgaat atctgtacac acccccttgtg tatatgcatg
cattttacta taatcttgat gaactattga atttttgtgg cgattaatgt
agGAATCTCA GAAGCTAGCC TTGGCCGAAT CGACAGGCCT GGACCAG

S. spp. A (L=97 bp %UT=63)

CCTTATCCGT CTgtatggat atctatacac cccctagata cgtctacatc
tgtatggata tctattctct cttggtcaac ttttttgaaa ttcgtggcaa
ctattgtagG AATCTCAGAA GCTAGCCCTA CCGAATCGAC AGGCCTGGAC
CAG

S. stomandrus A (L=94 bp %UT=64)

CCTTATCCAT CTgtatgcat ggatatgtct atacttctta catttgtatg
catgtctata tatccgctct tgatcaactt ttgaaacttc ctggcgacta
ttgtagGAAT CTCAGAAGCT TGCCCTAGCC GAATCAACTG GTCTGGACCA
G

S. thompsonii A (L=96 bp %UT=63)

CCTTATCCAT CTgtatgcat atctatacac cccctagata cgtctaacat
ctgcatggat agctattctc tcttgatcaa cttttgaaat tcgtggtaac
tattgtagGA ATCTCAGAAG CTAGCCCTAG CCGAATCGAC CGGCCTGGAC
CAG

Sa. tongwensis B (L=96 bp %UT=63)

CCTTATCCAT CTgtatgcat atctatacac cccctagata cgtctaacat
ctgcatggat agctattctc tcttgatcaa cttttgaaat tcgtggtaac
tattgtagGA ATCTCAGAAG CTAGCCCTAG CCGAATCGAC CGGCCTGGAC
CAG

Sa. tongwensis A (L=83 bp %UT=63)

CCTTATCCAT CTgtatggat atctatactt ctt?c?tttg tattccttct
attctctctt gatcaactat tgaaacttcg tggcgactat tgtagGAATC
TCAGAAGCTT GCCCTAGCCG AATCGACAGG CCTGGACCAG

S. thysanothus A (L=283 bp %UT=61)

CCTTATCCAT CTgtatggat atctctccca cctcccaccc tcccttccct
gagatacgtt tacatgtgtg catggatatt gatgtaactg aattagagat
tcagaagttt caaggagtct ttactggaga atctttaagt atccttactt
acgtcacgaa actcttagta gcatagcagt ctagtggatc tctattacag
ctttcggatc tctaatttgg attatctcgt ttttagtcac atctaatatc
tatcactcta gatcaagtat tgaactttcg tggcgactat ttcagGAATC
TCAGAAGCTA GCCCTAGCAG AATCAACAGG GCTGGACCAG

Sa. velutina A (L= bp %UT=)

CCTTATCCAT CTgtatggat atctatactt cttacatttg tattccatct
attctctctt gatcaactat tgaaacttcg tggcgattat tgtagGAATC
TCAGAAGCTT GCCCTAGCCG AATCGACAGG CCTGGACCAG

S. venosus A (L=83 bp %UT=68)

CCTTATCCAT CTgtatggat atctatacac cccctagata cgtctacatc
tgtatggata tctattctct cttgatcaac ttttgaaatt cgtggcaact
attgtagGAA TYTCAGAAGC TAGCCCTAGC CGAATCGACA GGCCTGGACC
AG

S. wendlandii A (L=94 bp %UT=67)

CCTTATCCAT CTgtatgaat atccatctat acacccccttg tgtatatgca
tgcattttac tacacactct tgatcaatta ttgaattttt gtggcgatta
atgtagGAGT CTCAGAAGCT AGCCCTAGCT GAATCGACAG GCCTGGACCA
G

S. wittei B (L=89 bp %UT=65)

CCTTATCCAT CTgtatgaac atctatacac cccttgtgta tatgcatgca
ttttactaca actcttgatc gagtattgaa tttttgtggc gattaatgta
gGAGTCTCAG AAGCTAGCCC TAGCTGAATC GACAGGCCTG GACCAG

S. wittei A (L=77 bp %UT=62)

CCTTATCCAT CTgtatggat atcttgaccc ccccacctat atatgtctta
cctttgcatg cattttactt tactacattg ctaatctagG AATCTCAGAA
GCTAGCCCTA GCCGAATCAA CGGGCCTTGA CCAG

Appendix 4.2: *Sknox* full intron sequence alignment

```
S. stomandrus B     CCTTATCCATCTGTATGCATGGATATGTCT--ATAC----------TTCTTACATTT--GTATGCATGTCTATATATCCGC--------
S. pallidiflorus B  CCTTATCCATCTGTATGCATGGATATGTCT--ATAC----------TTCTTACATTT--GTATGCATGTCTATATATCCGC--------
S. saxorum B        CCTTATCCATCTGTATGCATGGATATGTCT--ATAC----------TTCTTACATTT--GTATGCATGTCTRTATATCTGC--------
Sa.tongwensis A     CCTTATCCATCTGTATGGAT----AT--CT--ATAC----------TTCTTNCNTTT--GTATTCCT-TCTATTCTC------------
Sa. ionantha A      CCTTATCCATCTGTATGGAT----AT--CT--ATAC----------TTCTTRCRTTT--GTATTCCT-TCTATTCTC------------
Sa. velutina A      CCTTATCCATCTGTATGGAT----AT--CT--ATAC----------TTCTTACATTT--GTATTCCA-TCTATTCTC------------
S. dunnii B         CCTTATCCATCTGTATGGAT----AT--CT--ATAC----------TTCTTACATTT--GTATTCCT-TCTATTCTC------------
S. burundianus A    CCTTATCCATCTGTATGGAT----AT--CT--ATACACCCCCCC----------TATATATGTCTTACATTT--GCATGCATTTTACTTTACTATACACTATAC
S. dunnii A         CCTTATCCATCTGTATGGAT----AT--CT--AT----------ATATGTCTTACATTT--GCATGCATTTTACTTGACTATACTATAC--
S. wittei A         CCTTATCCATCTGTATGGAT----AT--CT--TGACCCCCCC----------ACCTATATATGTCTTACCTTT--GCATGCATTTTACTTTACTAC---------
S. rexii A          CCTTATCCATCTGTATGTAT----AT--CT--ACACACCCCCC----------TGTATATGTCTTACCTAT--GCATGCATTTTACTTTACTAT---------
S. hirticapsa A     CCTTATCCATCTGTATGTAT----AT--CT--ACACACCCCCC----------TGTATATGTCTTACCTAT--GCATGCATTTTACTTTACTAT---------
S. papangae A       CCTTATCCATCTGTATGGAT----AT--CT--ATACACTCCCCCCCCCCCCCCCCC------TAAAaTATAtATGTCTTAcATTT--GcATGCATTTTACTTTTCTATACTAtACGA
S. schliebenii A    CCTTATCCATCTGTATGGAT----AT--CA--ATACACCCCC----------TATATATGTCTTACATTT--GCATGCATTTTACTTTACTATACTATAC--
S. ibityensis A     CCTTATCCATCAGTATGGAT----AT--CT--ATACACCCCCCC----------ATATGTGTCTTACATTT--GCATGCATTTTACTTTTCTATACTACACGA
S. ibityensis B     CCTTATCCATCAGTATGGAT----AT--CT--ATACACCCCCC----------ATATATGTCTTACATTT--GCATGCATTTTACTTTTCTATACTATACGA
S. tongwensis B     CCTTATCCATCTGTATGCAT----AT--CT--ATACACCCCC----------TAGATAC----------GTCTAACATCT--GCATGGATAGCTATTCTC------------
S. dunnii C         CCTTATCCATCTGTATGCAT----AT--CT--ATACACCCCC----------TAGATAC----------GTCTAACATCT--GCATGGATAGCTATTCTC------------
S. modestus A       CCTTATCCATCTGTATGAAT----AT--CT--ATACACCCCC----------TAGATAC----------GTCTAACATCT--GCATGGAT----------AGCTATTCTC
S. thompsonnii A    CCTTATCCATCTGTATGCAT----AT--CT--ATACACCCCC----------TAGATAC----------GTCTAACATCT--GCATGGATAGCTATTCTC------------
S. beamp. A         CCTTATCCATCTGTATGGAT----AT--CT--NGACATCCCC----------TAGATAC----------GTCTTACATTT--GTATGGATATCTATCCTC------------
S. levis A          CCTTATCCATCTGTATGGAT----AT--CT--NNACACCCCC----------TAGATAC----------GTCTAACATCT--GTATGGANATC--TTCTC------------
S. venosus A        CCTTATCCATCTGTATGGAT----AT--CT--ATACACCCCC----------TAGATAC----------GTCT-ACATCT--GTATGGATATCTATTCTC------------
S. spp. 11 A        CCTTATCCGTCTGTATGGAT----AT--CT--ATACACCCCC----------TAGATAC----------GTCT-ACATCT--GTATGGATATCTATTCTC------------
S. papangae B       CCTTATCCATCTGTATGCAT----AT--CT--ATACACCCCTTA----------TGTATAT----------GCATGCATTTTAC-----TACAC-------
S. ibityensis C     CCTTATCCATCTGTATG-AT----AT--CT--ATACACCCCCT----------TATGTTATAT----------GCATGCATTTTAC---TAGACAC-------
S. itremensis A     CCTTATCCATCTGTATG-AT----AT--CT--ATACACCCC----------TTATGTATAT----------GCATGCATTTTAC---TC--CAC-------
S. schliebennii B   CCTTATCCATCTGTATGAAT----AT--CT--GTACACACCCC----------TTGTGTATAT----------GCATGCATTTTA-----CTATAA-------
S. pentherianus A   CCTTAWCCATYTGYATGAATATCTAT--CT--ATACACCCC----------TTGTGTATAT----------GCATGCATTTTAC-----AACACAC-----
S. wendlandii A     CCTTATCCATCTGTATGAATATCCAT--CT--ATACACCCC----------TTGTGTATAT----------GCATGCATTTTAC-----TACACAC-----
S. rexii B          CCTTATCCATCTGTATGAATATCTAT--CT--ATACACCCCC----------TGTGTATAT----------GCATGCATTTTGC-----TACACTC-----
S. wittei B         CCTTATCCATCTGTATGAAC----AT--CT--ATACACCCC----------TTGTGTATAT----------GCATGCATTTTAC-----TACA-AC-----
S. burundianus B    CCTTATCCATCTGTATGAATATCTAT--------ACACCCC----------TTGTGTATAT---------AT--GCATGCATTTTAC-----TACAC-------
S. hirticapsa B     CCTTATCCATCTGTATGAATATCTAT--------ACATCCC----------TTGTGTATAT----------GCATGCATTTTAC-----TACAC-------
S. modestus B       CCTTATCCATCTGTATGAATATCTAT--C------CACCCC----------TTGTGT----------GCATTTTACTACATAC----------
S. modestus C       CCTTATCCATCTGTATGAATATCTAT--C------CACCCC----------TTGTGTATAT----------GCATGCATTTTACTACACAC----------
S. primulifolius A  CCTTATCCATCTGTATGAAT----AT--CT--ATCCACCCCT----------TTGTGTATAT----------GCATGCATTTTAC-----TACACAC-----
```

S. stomandrus B	TCTTGATCAACTTTT--GAAACTT-CCTGGCGACTATTGTAGGAATCTCAGAAGCTTGCCCTAGCCGAATCAACTGGTCTGGACCAG
S. pallidiflorus B	TCTTGATCAACCTTT--GAAACCT-CGTGGCGACTATTGTAGGAATCTCAGAAGCTTGCCCtTGCCGAATCAACTGGTcTGGACCAG
S. saxorum B	TCTTGATCAACYTTT--GAAACYT-CGTGGCGACTATTGTAGGAATCTCAGAAGCTTGCCCTWGCCGAATCAACTGGTCTGGACCAG
Sa.tongwensis A	TCTTGATCAACTATT--GAAACTT-CGTGGCGACTATTGTAGGAATCTCAGAAGCTTGCCCTAGCCGAATCGACAGGCCTGGACCAG
Sa. ionantha A	TCTTGATCAACTATT--GAAACTT-CGTGGCGACTATTGTAGGAATCTCAGAAGCTTGCCCTAGCCGAATCGACAGGCCTGGACCAG
Sa. velutina A	TCTTGATCAACTATT--GAAACTT-CGTGGCGATTATTGTAGGAATCTCAGAAGCTTGCCCTAGCCGAATCGACAGGCCTGGACCAG
S. dunnii B	TCTTGATCAACTATT--GAAACTT-CGTGGCGATTATTGTAGGAATCTCAGAAGCTTGCCCTAGCCGAATCGACAGGCCTGGACCAG
S. burundianus A	TCTTGATCAACTATT--G---------------CTAATGTAGGAATCTCAGAAGCTTGCCCTAGCCGAATCAACGGGCCTTGACCAG
S. dunnii A	TCTTGATCAACTATT--G---------------CTAATCTAGGAATCTCAGAAGCTAGCCCTAGCCGAATCAACGGGCCTTGACCAG
S. wittei A	------------ATT--G---------------CTAATCTAGGAATCTCAGAAGCTAGCCCTAGCCGAATCAACGGGCCTTGACCAG
S. rexii A	------------ATT--G---------------CTAATCTAGGAATCTCAGAAGCTAGCCCTAGCAGAATCAACGGGCCTTGACCAG
S. hirticapsa A	------------ATT--G---------------CTAATCTAGGAATCTCAGAAGCTAGCCCTAGCAGAATCAACGGGCCTTGACCAG
S. papangae A	TCTTGATCAACTACT--GACATTT-CGTGGCTACTAATGTAGGAATCTCAGAAGCTAGCCCTAGCCGAATCAACGGGCCTTGACCAG
S. schliebenii A	TCTTGATCAACTATT--G---------------CTAATGTAGGAATCTCAGAAGCTAGCCCTAGCCGAATCAACGGGCCTTGACCAG
S. ibityensis A	TCTTGATCAACTATT--GACATTTC-GTGGCGACTAATGTAGGAATCTCAGAAGCTaGCCCTAGCCGAATCAACGGGCCTTGACCAG
S. ibityensis B	TCTTGATCAACTATT--GACATTTC-GTGGCGACTAATGTAGGAATCTCAGAAGCTAGCCCTAGCCGAATCAACGGGCCTTGACCAG
S. tongwensis B	TCTTGATCAACTTTT--GAAATT--CGTGGTAACTATTGTAGGAATCTCAGAAGCTAGCCCTAGCCGAATCGACCGGCCTGGACCAG
S. dunnii C	TCTTGATCAACTTTT--GAAATT--CGTGGTAACTATTGTAGGAATCTCAGAAGCTAGCCCTAGCCGAATCGACCGGCCTGGACCAG
S. modestus A	TCTTGATCAACTTTT--GAAATT--CGTGGTAACTATTGCAGGAATCTCAGAAGCTAGCCCTAGCCGAATCGACCGGCCTGGACCAG
S. thompsonnii A	TCTTGATCAACTTTT--GAAATT--CGTGGTAACTATTGTAGGAATCTCAGAAGCTAGCCCTAGCCGAATCGACCGGCCTGGACCAG
S. beamp. A	TCTTGATCAACTTTT--GAAATC--CGTGGCAACTATTGTAGGAATCTCAGAAGCTAGCCCTAGCNGAATCGACAGGCCTGGACCAG
S. levis A	TCTTGATCAACTTTT--GAAATT--CGTGGCAACTATTGTAGGAATCTCAGAAG??????????????????????????????????
S. venosus A	TCTTGATCAACTTTT--GAAATT--CGTGGCAACTATTGTAGGAATYTCAGAAGCTAGCCCTAGCCGAATCGACAGGCCTGGACCAG
S. spp. 11 A	TCTTGGTCAACTTTTTTGAAATT--CGTGGCAACTATTGTAGGAATCTCAGAAGCTAGCCCTACC-GAATCGACAGGCCTGGACCAG
S. papangae B	TCTTGATCAACTATT--GAACTTT-TGTGGCGATTAATGTAGGAATCTCAGAAGCTAGCCCTAGCCGAATCGACAGGCCTGGACCAG
S. ibityensis C	TCTTGATCAACTATT--GAACTTT-------------TGTRGGAATCTCAGAAGCTAGCCCTAGCCGAATCGACAGGCCTGGACCAG
S. itremensis A	TCTTTATCAACTATT--GAACTTY-CGTGGCGATTAATGTAgGAATCTCAGAAgCTAGCCCTAGCCGAATCGACAGGCCTGgACCAG
S. schliebennii B	TCTTGATGAACTATT--GAATTTTT-GTGGCGATTAATGTAGGAATCTCAGAAGCTAGCCTTGGCCGAATCGACAGGCCTGGACCAG
S. pentherianus A	TCATGATCAAGTATT--GAAGTTTTTGTGGCGATTAATGTACGAGTCTCAGAAGCTAGCCCTAGCTGAATCGACAGGCCTGGMCCAG
S. wendlandii A	TCTTGATCAATTATT--GAA-TTTTTGTGGCGATTAATGTAGGAGTCTCAGAAGCTAGCCCTAGCTGAATCGACAGGCCTGGACCAG
S. rexii B	TCTTGATCAAGTATT--GAA-TTTTTGTGGCGATTAATGTASGAGTCTCAGAAGCTAGCCCTAGCTGAATCGACAGGTCTGGACCAG
S. wittei B	TCTTGATCGAGTATT--GAA-TTTTTGTGGCGATTAATGTAGGAGTCTCAGAAGCTAGCCCTAGCTGAATCGACAGGcCTGGACCAG
S. burundianus B	TCCTGATCAAGTATT--GAA-CTTTTGTGGCGATTAATGTAGGAGTCTCAGAAGCTAGCCCTAGCTGAATCGACAGGTCTGGACCAG
S. hirticapsa B	TCTTGATCAACTATT--GAA-TATTTGTGGCGATTAATGTAGGAGTCTCAGAAGCTAGCCCTAGCTGAATCGACAGGCCTGGACCAG
S. modestus B	TCTTGATCAAGCATT--GAA-TTTTTGTGGCGATTAATGTAGGAGTCTCAGAAGCTAGCCCTAGCTGAATCGACAGGCCTGGACCAG
S. modestus C	TCTTGATCAAGTATT--GAA-TTTTTGTGGCGATTAATGTAGGAGTCTCAGAAGCTAGCCCTAGCTGAATCGACAGGCCTGGACCAG
S. primulifolius A	TCTTGATCAAGTATT--GAA-TTTTTGTGGCGATYAATGTAGGAGTCTCAGAAGCTAGCCCTAGCTGAATCGACAGGCCTGGACCAG

Appendix 5: Properties of data matrices and phylogenies presented in this thesis.

	Number of Characters	Number of variable characters	Number of Parsimony informative characters	number of trees	Length of shortest tree	Consistency index	Retention index
knox nucleotide matrix	189	159	140	1	1124	.286	.611
knox protein matrix	479	403	434	2	2029	.814	.818
Placement of Sdknox2 (nucleotide)	72	31	24	3	31.6	.944	.931
Placement of Sdknox2 (protein)	24	13	7	1	19	.947	.923
Clavata matrix	122	57	39	2	115	.922	.890
NAC matrix	116	51	31	1	99	.970	.914
Coding matrix	57	12	8	1598	23	.826	.942
Coding + intron matrix	126	45	30	5200	70	.829	.939
Repeats	106	62	19	107	85	.871	.676

Appendix 6: GenBank accession numbers of knox nucleotide sequences used in phylogeny

Gene	Species	Accession Number
KNAT1	*Arabidopsis thaliana*	U14714
KNAT2		U14175
KNAT3		X92392
KNAT4		X92393
KNAT5		X92394
STM		U32344
Brassica		AF23753
NTH1	*Nicotiana tabacum*	AB025573
NTH9		AB025713
NTH20		AB025714
NTH15		AB004785
NTH22		AB025715
NTH23		AB004797
Let12		AF000142
Thox1		U76409
Thox2		U76410
Leknox3		U32247
Tkn2		U76407
Tkn3		U76408
POTH1	*Solanum tuberosum*	U65648
Ssknox1	*Streptocarpus saxorum*	-
Ssknox3		BM34634
hirz	*Antirrhinum majus*	AY072736
ina		AY072735
Pisum	*Pisum sativum*	AF063307
Medicago		AF308454
SBH1	*Glycine max*	L13663
Knap3	*Malus domestica*	Z71980

Hvknox3	*Hordeum vulgare*	X83518
Wheat	*Triticum aestivum*	AF224500
Gene	Species	Accession Number
Oskn2	*Oryza sativa*	AF05180
OSH1		D16507
lg3	*Zea mays*	AF100455
KN1		X61308
rs1		L44133
Picea	*Picea abies*	AF063248
crknox1	*Ceratopteris richardii*	AB043955
crknox2		AB043956
crknox3		AB043957
mkn1	*Physcomitrella patens*	AF285148
mkn2		AF285147
mkn4		AF284817
Acetabularia	*Acetabularia acetabulum*	AF170172
Frog	*Xenopus laevis*	U68388
Mouse	*Mus musculus*	U68383
Worm	*Caenorhabditis elegans*	AJ000533