Examination of *CYCLOIDEA*-like genes in the Leguminosae

A thesis submitted to the University of Edinburgh for the degree of Doctor of Philosophy

Hélène Lucie Citerne

The University of Edinburgh
June 2004

DECLARATION

I hereby declare that this thesis is composed of work carried out by myself unless otherwise acknowledged, and that this thesis is of my own composition. This thesis has not in whole or in part been previously presented for any other degree.

Hélène L. Citerne

ACKNOWLEDGMENTS

I would like to thank my supervisors Toby Pennington and Justin Goodrich for their support during my project and for reading draft versions of this thesis, and Quentin Cronk, for making many aspects of this work possible and inviting me to the University of British Columbia. This project would not have been possible without the collaboration of Enrico Coen, John Innes Centre, Norwich and Da Luo, Shanghai Institute of Plant Physiology.

I am grateful to Michelle Hollingsworth and Alex Ponge for assistance in the RBGE lab, and also Catherine Baxter, John Innes Centre, and Karine Coenen, Institute of Cell and Molecular Biology, University of Edinburgh, for their help with RNA *in situ* hybridisation.

I would like to thank the horticultural staff at RBGE, particularly Steve Scott for looking after *Cadia* and *Clitoria* plants, and Clare Morter for growing many batches of lupins. I would also like to thank Maureen Warwick, along with Toby Pennington, for collecting *Cadia* flower material, and Debbie White for taking photos of these plants.

Thank you to Susan Barker at the University of Western Australia for welcoming me into her lab to carry out transgenic experiments. In particular, I'm grateful to Simone Chapple and Leon Hogdson for the training I've received and for continuing the work after I left.

Rick Ree, at the Field Museum, Chicago, was very generous with his time helping me with the PAML analyses.

Finally I would like to thank Karine Coenen, Sophie Neale, Stephen O'Sullivan, and Chun-Neng Wang, for discussions and encouragement.

This project was funded by the Carnegie Trust for the Universities of Scotland, with a contribution from the Systematics Association.

ABSTRACT

Zygomorphic flowers, with a single plane of symmetry, are thought to have evolved independently in diverse angiosperm lineages such as Lamiales and Leguminosae, possibly as an adaptation to insect pollinators. The majority of species belonging to the subfamily Papilionoideae of the Leguminosae have specialised zygomorphic flowers. However, a small number of papilionoid taxa derived from within zygomorphic clades, such as the genus *Cadia*, have evolved atypical radially symmetrical flowers. The genetic control of floral symmetry in the Leguminosae and the genetic basis for the apparent reversal to radial symmetry in *Cadia* were investigated using a candidate gene approach. In the model organism *Antirrhinum majus* (snapdragon, Lamiales), two paralogous genes *CYCLOIDEA* (*CYC*) and *DICHOTOMA* (*DICH*) determine dorsal (adaxial) floral identity and play a crucial role in the establishment of zygomorphy. The orthologue of *CYC/DICH* in *Arabidopsis thaliana TCP1* also has adaxial expression in the early stages of floral development. *CYC*-like genes may therefore be good candidates for the control of dorso-ventral floral symmetry in lineages outside of *Antirrhinum*. Using a phylogenetic approach, homologues of *CYC/TCP1* were identified in legume taxa from the major clades of the Papilionoideae, as well as from subfamilies Caesalpinioideae and Mimosoideae. LEGCYC genes have duplicated prior to the evolution of the Papilionoideae and form three main groups (LEGCYC1A, LEGCYC1B and LEGCYC2). Within these major gene groups, the precise relationships of paralogues between species from the main clades of the Papilionoideae was difficult to determine because of the rapid rate of sequence evolution outside of the conserved TCP and R domains characteristic of *CYC*-like genes. Nevertheless, the phylogenetic framework enabled the identification of orthologous gene pairs in the radially symmetrical papilionoid taxa *Cadia purpurea* and in a closely related species, *Lupinus nanus*, with typical zygomorphic flowers. LEGCYC1A and LEGCYC1B expression in *L. nanus* was

restricted to the adaxial part of the floral meristem and was maintained throughout flower development. This pattern is very similar to *Antirrhinum CYC* and suggests these genes are important for the development of bilateral symmetry in legumes. By contrast, radial symmetry in *C. purpurea* was correlated with an expansion of LEGCYC1B expression in the lateral and ventral petals. It appears therefore that radial symmetry in *Cadia* is not a reversal (*i.e.* with loss of LEGCYC expression during the later stages of floral development) but an evolutionary innovation involving homeotic-like transformation of lateral and ventral floral domains into regions with dorsal identity. Dorsalisation of the corolla is supported by morphological evidence, as the petals of *Cadia* are large and individually bilaterally symmetrical like the standard of typical papilionoid legumes. Patterns of molecular evolution of LEGCYC genes, using explicit codon-based models of evolution in a likelihood framework, were investigated in the clade containing *Lupinus* and *Cadia*. Results suggest positive selection may have acted at certain amino acid sites in *C. purpurea* LEGCYC1B, further implying changes in protein function correlated with changes in floral symmetry. To further establish the role of LEGCYC1A and LEGCYC1B in legume floral development, gene silencing experiments (mediated by RNA interference) were initiated in transformable *Lupinus angustifolius*.

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION 1

1.1 Advances in evolution and development 1

1.2 Organisation of reproductive structures in angiosperms 3

1.3 Types of floral symmetry 5

1.4 Evolution of floral symmetry 7

1.5 Genetic control of floral symmetry in *Antirrhinum* 9

1.6 *CYC* belongs to the TCP family of transcription factors 12

1.7 Role of *CYC* homologues in floral development 13

1.8 Evolution of floral symmetry in the leguminosae 14

1.9 Case study in the genistoid clade 20

1.10 Aims of research 21

CHAPTER 2: PHYLOGENOMIC INVESTIGATION OF *CYCLOIDEA*-LIKE GENES IN THE LEGUMINOSAE 22

2.1 INTRODUCTION 22

PART 1: PCR SURVEY OF *CYC*-LIKE GENES IN LEGUMINOSAE 24

2.2 MATERIALS AND METHODS 24

2.2.1 Taxon sampling and DNA extraction 24

2.2.2 Primer design 27

2.2.3 PCR conditions 27

2.2.4 Cloning and sequencing 28

2.2.5 Confirmation and expansion of results 29

2.2.5a Degenerate primers 29

2.2.5b. Survey of *CYC*-genes in a basal caesalpinioid legume: *Cercis griffithii* 30

2.3 RESULTS 30

2.3.1 PCR survey 30

2.3.2 Sequence data 33

2.3.2a Sequence survey using LEGCYC_F1-LEGCYC R1 33

2.3.2b Saturation cloning 35
2.3.2c Degenerate primers 35
2.3.3 Sequencing of *CYC*-like genes in *Cercis griffithii* 37

PART 2: LEGUME *CYC* GENES WITHIN THE TCP GENE FAMILY 38
2.4 MATERIALS AND METHODS 38
2.4.1 TCP sequence sampling 38
2.4.2 Phylogenetic analyses 38
2.5 RESULTS 39

PART 3: LEGUME *CYC* GENE PHYLOGENY 43
2.6 MATERIALS AND METHODS 43
2.6.1 Sequence sampling and alignment 43
2.6.2 Legume *CYC* sequence analyses 44
2.7 RESULTS 45
2.7.1 Evolution of LEGCYC genes: partial TCP and R nucleotide analyses 45
2.7.2 Evolution of LEGCYC genes: inclusion of sequence data between the TCP and R domains 50
2.8 DISCUSSION 54
2.8.1 Presence of *CYC/TCP1* orthologues in the Leguminosae 54
2.8.2 Problematic reconstruction of legume *CYC*-like gene evolution 55
2.8.3 Evidence for multiple duplication events within the Papilionoideae 56
2.8.4 The limitations and potential of *CYC*-like gene phylogenetics 57

CHAPTER 3: CHARACTERISATION OF *CYC*-LIKE GENES IN *CADIA PURPUREA* AND *LUPINUS NANUS* **58**
3.1 INTRODUCTION 58
3.2 MATERIALS AND METHODS 59
3.2.1 Specific amplification of *CYC*-like loci in *Cadia purpurea* and *Lupinus nanus* 59
3.2.2 Isolation of regions upstream and downstream of the initial fragment of LEGCYC1A and LEGCYC1B in *C. purpurea* and *L. nanus* using PCR based approaches 59
3.2.2a Inverse PCR 59

3.2.2b Standard PCR 60

3.2.2c Genome walking 60

3.2.3 Sequence compilation and comparison 63

3.2.4 Characterisation of intron and splice site 63

3.2.5 Characterisation of the 3'end of other LEGCYC genes in *C. purpurea* and *L. nanus*, with particular reference to LEGCYC2 64

3.3 RESULTS 65

3.3.1 Characterisation of LEGCYC1A and LEGCYC1B in *Cadia purpurea* and *Lupinus nanus* 65

3.3.2 Investigation of other LEGCYC genes including LEGCYC2 68

3.4 DISCUSSION 70

CHAPTER 4: EXPRESSION PATTERNS OF CYC-LIKE GENES IN *LUPINUS NANUS* AND *CADIA PURPUREA* **71**

4.1 INTRODUCTION 71

4.1.1 Comparative gene expression in closely related taxa 71

4.1.2 Expression of *CYCLOIDEA* (*CYC*) and *DICHOTOMA* (*DICH*) 71

4.2 MATERIALS AND METHODS 73

4.2.1 RNA *in situ* hybridisation 73

4.2.1a. Tissue fixation 73

4.2.1b. Probe synthesis 73

4.2.1c. RNA hybridisation 74

4.2.2 Reverse transcription (RT)-PCR 75

4.2.2a. RNA extraction and cDNA synthesis 75

4.2.2b. RT-PCR 76

4.3 RESULTS 78

4.3.1 *In situ* hybridisation 78

4.3.2 RT-PCR 83

4.4 DISCUSSION 88

4.4.1 Expression of LEGCYC genes in a typical papilionoid legume *Lupinus nanus* 88

4.4.2 Radial symmetry in *Cadia purpurea* as an evolutionary innovation 91

4.4.3 A complex expression pattern of LEGCYC genes in *C. purpurea* 93

4.4.4 Further work 94

CHAPTER 5: MOLECULAR EVOLUTION OF LEGCYC GENES IN THE GENISTOID CLADE 96

5.1 INTRODUCTION 96

5.2 MATERIALS AND METHODS 100

5.2.1 Taxon sampling 100

5.2.2. PCR and sequencing 101

5.2.3. Sequence alignment and phylogenetic analyses 102

5.2.4. Analyses of LEGCYC coding sequence evolution 103

5.2.5. Analysis of Lupinus nanus LEGCYC1A* 104

5.3 RESULTS 104

5.3.1. Range of LEGCYC primers 104

5.3.2. Phylogenetic analyses of LEGCYC paralogues in the genistoid clade 106

5.3.2a. Sequence data 106

5.3.2b. Phylogenetic analyses 107

5.3.3. Testing for positive selection 108

5.3.3a Site models 108

5.3.3b Branch models 111

5.4.4. Phylogenetic position of LEGCYC1A* 117

5.4 DISCUSSION **118**

5.4.1 Phylogenetic potential of LEGCYC genes in the genistoid clade 118

5.4.2 Selection pressures across LEGCYC paralogues 119

5.4.3 Limitations of this study and general conclusion 121

CHAPTER 6: GENE SILENCING IN *LUPINUS ANGUSTIFOLIUS* **123**

6.1 INTRODUCTION 123

6.1.1 Investigating gene function 123

6.1.2 Mechanism of RNA interference (RNAi) 124

6.1.3 Use of RNAi for the discovery of gene function 126

6.1.4 Experimental background 126

6.2 MATERIALS AND METHODS 127

6.2.1 Silencing construct design 127

6.2.2 Gene transfer in *Lupinus angustifolius* 129

6.2.2a *Agrobacterium tumefaciens* transformation 129
6.2.2b Explant preparation 128
6.2.3 Transformant screening 132
6.3 RESULTS 132
6.3.1 Frequency of transformation 132
6.3.2 Floral phenotypes of putative transformants 133
6.4 DISCUSSION 134
6.4.1 Transformation efficiency 134
6.4.2 Predicted results and limitations of this study 134
6.4.3 Future work 136

CHAPTER 7: CONCLUSIONS AND FUTURE WORK **137**
7.1 Summary of findings 137
7.1.1 Phylogenetic framework 137
7.1.2 Functional inferences from expression data 138
7.2 Future work 139
7.2.1 Detailed characterisation of LEGCYC function 139
7.2.2 Examination of other unusual papilionoid legumes 141
7.2.3 Evolution of floral symmetry in other lineages 142

REFERENCES **146**
APPENDIX 1 : MOLECULAR PROTOCOLS 162
APPENDIX 2 : PRIMERS 166
APPENDIX 3 : TCP AMINO ACID MATRIX 169
APPENDIX 4 : LEGCYC NUCLEOTIDE MATRIX 170
APPENDIX 5 : *C.PURPUREA* AND *L.NANUS* LEGCYC1A/LEGCYC1B SEQUENCES 175
APPENDIX 6 : OTHER SEQUENCES 181
APPENDIX 7 : GENISTOID LEGCYC1A/LEGCYC1B NUCLEOTIDE MATRICES 183
APPENDIX 8 : PUBLICATION 196

LIST OF TABLES AND FIGURES

CHAPTER 1: INTRODUCTION

Figure 1-1. Simplified model of interaction of floral organ identity genes, first proposed by Coen and Meyerowitz (1991). The floral meristem is divided into three overlapping regions of homeotic gene activity resulting in four concentric whorls of floral organs. A-class genes, including *APETALA1* (*AP1*) and *APETALA2* (*AP2*) in *Arabidopsis* affect development in the outer two whorls (sepals and petals), B-class genes, such as *APETALA3* (*AP3*) and *PISTILLATA* (*PI*) affect development in whorls 2 and 3 (petals and stamens), and C-class genes such as *AGAMOUS* (*AG*) affect development in the inner two whorls (stamens and carpels). Some A and C-class genes have been found to be mutually antagonistic. This model has been extended with the discovery of other floral organ identity genes (*e.g.* redundant E class *SEPALLATA* genes specifying petal, stamen and carpel development; reviewed in Theiβen *et al.*, 2002). 5

Figure 1-2. The two main types of floral symmetry: actinomorphy and zygomorphy in relation to corolla shape, compared with absence of symmetry (asymmetry); reproduced from Endress (2001). 6

Figure 1-3. *Antirrhinum majus* flowers of wild type and *dich*, *cyc* and *cyc/dich* mutant (E. Coen, John Innes Centre, Norwich). Loss of *CYC* function has a greater effect on phenotype than loss of *DICH*, although loss of both genes is required for a fully radial phenotype. 10

Figure 1-4. Representation of major legume lineages, showing the relationship of the monophyletic subfamilies Papilionoideae and Mimosoideae, and a grade of caesalpinioid tribes. Redrawn from Doyle and Luckow (2003) and Wojciechowski (2003); caesalpinioid tribes defined in Wojciechowski (2003). 15

Figure 1-5. Examples of flowers from the three subfamilies of the Leguminosae (from Watson & Dallwitz, 1992). Transverse sections along the median axis are shown for *Genista* and *Cercis*. *Genista* has typical papilionoid flowers, with a reflexed adaxial petal, and differentiated lateral and ventral petals enclosing the stamens and carpel. *Acacia retinoides*, like many mimosoid species, has a reduced radially symmetrical perianth and a proliferation of free stamens. Caesalpinoids are more variable in floral morphology, usually more open and with less differentiated petals than papilionoids. *Cercis*, shown here, has flowers which superficially resemble those of papilionoids. 16

Figure 1-6. Current phylogeny of the Papilionoideae, redrawn from Wojciechowski (2003). Taxa with atypical non-papilionoid flowers (listed in Pennington *et al.*, 2000) are in bold, and in addition those with more or less radially symmetrically flowers are highlighted. The swartzioid clade is sister to other papilionoid lineages. It was estimated that twelve instances of reversals from zygomorphy to actinomorphy occurred in the Papilionoideae. * denotes clades with over 50% bootstrap support (from different sources of molecular data, summarised by Wojciechowski, 2003). 18

Figure 1-7. Floral diversity in the Papilionoideae. A-C: typical zygomorphic papilionaceous flowers, adapted to bee pollination. A: *Lupinus nanus*, B: *Lotus japonicus,* C: *Cytisus* sp., showing bee pollination. D-G: Papilionoid legumes with unusual open flowers. D: *Cyathostegia matthewsii*, E: *Swartzia pinnata* have reduced/lost corolla parts and a proliferation of stamens. F: *Cadia purpurea*, G: *Acosmium panamense*, have near radially symmetric flowers. Photos: A, F: D. White, RBGE; C: Q.Cronk, UBC Botanical Garden; B: S. Suehiro, Japan; D: G.P. Lewis, RBG Kew; E,: T. Pennigton, RBGE, G: L. Pauwels, Belgium. 19

Figure 1-8. Lateral view of actinomorphic (A) and zygomorphic (B) flowers from two genera from the genistoid clade A. *Cadia purpurea* (from Polhill, 1981) and B. *Lupinus sabinii* (from the Rare Plants in Washington, University of Washington, http://courses.washington.edu/rare care/RarePlantsinWashington.htm). 20

CHAPTER 2: PHYLOGENOMIC INVESTIGATION OF *CYCLOIDEA*-LIKE GENES IN THE LEGUMINOSAE

Figure 2-1. Schematic representation of the relationship of some of the major groups in the Papilionoideae as defined by current molecular evidence (Doyle *et al.*, 1997; Hu *et al.*, 2000; Kajita *et al.*, 2001; Pennington *et al.*, 2001), with representative taxa used in the legume *CYC* sequence analyses. 24

Table 2-1. List of taxa included in the PCR survey of *CYC*-like genes using primers LEGCYC_F1 and LEGCYC_R1. Taxa are listed according to subfamily (Caesalpinioideae, Mimosoideae, Papilionoideae). Within the Papilionoideae, the major clades are shown (* names follow the nomenclature of Pennington *et al.*, 2001) based on recent phylogenetic evidence (Doyle *et al.*, 1997; Hu *et al.*, 2000, Kajita *et al.*, 2001; Pennington *et al.*, 2001). ** Source number refers to either RBGE living collection number (*e.g.* 1996 0942A) or collector's voucher number from wild collections (*e.g.* R.T. Pennington 473), with the exception of *Pisum sativum* DNA from genetic line 399 grown at the JIC. All herbarium vouchers at RBGE. 25

Figure 2-2. Representation of the legume *CYC*-like open reading frame (ORF), based on a 27

cDNA sequence from *Lotus japonicus* (Lotus japonicus 2; D. Luo, pers. comm.), with sequences of the conserved TCP and R domains shown. Sequence in red/bold mark the priming location of the legume *CYC* primers LEGCYC_F1 and LEGCYC_R1.

Figure 2-3. PCR products (3µl load), amplified using primers LEGCYC_F1 and LEGCYC_R1 in a range of legumes, separated on a 2% agarose gel for 2 ½ hours at 80V. Products are run against a 1 kb ladder (L). –ve: negative control. Taxa corresponding to each lane are given in table 2-2. 31

Table 2-2. Results of the PCR survey using primers LEGCYC_F1 and LEGCYC_R1 on an array on taxa from the three subfamilies of the Leguminosae. The presence and number of bands visible on a 2% agarose gel run for 2 ½ hours at 80V is given for each taxa. The lane number refers to figure 2-3, some products are not shown (ns). Products much larger than 500 bp are given in parentheses. 32

Table 2-3. Number of sequence types with a TCP and R domain obtained from cloned PCR products amplified using primers LEGCYC_F1 and LEGCYC_R1. GenBank accession numbers corresponding to the partial gene nucleotide sequence are given. Two other *CYC*-like sequences were obtained with different primers from *Lupinus* species and included in the phylogenetic analyses (part 3, this chapter): Lupinus sp. 1 (AY225832) amplified with primers LEGCYC_F2 and LEGCYC_R2 (described in section 2.2.3a), and L. nanus 1 (AY225836) with locus specific primers (see chapter 3). 34

Figure 2-4. Number of clones sequenced from *Cadia purpurea*, *Lupinus* sp. and *Lupinus nanus* from PCR reactions using different primer combinations, including the highly degenerate primers F2, F4 and R2. Cloned PCR products have been grouped into different *CYC*-like sequence types (*i.e.* with a TCP and R domain), plus those which are not TCP genes. Numbers referring to sequence type do not imply homology between *C. purpurea* and *Lupinus* sequences. **2-4a.**☐ *C. purpurea* F1-R1,☐ *C. purpurea* F2-R2, ☐ *C. purpurea* F4-R2. Sequence type I = Cadia 1, sequence type II = Cadia 2, sequence type III = Cadia 3, sequence type IV = Cadia 4. **2-4b.**☐ *L. nanus* F1-R1,☐ *Lupinus* sp. F1-R1, ☐ *Lupinus* sp. F2-R2. Sequence type 1 = Lupinus nanus 2, Lupinus sp. 2 ; sequence type 2 = Lupinus nanus 3, Lupinus sp. 3 ; sequence type 3= Lupinus sp. 4 ; sequence type 4 = Lupinus sp. 1 (sequence names listed in table 2-3). Degenerate primers were not found to amplify CYC-like genes specifically in *C. purpurea*, but did amplify a *CYC*-like gene in *Lupinus* sp. that was not amplified by LEGCYC_F1-LEGCYC_R1. 36

Figure 2-5. An alignment of the predicted amino acid sequence from Ceratonia 1 and Cercis 1, and Cadia 4 and Cercis 2. Identical amino acids are in black boxes, while amino acids with similar charge or hydrophobicity are in grey. The partial TCP and R domains are shown for both sequence pairs. 37

Figure 2-6. Unrooted phylogram of protein ML analysis using TREEPUZZLE v5.0 (Schmidt *et al.*, 2000) of the TCP domain data set including representative legume sequences. The *CYC-TB1* and PCF groups described in Cubas (2002) are recovered here, as well as a group containing *CIN*-like genes (Palatnik *et al.*, 2003). Support values were obtained using MrBayes (Hulsenbeck and Ronquist, 2001); asterisks * indicate that a clade was recovered in < 50% of Bayesian trees. Results support a LEGCYC clade (highlighted in green, excluding Cadia 4) as sister to the *CYC/TCP1* clade. All TCP genes, unless otherwise indicated, are from *Arabidopsis*; *PCF* from rice; *TB1* from maize; *LCYC* from *Linaria vulgaris*, *CYC* and *DICH* from *Antirrhinum*; *AUX* from cotton (accession numbers in appendix 3). 41

Figure 2-7. **2-7a.** Fifty percent Majority Rule (MR) consensus tree of the protein distance analysis using the PAM-Dayhoff model of protein substitution (PROTDIST; Felsenstein, 1993) of the TCP domain. Values > 50% of the 100 jackknife replicates are given at branch nodes. Taxa as in legend to figure 2-6. **2-7b.** Fifty percent MR consensus tree of protein maximum parsimony analysis (PROTPARS; Felsenstein, 1993) of the TCP domain. Support values above 50% from the 100 jackknife replicates are shown. Maximum parsimony fails to resolve groups recovered in protein-ML, Bayesian-and distance-analyses. Although-it does not contradict any of the-results from other methods, it offers no support for a *CYC-TB1* clade, and only weak support (54%) for a LEGCYC clade. 42

Figure 2-8. Strict consensus of 194 most parsimonious trees of partial TCP and R nucleotide sequences (CI = 0.321, RI = 0.567), rooted on *Antirrhinum* CYC and DICH. Support values above 50% from the 1000 bootstrap replicates are shown below branches. 46

Figure 2-9. Analyses of 29 partial legume TCP and R domain nucleotide sequences. **2-9a.** Strict consensus of 168 most parsimonious trees (CI = 0.424, RI = 0.636), with bootstrap values shown below branches. **2-9b.** Bayesian analysis 50% MR tree of the legume TCP and R nucleotide sequences allowing for codon specific nucleotide substitution. Major clades I and II within LEGCYC are indicated with high Bayesian support. Both consensus trees are rooted on *Antirrhinum CYC* and *DICH*. 48

Figure 2-10. One of three most likely trees of the TCP plus R data set, analysed with the parameters of the best-fit model TIM + I + G selected by the Akaike Information Criterion. All trees have an identical topology, but differ in branch lengths. Group II (marked by the red bar), also recovered by maximum parsimony and Bayesian analysis of the same data, is nested here within a grade of LEGCYC sequences. 49

Figure 2-11. Maximum parsimony and ML analyses of 38 partial legume *CYC*-like sequenc some sequence data from the hypervariable region. Major groups recovered from the previous anal and group II) are shown, as well as one putative duplication event in group I is marked by IA an containing genistoid (in red) and robinioid (in blue) sequences are highlighted suggesting th duplication events. **2-11a.** Unrooted phylogram of the single most parsimonious tree (748 steps, CI = 0.601). Bootstrap values are given for branches with > 50% support. **2-11b.** Unrooted phylogra analysis using the GTR + I + G model of nucleotide substitution. Support values at each node were Bayesian analysis of the data set and represent the frequency of each node in the MR consensus tree. 51

Figure 2-12. Comparison of the partial TCP domain amino acid sequence from group I and II *CYC*-like sequences in legumes. Asterisk highlights group-specific changes; above and below bold sequences are amino acid differences found less frequently in these groups. 52

Table 2-4. Descriptive values of the maximum parsimony analyses carried out with different nucleotide data sets: 1: all LEGCYC, *Antirrhinum CYC*, *DICH*, and *Arabidopsis TCP1* partial TCP and R nucleotide data (figure 9); 2: partial TCP and R nucleotide data of a subset of LEGCYC sequences (figure 10); 3: inclusion of the hypervariable region between the TCP and R domain, aligned against a subset of LEGCYC sequences (figure 13). MP trees: most parsimonious trees, CI: consistency index, RI: retention index. 53

CHAPTER 3: CHARACTERISATION OF *CYC*-LIKE GENE SEQUENCES IN *CADIA PURPUREA* AND *LUPINUS NANUS*

Table 3-1. Summary of the different PCR approaches used to isolate regions flanking known fragments of two *CYC*-like genes, LEGCYC1A and LEGCYC1B, in *Lupinus nanus* and *Cadia purpurea*. Details of template preparation for inverse PCR and genome walking are given in sections 3.2.2a and 3.2.2c respectively. Primer sequences and location are given in appendix 2. PCR mix was as follows in all reactions: sterile distilled water, polymerase buffer, $MgCl_2$ (2.5mM), dNTP's (20μM), primers F1 and R1 (0.5μM each), 1 unit *Taq* polymerase (Bioline Ltd., London NW2, UK). * The annealing/extension temperature is decreased by 1°C per cycle for the first eight 62

cycles of the genome walking PCRs.

Figure 3-1. Schematic representation of the LEGCYC open reading frame (ORF), showing the TCP and R domains, and the short intron. The binding sites of general primers LEGCYC_F3, LEGCYC_R1, LEGCYC_R5 and LEGCYC_R8 are shown. 64

Figure 3-2. An alignment of the predicted amino acid sequence of the complete open reading frame of Cadia 1, Lupinus nanus 1 (LEGCYC1B), Cadia 2, Lupinus nanus 2 (LEGCYC1A). Identical amino acids are in black boxes, while amino acids with similar charge or hydrophobicity are in grey. The TCP and R domains are shown, as well as the EVV motif and another putative helix domain ("new domain") which are both found in *Antirrhinum DICH.* 66

Figure 3-3. Pairwise distances of nucleotide sequences (excluding the intron: hatched region) between *Lupinus nanus* and *Cadia purpurea* LEGCYC1A and LEGCYC1B orthologues respectively. Loci are divided into five regions: three hypervariable regions and the TCP and R domains (in grey). 67

Figure 3-4. PCR products (3µl load) amplified in *Cadia purpurea* and *Lupinus nanus* using the forward primer in the TCP domain LEGCYC_F3 in combination with LEGCYC_R1 (in the R domain), LEGCYC_R5 and LEGCYC_R8 (3' of the intron). All primer combinations amplify three distinct bands in both taxa. C: *Cadia purpurea*, L: *Lupinus nanus*, –ve: negative control (no DNA in sample); 1Kb: 1Kb ladder (Bioline Ltd., London NW2, U.K.). 68

CHAPTER 4: EXPRESSION PATTERNS OF *CYC*-LIKE GENES IN *LUPINUS NANUS* AND *CADIA PURPUREA*

Figure 4-1. RNA *in situ* hybridisation of longitudinal sections of wild type *Antirrhinum* inflorescence (a) and flowers (b, c) probed with *CYC*. A signal can be detected in the adaxial region of the floral meristem prior to organogenesis through to organ differentiation. At early stages, the signal can be detected in the adaxial sepal primordia and the dorsal region of the floral dome (b). At later stages, the signal is detected in the dorsal petal and staminode (c). b: bract, ds: dorsal sepal, vs: ventral sepal, d: dorsal petal, l: lateral petal, st: stamen: std: staminode, c: carpel. Scale bar 100 µm. Reproduced from Luo *et al.*, 1996. 72

Figure 4-2. Dissected mature flowers of *Ulex europaeus* (4-2a), a close relative of *Lupinus* with similar typical papilionoid flowers, and *Cadia purpurea* (4-2b). Organs in the three outer 75

whorls are divided into dorsal (D), lateral (L) and ventral (V) domains. Strong differentiation in the calyx, corolla and androecium (ANDR) is found in typical papilionoid flowers such as those of *Ulex*, whereas no differentiation is observed in these whorls in *C. purpurea*. The gynoecium (GYN) in both taxa is typical of the Papilionoideae.

Figure 4-3. Expression pattern of LEGCYC1A (4-3a) and LEGCYC1B (4-3b) in *Lupinus nanus* inflorescences fixed in PFA (hybridisation carried out at ICMB; appendix 1B-D). Longitudinal sections of *L. nanus* inflorescences show floral meristems (fm) in the axil of bracts (B). The adaxial (Ad) and abaxial (Ab) regions are shown in relation to one floral meristem (4-3a). The early stages of organogenesis can be seen in more developmentally advanced flowers at the base of the inflorescence. RNA from LEGCYC1A and LEGCYC1B is detected in the adaxial part of floral meristems prior to organogenesis, as well as during floral organ development. Negative control (sense probe) shown in figure 4-3c. 79

Figure 4-4. RNA *in situ* hybridisation of LEGCYC1A (A-D) and LEGCYC1B (E-H) in the developing flowers of *Lupinus nanus* (hybridisation carried out at ICMB). The flowers are subtended by bracts (B) on the abaxial (ventral) side. Both genes are expressed in the flower meristem (fm) prior to organogenesis (figures A, E), and in the adaxial sepal (AdS) as it develops (figures B, F). In more advanced developmental stages (figures C-D, G-H), expression is found in the adaxial petal (AdP). Although both copies have a similar expression pattern, LEGCYC1B has a wider expression domain than LEGCYC1A, particularly in later developmental stages. St: stamen, AbS: abaxial sepal. 80

Figure 4-5. RNA *in situ* hybridisation of LEGCYC1A and LEGCYC1B in *Lupinus nanus* flowers fixed in FAA (hybridisation carried out at JIC). Patterns of expression are in agreement with *in situ* hybridisation of LEGCYC1A and LEGCYC1B in inflorescence material fixed in 4% PFA (figures 4-3 and 4-4). As in figures 4-3 and 4-4, LEGCYC1B was found to have a larger expression domain compared to LEGCYC1A, particularly at later stages (B and D). fm= floral meristem, B= bract (subtending the flower on the abaxial side), AdS = adaxial sepal, AdP= adaxial petal, AbP= abaxial petal, St= stamen. 81

Figure 4-6. RNA *in situ* hybridisation in *Cadia purpurea* flower material. Although no hybridisation was detected using either LEGCYC1A or LEGCYC1B antisense probes (not shown), a histone probe used as a positive control (4-6a) may be showing hybridisation in a region of intense cell division, the pollen sacs in the stamens (St), compared to the negative control (using a LEGCYC1B sense probe) (4-6b). 82

Figure 4-7. RT-PCR analysis of LEGCYC1A and LEGCYC1B expression in developing vegetative (leaf) and floral tissue in *Cadia purpurea* and *Lupinus nanus*, with amplification of actin cDNA used as a control. Lanes with cDNA amplification are marked by a line. Results in *L. nanus* confirm that both LEGCYC1A and LEGCYC1B are florally expressed genes, however LEGCYC1A is also transcribed in vegetative leaf tissue. Results in *C. purpurea* suggests that both LEGCYC copies are expressed florally, with LEGCYC1A also expressed in leaf tissue as in *L. nanus*. 84

Figure 4-8. RT-PCR analysis of LEGCYC1A and LEGCYC1B expression in the different whorls of the developing flower of *Cadia purpurea* and *Lupinus nanus*, with amplification of actin cDNA used as a control. Results in *L. nanus* are in agreement with the *in situ* hybridisation pattern, with both LEGCYC1A and LEGCYC1B transcribed in the dorsal region. Results in *C. purpurea* suggest that whereas LEGCYC1A is weakly expressed in the dorsal petal, LEGCYC1B is expressed in all petals, and correlates with the lack of differentiation within the corolla. Neither LEGCYC1A nor LEGCYC1B seem to be transcribed in the androecium or gynoecium, whereas LEGCYC1A appears to be transcribed in the dorsal and lateral region of the calyx. DS = dorsal sepal, LS = lateral sepals, VS = ventral sepals, DP = dorsal petal, LP = lateral petal, VP = ventral petal, DSt = dorsal stamens, LSt = lateral stamens, VSt = ventral stamens, G = gynoecium, gDNA = genomic DNA, -ve = negative control. Lanes with PCR products amplified from cDNA are marked by a line. 86

Figure 4-9. RT-PCR analysis of LEGCYC2 expression in developing vegetative (leaf) and floral tissue in *Cadia purpurea*, with amplification of actin cDNA used as a control. Lanes with cDNA amplification are marked by a line. An apparently low level of LEGCYC2 transcripts was detected in floral tissue. 87

Figure 4-10. RT-PCR analysis of LEGCYC2 expression in the dissected calyx, corolla, androecium and gynoecium of *Cadia purpurea*, with amplification of actin cDNA used as a control. Lanes with cDNA amplification are marked by a line. LEGCYC2 transcripts were detected in the calyx and corolla, with no apparent asymmetry, but not in the androecium or gynoecium. 87

Figure 4-11. Summary of eudicot phylogeny (based on results from Soltis *et al.*, 1999). Representative taxa with known asymmetric expression of *CYC*-like gene in axillary meristems are shown in green. The occurrence of this adaxial expression pattern in distantly related species may suggest that it facilitated the evolution of zygomorphy in distantly related lineages, through 89

modifications of *CYC*-like gene regulation. Phylogeny reproduced from Cronk (2001). R = rosid, ER1 = eurosid 1, ER2 = eurosid 2, A = asterid, EA1 = euasterid 1, EA2 = euasterid 2.

Figure 4-12. Simplified model of the control of floral symmetry in papilionoid legumes. A typical papilionoid flower (with only petals shown, left) can be divided into dorsal, lateral and ventral domains, where LEGCYC is a marker for dorsal identity. The evolution of radial symmetry in *Cadia* appears to have resulted from the expansion of the expression domain of one LEGCYC gene, so that all petals have dorsal identity (right). 93

CHAPTER 5: MOLECULAR EVOLUTION OF LEGCYC GENES IN THE GENISTOID CLADE

Figure 5-1. Summary of phylogenetic relationships within the genistoid clade (redrawn and modified from Wojciechowski, 2003), based on results from nrDNA ITS and *rbc*L (Crisp *et al.*, 2000; Kajita *et al.* 2001), and *trn*L intron (Pennington *et al.*, 2001) analyses. * denotes clades with bootstrap support greater than 50%, based mainly from Crisp *et al.*, (2000), and Pennington *et al.* (2001). Taxa highlighted in yellow were sampled for the LEGCYC sequence analyses. Taxa underlined have near-radially symmetrical flowers; their distribution suggests that radial symmetry evolved independently in the genistoid clade. Tribes are given on the right. The core genistoid clade is defined by Crisp *et al.* (2000) and Wojciechowski, 2003; a broader definition, with *Ormosia* as sister to all other genistoids, is given by Pennington *et al.* (2001). 99

Table 5-1. List of taxa from the core genistoid clade and sister group (*sensu* Wojciechowski, 2003) used to test the primers LEGCYC_F5-LEGCYC_R4/R3 and LEGCYC_iR4/iR3-LEGCYC_R8 specific to LEGCYC1A and LEGCYC1B (see appendix 2). 101

Table 5-2. Amplification results using primer combinations specific to LEGCYC1A (LEGCYC_R4/iR4) and LEGCYC1B (LEGCYC_R3/iR3) in a range of genistoid taxa. √ = amplification of a single band of the expected size, √ mul = amplification of multiple bands, Ø = no amplification. 105

Figure 5-2. One of the two most parsimonious trees of LEGCYC1A nucleotide matrix (447 steps, CI = 0.859, RI = 0.795) rooted on *Bowdichia*, and of LEGCYC1B nucleotide matrix (658 steps, CI = 0.781, RI = 0.711) rooted on *Ormosia*, with bootstrap support shown in bold. * marks branches which collapse in the strict consensus tree. 108

Table 5-3. Parameter estimates for LEGCYC1A and LEGCYC1B under site models. *p* is 110

the number of free parameters for ω. lnL is the log likelihood of each model. p_n describes the proportion of sites having ω_n. For M7 and M8, p and q describe the beta distribution of ω values. None of these models detected sites under positive selection across the entire phylogeny in either locus.

Figure 5-3. Cladograms of LEGCYC1A and LEGCYC1B showing the foreground ω_2 value obtained under model B for each branch. Branches with ω_2 values greater than one, indicative of positive selection on some sites on that particular lineage, are in bold. For LEGCYC1B, only *Cadia* has an ω_2 value much greater than 1, whereas for LEGCYC1A, these are scattered across the phylogeny. 113

Table 5-4. Parameter estimates from the 2-ratio and branch-site models for selected LEGCYC1A and LEGCYC1B foreground branches where $\omega > 1$ under one of these models. p is the number of free parameters for ω. lnL is the log likelihood of each model. p_n describes the proportion of sites having ω_n. For the two-ratio model, ω_0 is the background estimate and ω_1 the foreground estimate. In the branch-site models, ω_2 is the additional parameter for a site class in the foreground branch and p_2 the proportion of sites in this class. For LEGCYC1B, only the *Cadia* branch was found to have a higher non-synonymous rate, whereas for LEGCYC1A more branches showed a signature of positive selection (also table 5-5). The location of positively selected sites (with a posterior probability $P > 0.5$) is shown in figure 5-4. 114

Table 5-5. Parameter estimates for *Sophora* LEGCYC1B and *Bowdichia* LEGCYC1A from the two-ratio and branch-site models. Both branches have ω_2 greater than 1 under the model B, although the d_N/d_S is close to 1 for the *Sophora* branch suggesting a proportion of sites are evolving neutrally. p is the number of free parameters for ω. lnL is the log likelihood of each model. p_n describes the proportion of sites having ω_n. For the two-ratio model, ω_0 is the background estimate and ω_1 the foreground estimate. In the branch-site models, ω_2 is the additional parameter for a site class in the foreground branch and p_2 the proportion of sites in this class. Position and codon translation of sites identified in the ω_2 site class are given, along with their posterior probability (P). The location of positively selected sites (with a posterior probability $P > 0.5$) is shown for the *Bowdichia* branch in figure 5-4. 115

Figure 5-4. Location of the inferred non-synonymous mutations (with a posterior probability greater than 0.5 under model A or B) along the partial LEGCYC coding region, using *Genista tenera* sequences as reference. The predicted secondary structure (NNPREDICT ; Kneller *et al.*, 116

1990) is given for each locus, with helix and beta-strands regions shown., and the helix-loop-helix region of the TCP domain highlighted. Ancestral and derived amino acids are shown below and above the line respectively. For LEGCYC1B, derived amino acids are shown for the *Cadia purpurea* sequence. For LEGCYC1A, derived amino acids are shown for the *Lupinus digitatus/L. angustifolius* branch (red), *Bowdichia vigilioides* (blue) and *L. nanus* (green). One mutation was inferred in the TCP domain for *B. vigilioides* and one for the *L. digitatus/L. angustifolius* lineage.

Figure 5-5. Unrooted phylogram of one most parsimonious tree out of two MP trees of 383 steps (CI = 830, RI = 733) of sequences amplified by LEGCYC1A specific-primers (LEGCYC_iR4/R4) and *L. nanus* LEGCYC1A*. The branch marked with * collapsed in the strict consensus tree. 117

CHAPTER 6: GENE SILENCING IN *LUPINUS ANGUSTIFOLIUS*

Figure 6-1. Current model of RNA interference (redrawn from the Ambion RNAi resource: http://www.ambion.com/techlib/append/RNAi_mechanism.html). Similar models have been described in plants (Waterhouse *et al.*, 2001), animals (*e.g.* nematodes, Montgomery *et al.*, 1998) and fungi (Pickford *et al.*, 2002). Upon introduction into an organism, long double stranded RNAs (dsRNAs) are processed by a dicer-containing complex into 21-25 bp small interfering RNAs (siRNAs). These siRNAs assemble with an endonuclease-containing complex, known as RNA-induced silencing complexes (RISCs). The anti-sense strand of the siRNA guides the RISC to complementary mRNA, where cleavage is induced. 125

Figure 6-2. Plasmid maps showing the transformed pFGC514 RNAi vector (ChromDB, Arizona, USA) with inserted *CYC* fragments (in yellow), generated with BioEdit v5.0.9 (Hall, 2001). Details of the portion transferred to *L. angustifolius* generating *CYC*-specific dsRNA fragments are given in figure 6.3. The plasmids have a kanamycin resitant gene (Km) for selection of *Agrobacterium tumefaciens*. CaMV p35S: cauliflower mosaic virus promoter, CHSA intron: 1,353 bp fragment from the petunia Chalcone synthase A gene, OCS -3': poly adenylation signal sequence from *A. tumefaciens*, for trancription termination. The selectable marker BAR gene conveys resistance to the herbicide Basta. pMAS 1': plant promoter from *A. tumefaciens*, MAS 3': poly adenylation signal sequence from *A. tumefaciens*. LB: left border repeat from T-DNA; RB: right border repeat from T-DNA. 128

Figure 6-3. Schematic outline of the intron-spliced hairpin RNA construct tranferred to lupins for RNA-mediated gene silencing, from the pFGC5149 vector (ChromDB, Arizona, USA), 129

modified with GATEWAY adaptors for directional insertion of DNA target sequence (TS). The target sequence (TS) fragments are inserted in opposite orientation to form a dsRNA structure. Primers pFGCF1/R1/F2/R2 specifically bind to regions flanking the two cloning sites of pFGC5941, and are therefore transgene specific. Abbreviations are given in figure 6-2.

Figure 6-4. Stages in *Lupinus angustifolius* transformation and explant regeneration (following the protocol of Pigeaire *et al.*, 1997). *L. angustifolius* seeds were germinated overnight (A), dissected to expose the apical meristem (B), and co-cultivated with *Agrobacterium* containing the dsRNA construct (C). Explant were regenerated over two days (D). Shoots were then dissected and placed on selective medium containing PPT (20mg/l), the active ingredient of the herbicide Basta (E). Surviving shoots (F) were then subcultured on selective medium (G). When explants reached a certain size (~ 5cm in height), roots were induced (H). At this stage, sterile flowers were observed (I). 131(

Figure 6-5. Amplification of transgene in surviving explants (L. Hogdson, UWA) using the pFGC5149 specific primers pFGC-F2 and pFGC-R2. Lanes with products from plants transformed with the LEGCYC1A construct are marked by ▬, lanes with products from plants transformed with the LEGCYC1B construct are marked by ═. –ve: negative control, +: positive control (plasmid DNA), L: 100 bp ladder. 132

Figure 6-6. Mature flower of T0 plant with LEGCYC1A inverted repeat insert (A) and wild type (B) *L. angustifolius* cv. Merrit. Although no differences were visible, T0 plants are often chimeric and therefore seldom informative in transformation experiments. 133

CHAPTER 7: CONCLUSIONS AND FUTURE WORK

Figure 7-1. Wild type (A) and mutant (B) *Clitoria ternatea* flowers. In the mutant, all petals are equal and resemble the wild type standard. 142

Figure 7-2. Schematic representation of the the Leguminosae and sister clades, based on molecular data (from Doyle & Luckow, 2003). The Polygalaceae (*Polygala paucifolia*; Ken Systma, UW Madison, dept Botany Plant Systematics Collection) have strongly zygomorphic flowers, whereas Surianaceae (*Suriana maritima*; Tim Motley University of Hawaii Botany dept.) and *Quillaja* (*Quillaja saponaria*; San Marcos growers) have radially symmetric flowers. 144

CHAPTER 1: Introduction

1.1 Advances in evolution and development

One of the fundamental questions in evolutionary biology is concerned with the processes underlying the origin of novel phenotypic characters. At the proximate (intrinsic) level, these changes are the result of modifications in the genome. Three critical questions surround the genetic basis of morphological evolution (Doebley & Lukens, 1998):

1) Are traits controlled by many genes of small individual effects, as proposed by the neo-darwinian synthesis, or are changes in a few genes of large effect sufficient for the establishment of new traits?
2) Are certain classes of genes particularly important contributors to the evolution of new traits? If so, are these regulatory genes such as transcription factors or RNA binding proteins, or downstream genes controlled by these regulatory genes?
3) What types of changes are responsible for genetic modifications that are evolutionarily significant: mutations in the protein coding region, or changes in the *cis*-regulatory elements controlling spatial and temporal gene expression?

The considerable advances in molecular genetics from a few model species have provided a starting point for studying morphological diversity and evolution at the molecular level. Pioneer work carried out in *Drosophila* led to the discovery of homeobox (*HOX*) genes, a family of transcription factors that regulate anterior posterior segment identity (Lewis, 1978). *HOX* gene homologues have subsequently been found in numerous animal groups, where changes in gene regulation were found to have shaped large-scale changes in animal body plan

and parts (reviewed in Caroll, 2000). In plants, genes controlling the developmental fate of meristems and primordia have been isolated (*e.g.* Carpenter & Coen, 1990; Coen & Meyerowitz, 1991; Vollbrecht *et al.*, 1991). These genes have been termed homeotic because they replace one member of a series of meristic units with another. Since the early 1990s, many developmental genes have been isolated and their function characterised in model plant species, in particular *Arabidopsis thaliana* L. (Brassicaceae; eurosid II), snapdragon (*Antirrhinum majus* L., Veronicaceae, Lamiales; euasterid I) and maize (*Zea maize* L., Poaceae; commelinids) (ordinal and familial classification based on the Angiosperm Phylogeny Group (APG), 2003).

The types of changes in architecture and floral morphology that occur in mutants of model organisms resemble those that distinguish species and that may generate new lineages through evolutionary time. It is pertinent to ask whether changes in developmental genes can account for natural diversity, and what the nature of these changes is. Genes that control development have been implicated in the evolution of novel phenotypes (reviewed in Baum, 1998; Doebley & Lukens, 1998; McSteen & Hake, 1998; Cronk, 2001; Shepard & Purugganan, 2002). In particular, comparative studies of the genetic changes responsible for morphological diversity, both at the subspecies level and between major lineages, have found that changes in the *cis*-regulatory regions of transcription factors are important in evolution (reviewed in Caroll, 2000). For example, the transcription factor *TEOSINTE BRANCHED 1* (*TB1*), controlling axillary meristem growth, was identified as a primary determinant of the morphological differences between domesticated maize and its wild relative teosinte (Doebley *et al.*, 1997). Although no evidence of selection was detected in the coding region of *TB1* alleles, sequence diversity of the 5'-flanking region was extremely low in domesticated maize compared to its wild relative teosinte, suggesting that changes in *cis*-regulatory regions, associated with changes in architecture, were selected for during the domestication of maize (Wang *et al.*, 1999). *Cis*-regulatory changes can also be important at higher taxonomic levels. For instance, sequence

variation at a few nucleotide positions in the highly conserved enhancer region of orthologous mammalian and avian *HOX* genes, implicated in modifications of axial morphology, was associated with spatial and temporal changes in expression during embryo development (Belting *et al.*, 1998). There is now a growing interest in expanding this knowledge to other species less amenable to genetic studies but displaying patterns of morphological variation that could be accounted for by changes in the expression of developmental genes.

The aim of this project is to investigate whether developmental genes controlling floral morphology and initially characterised in *Antirrhinum majus* (Lamiales, euasterid I), have a similar role in a distantly related plant lineage, the Leguminosae (eurosid I; APG, 2003). In addition to examining macro-evolutionary processes between distant plant lineages, candidate genes are contrasted in two closely related species within the Leguminosae which differ in floral morphology.

1.2 Organisation of reproductive structures in angiosperms

Flowering plants exhibit high levels of morphological and architectural variation despite being structurally simple. The development of parts occurs in meristematic regions where cells divide and differentiate. At these meristematic regions, organ primordia, producing leaves or floral organs, and secondary meristems (*e.g.* producing inflorescences) are formed. The indeterminate nature of plant growth allows for much morphological variation to be affected by changes in the fate of meristematic regions (McSteen & Hake, 1998). Recent advances in developmental genetics have led to the isolation of genes controlling meristem growth and identity. Mutations in those genes have been found to alter branching pattern (Carpenter & Coen, 1990), inflorescence structure (Bradley *et al.*, 1996) and floral organisation (Coen & Meyerowitz, 1991) to cite only a few examples.

Some of the most intensely studied developmental genes are those which are involved in floral development. Floral development begins with the transition from shoot vegetative meristem to inflorescence meristem at the flank of which determinate floral meristems form, differentiating to produce the perianth and reproductive organs. The organisation of the different organs within a flower is broadly invariant across angiosperms, where concentric regions are occupied by different floral organs in the following order beginning with the outermost whorl: sepals – petals – stamens – carpels. The genetic control of floral organ identity has been established in the distantly related *Antirrhinum* (euasterid I) and *Arabidopsis* (eurosid II) and is thought to be highly conserved in higher flowering plants (Lawton-Rauh *et al.*, 2000; Theiβen *et al.*, 2002). Transcription factors with a characteristic MADS-box domain, classified into A, B or C type, interact to determine floral organ identity (Coen & Meyerowitz, 1991). In wild type flowers, activity of class A genes is restricted to the outer whorls and A function alone defines sepals. The combined expression of A and B-class genes specifies petal identity. The conjunction of B and C-class genes specifies stamens, whereas C class expression alone results in the formation of carpels. Class A and C genes negatively regulate each other, so that in class A mutants, class C activity expands to the two outer whorls (Bowman *et al.*, 1991) (figure 1-1).

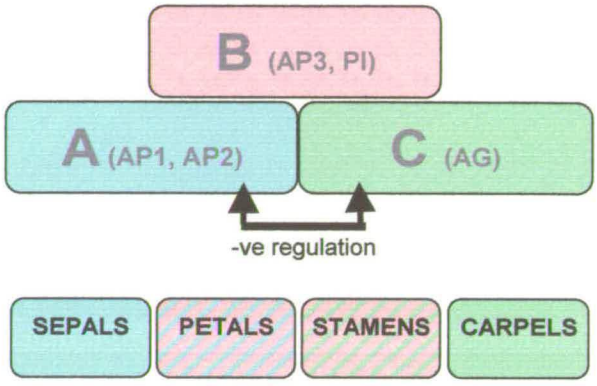


Figure 1-1. Simplified model of interaction of floral organ identity genes, first proposed by Coen and Meyerowitz (1991). The floral meristem is divided into three overlapping regions of homeotic gene activity resulting in four concentric whorls of floral organs. A-class genes, including *APETALA1* (*AP1*) and *APETALA2* (*AP2*) in *Arabidopsis* affect development in the outer two whorls (sepals and petals), B-class genes, such as *APETALA3* (*AP3*) and *PISTILLATA* (*PI*) affect development in whorls 2 and 3 (petals and stamens), and C-class genes such as *AGAMOUS* (*AG*) affect development in the inner two whorls (stamens and carpels). Some A and C-class genes have been found to be mutually antagonistic. This model has been extended with the discovery of other floral organ identity genes (*e.g.* redundant E class *SEPALLATA* genes specifying petal, stamen and carpel development; reviewed in Theißen *et al.*, 2002).

Within this conserved organisation, there is much variation in the presence, number and form of floral organs. One particular point of interest is the differentiation of organs within the same whorl, depending on positional cues. This differential growth, because it is often expressed along a defined axis, results in various patterns of floral symmetry.

1.3 Types of floral symmetry

The evolution of floral morphology has been of considerable interest, as it is interlinked with modes of pollination and therefore speciation. Changes in floral symmetry in particular are associated with specialised pollination mechanisms, which have promoted angiosperm diversification (Endress, 1999).

Floral symmetry is usually determined with respect to the centre of the receptacle, thereby only considering the flower in a two-dimensional perspective (Neal *et al.*, 1998). Three basic types of symmetry have been defined by Weberling (1989a):

1) translational, where repetition occurs along a straight line, *e.g.* successive whorl of similar floral organs
2) rotational, where a pattern is repeated twice or more over 360° around the principal axis through the centre
3) reflectional, where patterns are repeated as mirror images

In practice, the translational component is seldom taken into account when describing floral symmetry (Neal *et al.*, 1998). Although some inconsistency can be found in terminology, symmetrical flowers are commonly divided into two main categories: actinomorphic (regular, polysymmetric, or symmetrical) and zygomorphic (irregular, monosymmetric or bilaterally symmetrical). Actinomorphy is characterised by multiple planes of symmetry, and is a combination of rotational and reflectional symmetry, whereas zygomorphy only involves reflectional symmetry over one plane (figure 1-2).

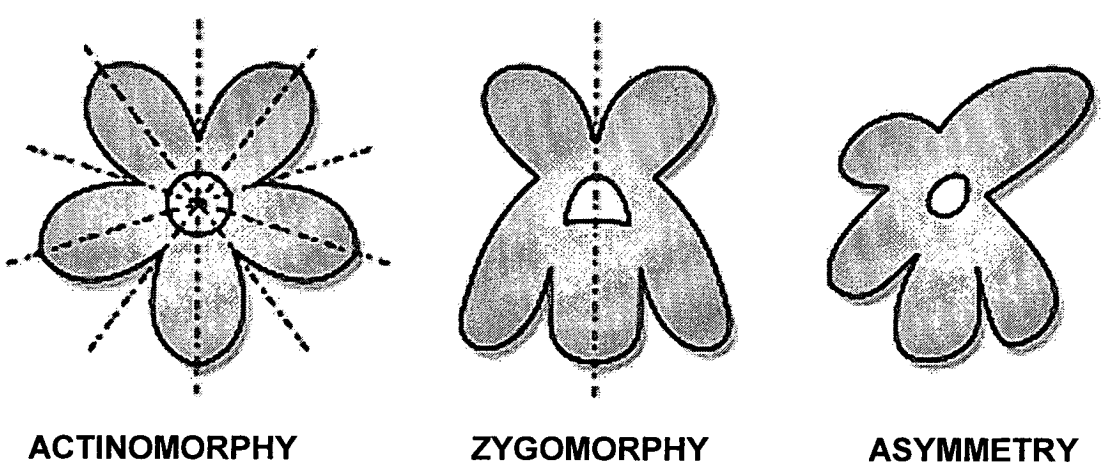


Figure 1-2. The two main types of floral symmetry: actinomorphy and zygomorphy in relation to corolla shape, compared with absence of symmetry (asymmetry); reproduced from Endress (2001).

Much variation exists within this broad descriptive framework, suggesting that different developmental mechanisms may be involved. For instance, zygomorphic development can be manifest at different stages of ontogeny. In most predominantly monosymmetric lineages, such as Scrophulariaceae *s.l.*, Orchidaceae, and subfamily Papilionoideae of the Leguminosae, zygomorphy is often evident during organ initiation, whereas in zygomorphic taxa derived from mainly radially symmetrical lineages, such as Ranunculaceae, bilateral symmetry is apparent only later in development as the organs enlarge (Tucker, 1999). Frequently, the designation of symmetry is based upon corolla shape, which may be simplistic as the symmetry of other floral organs may have important ecological significance (Neal *et al.*, 1998). Different patterns, or absence, of symmetry may occur between organ types within the same flower. For instance deflection of the style and/or anthers away from the axis of symmetry, known as enantiomorphy if the shift is lateral, is a common phenomenon (Jesson & Barrett, 2002). Deviations from radial symmetry may be facilitated by gravity (Weberling, 1989a; Neal *et al.*, 1998) as well as being controlled genetically (Luo *et al.*, 1996).

Zygomorphic flowers often develop in an asymmetric environment such as indeterminate racemose inflorescences *e.g.* Scrophulariaceae *s.l.*, or dense flower clusters (capitula) as found in the Asteraceae. This implies that a polarised environment may in many cases provide the cues necessary for floral dorso-ventral differentiation (Coen & Nugent, 1994). However, taxa which produce terminal zygomorphic flowers, such as *Schizanthus* (Solanaceae, Lamiales) are known. In Leguminosae, flowers are borne on diverse inflorescences that are variations on the indeterminate raceme (Weberling, 1989b).

1.4 Evolution of floral symmetry

The organisation of floral organs in concentric whorls is thought to be an advanced condition over spiral organisation and is a prerequisite for the evolution of zygomorphy. Early

flowering plants with a whorled phyllotaxy are believed to first have had regular flowers; irregularity is considered a derived condition (Neal *et al.*, 1998). The fossil record suggests that actinomorphy predates zygomorphy by around 30 to 40 million years (Crane *et al.*, 1995). It is commonly thought that zygomorphy has evolved independently numerous times, perhaps on as many as 25 separate occasions, and has contributed to the evolution of major angiosperm lineages such as Lamiales, Asteraceae, Leguminosae and Orchidaceae (Stebbins, 1974). Different lines of evidence support this multiple-gain hypothesis. First of all, angiosperm phylogenies reveal that zygomorphy occurs in highly divergent taxa and is more sporadically represented than actinomorphy (Neal *et al.*, 1998), implying that independent gains produce a more parsimonious scenario. In addition, the structural variety of bilaterally symmetrical flowers suggests that different mechanisms are implied in the evolution of zygormorphy in different groups. Bilateral symmetry is also viewed as a specialised adaptation to animal pollinators which are receptive to visual cues, and therefore selection would favour its repeated evolution (Giurfa *et al.*, 1999). Bilaterally symmetrical flowers became abundant in the Upper Cretaceous coinciding with the evolution of social insects. This association with specific pollinators underpinned the success and rapid radiation of diverse zygomorphic lineages (Dilcher, 2000). However, many arguments have been presented against the generality of the multiple-gain hypothesis. For instance, whereas actinomorphic mutants of normally zygomorphic species have been recorded in a variety of groups such as Orchidaceae and a wide range of eudicot clades (Rudall & Bateman, 2003), there is no indication of mutations producing bilateral symmetry in actinomorphic plants (Coen & Nugent, 1994; Donoghue *et al.*, 1998). This would suggest that the probability of losing zygomorphy is greater, at least when considering changes in developmental genes of large effect. This must be balanced against the cost in terms of pollination efficiency, which has been demonstrated experimentally (Giurfa *et al.*, 1999). There is no agreement as to the relative importance of these factors in the evolution of zygomorphy

(Coen & Nugent, 1994). Examining character evolution by mapping traits on phylogeny reconstructions may be contentious, especially if the phylogeny is based on morphology and is therefore not independent from the trait in question (Coen & Nugent, 1994). A phylogenetic approach must also take into consideration the relative likelihood of character change (Ree & Donoghue, 1999), as previously mentioned.

The question of evolution of zygomorphy is far from resolved. Elucidating the genetic control of zygomorphy in plants from different groups may provide a breakthrough in understanding its evolution. If similar genes are found to control zygomorphy in different taxa, this would suggest that either zygomorphy is more ancient than suspected, or that the same genes have been recruited more than once (Coen & Nugent, 1994).

1.5 Genetic control of floral symmetry in *Antirrhinum*

The first record of actinomorphic mutants was made in *Linaria vulgaris* L. (Veronicaceae, Lamiales) by Linnaeus (1749) who classified them in the separate genus *Peloria* (from the greek peloros meaning monstrous), although their similarity with *Linaria* was already acknowledged. In these mutants, the five petals resembled the single lower spurred petal of wild type. The term peloric was subsequently adopted to describe actinomorphic mutants. Peloric forms of *Antirrhinum majus* (Veronicaceae, Lamiales) and *Sinningia speciosa* (Lodd.) Hiern (Gesneriaceae, Lamiales) were also recognised by Darwin (1868). The control of zygomorphy was until recently understood from classical genetic experiments in *Antirrhinum majus*, which suggested that a few genes were involved (Stubbe, 1966). The recent characterisation of members of an active transposon family in *Antirrhinum majus* has made this species amenable to transposon mutagenesis experiments (Carpenter & Coen, 1990), a technique which directly links genes to their effect on phenotype.

The genetic basis of floral symmetry has been extensively examined in *Antirrhinum majus* (Luo *et al.*, 1996; Almeida *et al.*, 1997; Luo *et al.*, 1999). Wild type *Antirrhinum* flowers are pentamerous in the three outer whorls and strongly zygomorphic along the dorso-ventral axis. Zygomorphy is most pronounced in the petals and stamens which can be divided into three types according to their position: dorsal (adaxial), lateral and ventral (abaxial). All three types of petals have a distinctive shape as well as pigmentation and presence/absence of hairs, and differ in size with the dorsal lobes being the largest (figure 1-3). Although all stamen primordia are initiated, the dorsal stamen is aborted early in development. Two nuclear genes *CYCLOIDEA* (*CYC*) and *DICHOTOMA* (*DICH*) play a key role in establishing dorso-ventral differentiation of floral organs in *Antirrhinum* (Luo *et al.*, 1996; Luo *et al.*, 1999). Double mutants for both *CYC* and *DICH* have a fully radially symmetric phenotype characterised by ventralisation of the corolla lobes (*i.e.*, all lobes resemble the wild type phenotype of the ventral petal) and complete equal development of all five stamens (figure 1-3). *CYC* has the greatest affect on phenotype, with mutants showing a ventralisation of lateral regions, whereas *DICH* mutants show only weak departure from the wild type in the dorsal petals (figure 1-3).

Figure 1-3. *Antirrhinum majus* flowers of wild type and *dich*, *cyc* and *cyc/dich* mutant (E. Coen, John Innes Centre, Norwich). Loss of *CYC* function has a greater effect on phenotype than loss of *DICH*, although loss of both genes is required for a fully radial phenotype.

CYC and *DICH* are two closely related, partially functionally redundant transcription factors with overlapping expression patterns in the adaxial region of the developing *Antirrhinum* flower (Luo *et al.*, 1996; Luo *et al.*, 1999). Both genes are expressed prior to organogenesis in the dorsal region of the floral meristem and during the early stages of development affect growth rate and primordium initiation. During later stages, *CYC* expression can be detected only in the two dorsal petals and the adaxial staminode (Luo *et al.*, 1996), whereas *DICH* is restricted to the dorsal half of the dorsal petals (Luo *et al.*, 1999) (further details of expression patterns are given in chapter 4). Early expression of *CYC* affects primordium initiation and retards primordia growth in the abaxial region, whereas late expression affects organ morphology in a whorl-specific manner, causing abortion of the dorsal stamen but enlargement of the dorsal petals (Luo *et al.*, 1996). The role of *DICH* on floral morphology appears to be restricted to the elaboration of asymmetric dorsal petals (Luo *et al.*, 1999).

CYC and *DICH* are known to interact with other genes affecting the morphology of *Antirrhinum* flowers. For instance, ventralisation of the mutant flower suggests that *CYC* and *DICH* restrict the expression of gene(s) conferring ventral identity to the abaxial side of the flower (Almeida *et al.*, 1997). Such a gene, the MYB transcription factor *DIVARICATA* (*DIV*), was isolated in *Antirrhinum*, and its activity was shown to be restricted by both *CYC* and *DICH* (Almeida *et al.*, 1997; Galego & Almeida, 2002). The gene *RADIALIS* (*RAD*) is also suspected to interact with *CYC*, *DICH* and *DIV*. Current preliminary models suggest that *RAD* may be regulated by *CYC* and antagonises the expression of *DIV* in the lateral domain of the developing flower (E. Coen, pers. comm.).

CYC has a differential effect on the growth of organs in different whorls (Coen & Meyerowitz, 1991). This effect is dependent on the level of *CYC* activity, but is controlled by organ identity and not whorl position (Coen & Meyerowitz, 1991). In *ovulata* mutants, which have stamens in place of petals, the two dorsal stamens are aborted (Coen & Meyerowitz, 1991).

At later developmental stages, it appears that B function genes, in the absence of C function, interact with *CYC* to increase cell division, whereas combination of B and C function with *CYC* has the opposite effect (Luo *et al.*, 1996). Organ identity genes not only regulate the effect of *CYC* on cell division, but also affect the region of expression of *CYC* (Clark & Coen, 2002). For instance, ectopic expression of *CYC* was found in whorl 4 in *plena* mutants which have petaloids instead of carpels (Clark & Coen, 2002). However, what cues trigger the establishment of the dorso-ventral axis along which *CYC* and *DICH* are differentially expressed still remain to be uncovered (Clark & Coen, 2002).

1.6 *CYC* belongs to the TCP family of transcription factors

CYC and *DICH* belong to a family of putative transcription factors characterised by a basic helix-loop-helix (bHLH) DNA binding region (Cubas *et al.*, 1999a, Kosugi & Ohashi, 2002). This domain is referred to as the TCP domain after the first characterised members of this family *TEOSINTE BRANCHED 1* (*TB1*) in maize, *CYC* in *Antirrhinum* and *PROLIFERATING CELL FACTORS* (*PCF*s) in rice (Cubas *et al.*, 1999a). In maize, *TB1* affects the fate of axillary meristems by suppressing growth at the lower nodes and by promoting the development of female inflorescences at the upper nodes (Doebley *et al.*, 1997). In the wild relative of maize, teosinte, most meristematic nodes along the main stem produce elongated lateral branches which are terminated by male inflorescences, whereas female inflorescences are borne on secondary branches. Maize differs radically from teosinte by producing lateral branches, which are terminated by female inflorescences, at only a few nodes along the stem. It has been shown that differences in levels of expression of *TB1* are largely responsible for producing the distinctive phenotypes of maize and teosinte (Doebley *et al.*, 1997; Hubbard *et al.*, 2002). In rice, certain *PCF*s are known to control cell proliferation by binding of the TCP domain to promoter elements of *PROLIFERATING CELL NUCLEAR ANTIGENS* (*PCNA*) that control cell cycling

in meristematic regions (Kosugi & Ohashi, 1997). Like *PCFs*, *CYC* has been shown to modulate the transcription of cell cycle genes (Gaudin *et al.*, 2000).

Many other TCP genes have been isolated in a variety of taxa. In *Arabidopsis*, 24 members are known, some of which are expressed in floral meristems (Cubas *et al.*, 1999a; Cubas, 2002). Some members of this gene family, including *CYC*, *DICH* and *TB1* and their *Arabidopsis* homologues, but excluding rice *PCFs*, have another conserved region, known as the R domain, which is arginine-rich and is predicted to form a hydrophilic α helix that may be functionally important (Cubas *et al.*, 1999a). *CYC/TB1*-like genes are clearly present in a wide range of angiosperms including monocots and eudicots, share certain properties affecting cell growth and division, and may therefore be developmentally important in many species.

1.7 Role of *CYC* homologues in floral development

CYC-like genes have been implicated in modifications of floral symmetry in taxa closely related to *Antirrhinum*. Diverse genetic changes have underlied these morphological innovations. The first naturally occurring actinomorphic mutants to be characterised genetically are from polymorphic populations of *Linaria vulgaris* (Veronicaceae, Lamiales) (Cubas *et al.*, 1999b). The peloric mutants of *L. vulgaris* resemble in many respects the radial *Antirrhinum* mutants by having five rather than four functional stamens and a ventralised phenotype for both petals and stamens, suggesting a loss of function of *CYC*-like genes. The homologue of *Antirrhinum CYC*, *LCYC* has been isolated and implicated in the control of zygomorphy in *Linaria* (Cubas *et al.*, 1999b). However, loss of expression was not found to result from a genetic mutation, but was correlated with extensive methylation of *LCYC* (Cubas *et al.*, 1999b). In another close relative of *Antirrhinum*, *Mohavea*, evolutionary changes in floral corolla morphology and stamen abortion

correlate with an expansion of expression of both *CYC* and *DICH* orthologues from the dorsal into the lateral domain (Hileman *et al.*, 2003).

One of the fundamental questions regarding the evolution of floral symmetry is whether *CYC*-like genes are involved in the control of this trait beyond *Antirrhinum* and its close relatives. In the Asteraceae, the second largest family of flowering plants, zygomorphy has evolved independently from the Lamiales (Donoghue *et al.*, 1998). Nevertheless, *CYC*-like genes have been implicated in the production of zygomorphic flowers in *Senecio vulgaris* L. from this clade (Gillies *et al.*, 2002). In *Arabidopsis*, a species with radially symmetrical flowers, the homologue of *CYC*, *TCP1*, is expressed transiently in the adaxial region of axillary meristems, including floral meristems (Cubas *et al.*, 2001). As *TCP1* is expressed only in the very early stages of floral development, this may account in part for the lack of dorsoventral asymmetry in *Arabidopsis* (Cubas *et al.*, 2001). This early adaxial expression pattern, shared by distant species with different floral morphology, may represent an ancestral state that has been modified repeatedly to generate zygomorphic flowers (Cubas, 2002). To test this hypothesis, the role of *CYC* homologues is investigated here in the Leguminosae.

1.8 Evolution of floral symmetry in the Leguminosae

The Leguminosae is an important plant family where zygomorphy is believed to have evolved separately from the Lamiales (Stebbins, 1974; Donoghue *et al.*, 1998). With approximately 20,000 species, it is the third most species-rich angiosperm family, after two other predominantly zygomorphic families Asteraceae and Orchidaceae. This family is traditionally divided into three subfamilies: Caesalpinioideae, Mimosoideae and Papilionoideae. Whereas current molecular evidence supports the monophyly of the Papilionoideae and Mimosoideae with their derived floral characteristics, the Caesalpinioideae were found to be a diverse

assemblage of unrelated (paraphyletic) lineages which have diverged early in the history of the family (reviewed in Wojciechowski, 2003, see figure 1-4).

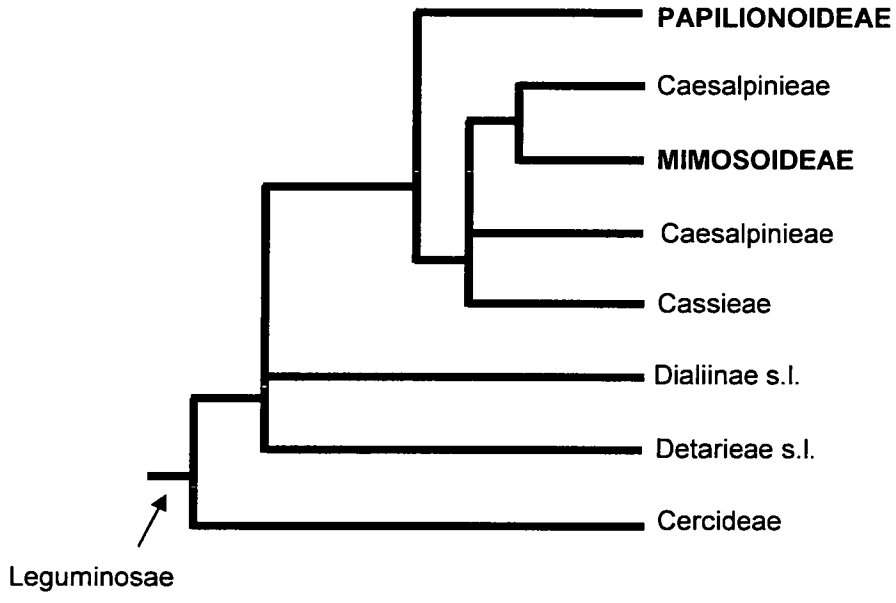
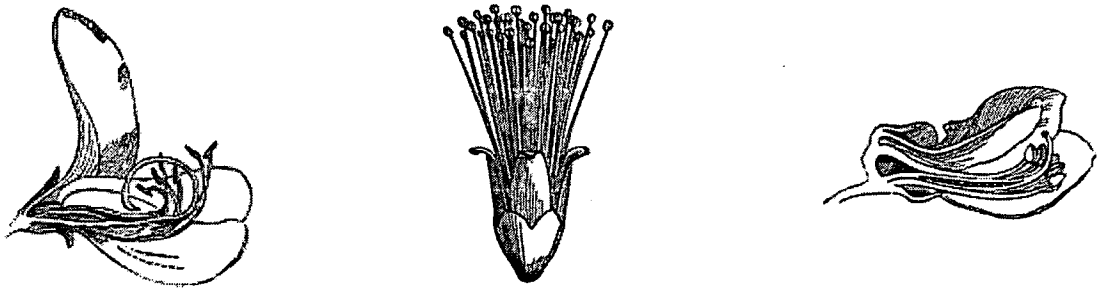


Figure 1-4. Representation of major legume lineages, showing the relationship of the monophyletic subfamilies Papilionoideae and Mimosoideae, and a grade of caesalpinioid tribes. Redrawn from Doyle and Luckow (2003) and Wojciechowski (2003); caesalpinioid tribes defined in Wojciechowski (2003).

The greatest number of species (*ca.* 14,000 species in 476 genera (Doyle & Luckow, 2003)) is found in the subfamily Papilionoideae. Although widely distributed and extremely diverse in habit and ecology, papilionoids are characterised by highly distinctive zygomorphic flowers with an enlarged dorsal (standard) petal, differentiated lateral (wing) and ventral (keel) petals housing the fertile parts, and imbricate aestivation with the reflexed adaxial petal outside the lateral petals in bud (figure 1-5). This specialised floral form, an adaptation to bee pollination, contrasts with that of the other two subfamilies Caesalpinioideae and Mimosoideae. Mimosoid flowers are typically actinomorphic, with reduced outer whorls and often numerous

showy stamens (figure 1-5). Caesalpinoids differ from papilionoids by having ascending imbricate aestivation (the lateral petals are outside the adaxial petal), and display much more variation in floral symmetry ranging from near-radial to zygomorphic. Many members of the Caesalpinioideae have reduced or absent floral parts (Tucker, 2003). However, within the basal-most lineage of the Leguminosae (Cercideae), the genus *Cercis* L. has "pseudo-papilionaceous" flowers (figure 1-5), which are believed to have evolved by convergence (Tucker, 2002a).



Genista
PAPILIONOIDEAE

Acacia retinoides
MIMOSOIDEAE

Cercis
CAESALPINIOIDEAE

Figure 1-5. Examples of flowers from the three subfamilies of the Leguminosae (from Watson & Dallwitz, 1992). Transverse sections along the median axis are shown for *Genista* and *Cercis*. *Genista* has typical papilionoid flowers, with a reflexed adaxial petal, and differentiated lateral and ventral petals enclosing the stamens and carpel. *Acacia retinoides*, like many mimosoid species, has a reduced radially symmetrical perianth and a proliferation of free stamens. Caesalpinoids are more variable in floral morphology, usually more open and with less differentiated petals than papilionoids. *Cercis*, shown here, has flowers which superficially resemble those of papilionoids.

Within the Papilionoideae, a few genera have flowers that differ from the distinctive entomophilous papilionoid form. In particular, a small number a taxa have open near radial flowers. Their traditional taxonomic position has been influenced by perceptions of evolutionary

advancement, particularly in floral characters. These include increasing petal and stamen fusion, and a progression from unspecialised open radial flowers to truly zygomorphic papilionoid flowers (Polhill, 1981; see examples figure 1-7). These atypical taxa have therefore been considered pleisiomorphic (primitive) members of the subfamily, even transitional between caesalpinioids and papilionoids, and were grouped together into two basal tribes, the Swartzieae and Sophoreae (Polhill, 1981). Recent molecular evidence, however, suggests that these unusual taxa are not related and that many are derived from within clades of typical papilionoid taxa (Pennington *et al.*, 2000; see figure 1-6). In addition, detailed morphological examination has shown that these taxa are morphologically diverse, and do not share any unifying features (Pennington *et al.*, 2000). The swartzioid clade, as defined from molecular phylogenies (*e.g.* Pennington *et al.*, 2001), was found to be sister to the rest of the Papilionoideae, and is characterised by genera with highly unusual flowers, a morphology that may be pleisiomorphic in this subfamily. These taxa frequently have a proliferation of free stamens (*i.e.* an increase in number compared to typical papilionoid flowers with 10 stamens), and have often lost the lateral and ventral petals.

From phylogenetic evidence, papilionoid taxa lacking dorso-ventral differentiation appear to have evolved repeatedly from zygomorphic ancestors. It was estimated that twelve independent instances of reversals from a zygomorphic papilionoid flower to a more radial form occurred during the evolution of the Papilionoideae (Pennington *et al.*, 2000; figure 1-6). Within certain clades, such as the genistoid or the dalbergioid clade, detailed molecular phylogenies have shown that the atypical near-radial flowers of taxa such as *Cadia* and *Dicraeopetalum* (genistoid), and *Etaballia*, *Inocarpus* and *Riedeliella* (dalbergioid) were all derived independently (Pennington *et al.*, 2000; Lavin *et al.*, 2001). Although unusual taxa are nested within relatively derived lineages, no atypical flowers are found in the major clades containing model legumes (*e.g.* Phaseoleae and Hologalegina).

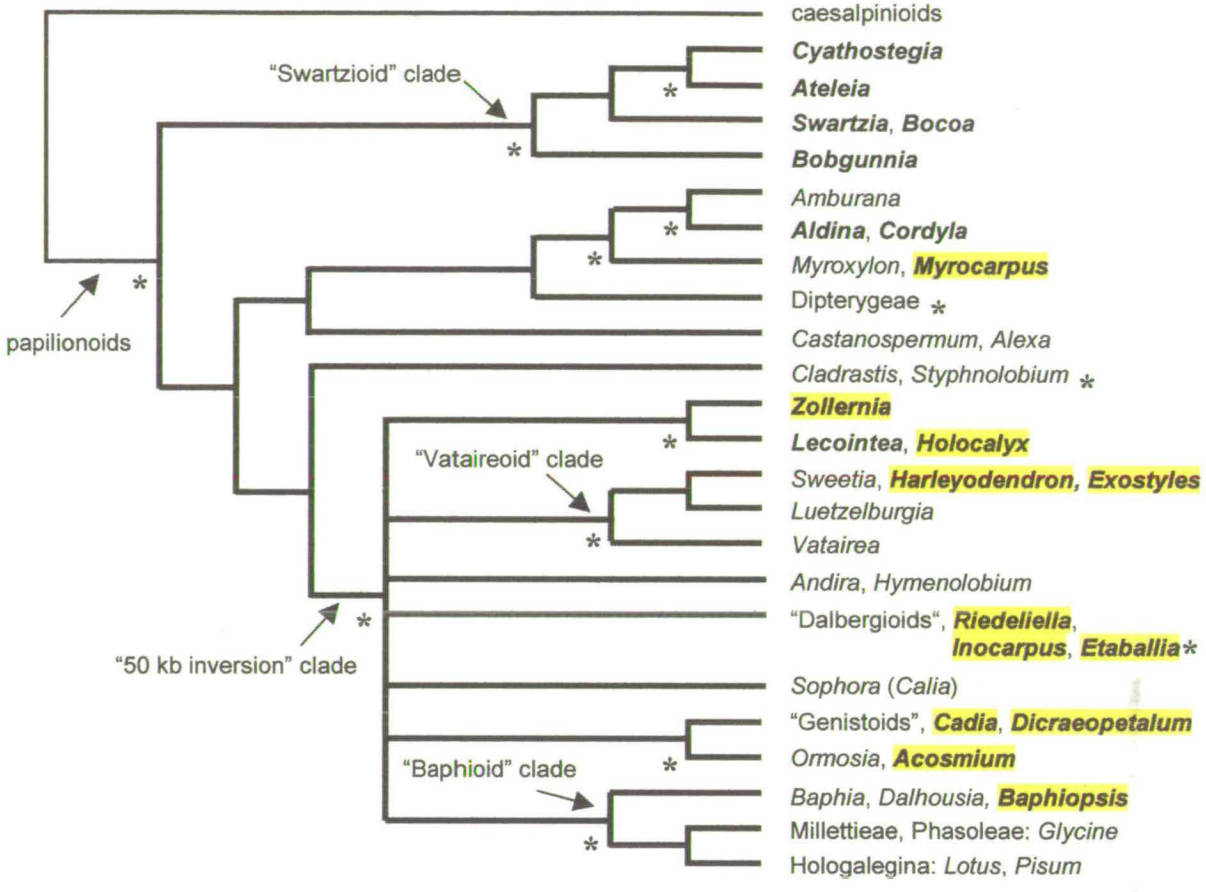


Figure 1-6. Current phylogeny of the Papilionoideae, redrawn from Wojciechowski (2003). Taxa with atypical non-papilionoid flowers (listed in Pennington *et al.*, 2000) are in bold, and in addition those with more or less radially symmetrically flowers are highlighted. The swartzioid clade is sister to other papilionoid lineages. It was estimated that twelve instances of reversals from zygomorphy to actinomorphy occurred in the Papilionoideae. * denotes clades with over 50% bootstrap support (from different sources of molecular data, summarised by Wojciechowski, 2003).

Figure 1-7. Floral diversity in the Papilionoideae. A-C: typical zygomorphic papilionaceous flowers, adapted to bee pollination. A: *Lupinus nanus*, B: *Lotus japonicus,* C: *Cytisus* sp., showing bee pollination. D-G: Papilionoid legumes with unusual open flowers. D: *Cyathostegia matthewsii*, E: *Swartzia pinnata* have reduced/lost corolla parts and a proliferation of stamens. F: *Cadia purpurea*, G: *Acosmium panamense*, have near radially symmetric flowers. Photos: A, F: D. White, RBGE; C: Q.Cronk, UBC Botanical Garden; B: S. Suehiro, Japan; D: G.P. Lewis, RBG Kew; E,: T. Pennigton, RBGE, G: L. Pauwels, Belgium.

1.9 Case study in the genistoid clade

These reversals from typical zygomorphic to radial open flowers provide a framework for investigating the control of floral symmetry in papilionoid legumes. This project focuses on *Cadia* Forsk., cultivated and flowering at the Royal Botanic Garden Edinburgh, a genus of seven species of small shrubs from Arabia, Madagascar and Eastern Africa. *Cadia* has atypical actinomorphic pendent flowers with unstable petal aestivation in solitary or few-flowered axillary racemes (see figure 1-7G and 1-8). These flowers produce abundant nectar, but no scent, suggesting these may be pollinated by birds (Pennington *et al.*, 2000). Although this genus has "always troubled botanists whether it ought to be referred to Papilionoiaceae or Caesalpiniaceae" (van der Maesen, 1970), recent molecular data suggest it is nested within the genistoid clade of Papilionoideae (Pennington *et al.*, 2001). Within this clade, the genus *Lupinus* L., with its typical zygomorphic papilionoid flowers in racemose inflorescences (see figures 1-7A and 1-8), makes an ideal comparative organism as it has been studied for agricultural purposes and is currently being developed for genetic transformation (Pigeaire *et al.*, 1997).

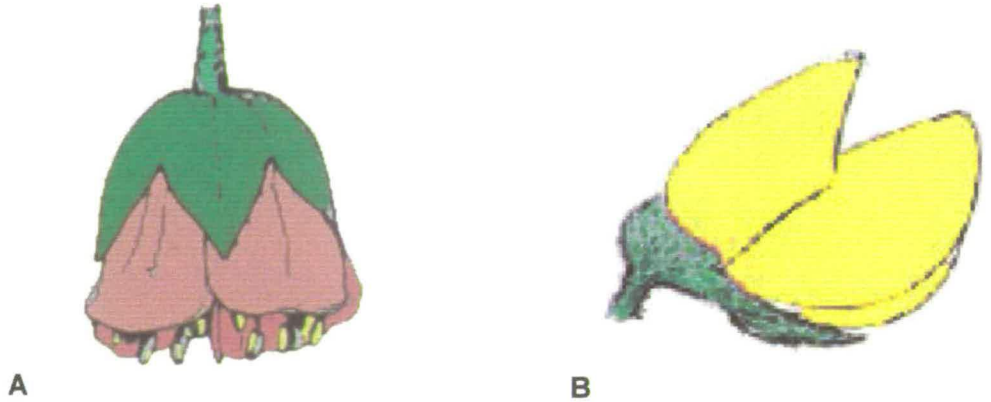


Figure 1-8. Lateral view of actinomorphic (A) and zygomorphic (B) flowers from two genera from the genistoid clade A. *Cadia purpurea* (from Polhill, 1981) and B. *Lupinus sabinii* (from the Rare Plants in Washington, University of Washington, http://courses.washington.edu/rarecare/RarePlantsinWashington.htm).

1.10 Aims of research

This project investigates the evolution and function of *CYC*-like genes in the Leguminosae, with particular emphasis on the subfamily Papilionoideae, where the vast majority of species has strongly zygomorphic flowers. This study aims to assess the importance of *CYC*-like genes in the repeated evolution of floral symmetry in the angiosperms. In addition, the hypothesis that changes in legume *CYC* expression may be responsible for the evolution of actinomorphic flowers in papilionoid taxa is tested by comparing the expression pattern of orthologous candidate genes in *Cadia purpurea*, with unusual radially symmetrical flowers, and *Lupinus nanus*, a small lupin with typical papilionoid flowers. The work presented in this thesis aims to:

1) Isolate *CYC*-like genes in an array of legume taxa and place them in a phylogenetic context (chapters 2 and 3). Taxa sampled include the two closely related species *C. purpurea* and *L. nanus* that differ in their floral symmetry.
2) Characterise the expression pattern of *CYC*-like genes in a typical papilionoid legume, *L. nanus*, and contrast the expression pattern of their homologues in *C. purpurea* (chapter 4).
3) Investigate sequence evolution of *CYC*-like genes in the genistoid clade, to which *Cadia* and *Lupinus* belong (chapter 5).
4) Further characterise legume *CYC* function in *Lupinus* by gene silencing using RNA interference (chapter 6).

CHAPTER 2: Phylogenomic investigation of *CYCLOIDEA*-like genes in the Leguminosae

2.1 INTRODUCTION

As functional gene studies expand from model organisms to related species, it becomes necessary to identify the functional counterparts of genes well-characterised in model species. The phylogenomic method proposes that orthology (*i.e.* the relationship of gene duplicates that have originated by speciation and therefore have a common descent) is a likely predictor of functional equivalence (Eisen,1998; Eisen & Wu, 2002). Modern phylogenetic techniques now often permit robust determination of orthology relations of genes. In particular, implementation of more realistic models of sequence evolution by maximum likelihood or Bayesian approaches may provide greater accuracy in tree reconstruction (Holder & Lewis, 2003).

A phylogenetic approach has been used to investigate orthologues of *Antirrhinum CYCLOIDEA* (*CYC*) in the Leguminosae. Prior to this study, putative *CYC*-like genes were isolated by Da Luo (Shanghai Institute of Plant Physiology) in the model legumes *Lotus japonicus*, soybean (*Glycine max*) and pea (*Pisum sativum*). In the case of *L. japonicus*, two of these genes were found to be expressed adaxially in the early stages of flower development (D. Luo, unpublished data). This study aims to expand these findings to other taxa from other major papilionoid groups such as the dalbergioid and genistoid clades as well as basal lineages (as defined by Wojciechowski, 2003) where most of the papilionoid floral morphological variation lies. Papilionoid species with unusual flower morphology were sampled here, such as *Acosmium subelegans* (Mohl.) Yakovlev and *Cadia purpurea* (Picc.) Aiton, from the genistoid clade, with radially symmetrical flowers, and *Swartzia jorori* Harms, from the basal papilionoid grade, which has no lateral or ventral petals (described in Pennington *et al.*, 2000). Inclusion of

papilionoid taxa with atypical flowers in a phylogenetic study of candidate genes for the control of floral symmetry is useful for understanding the origin of derived modifications in this trait. In addition, a few representatives from the other two subfamilies, Caesalpinioideae and Mimosoideae, were included in this study, with one species from the basal-most clade in this family Cercideae, *Cercis griffithii* Boiss. (Wojciechowski, 2003). The inclusion of a basal legume such as *Cercis* may provide a framework for understanding the pleisiomorphic (ancestral) condition of *CYC*-like genes in this family.

In view of functional analyses, particular emphasis was placed in identifying homologues of the two *CYC*-like copies from *L. japonicus* in a taxon from the genistoid clade with unusual near-radially symmetrical flowers (*Cadia; C. purpurea*) and a close relative of *Cadia* with typical zygomorphic flowers, *Lupinus* (*L. nanus*). Based on preliminary expression data in *L. japonicus*, these are prime candidates for the control of floral symmetry in legumes.

This study was divided into three parts: a survey of putative *CYC*-like genes in an array of legume taxa using PCR, then placement of key legume sequences in the context of the TCP gene family, and finally a detailed phylogenetic analysis of *CYC*-like genes in members of the subfamily Papilionoideae. The main results were published in Plant Physiology in March 2003 (Citerne *et al.*, 2003; appendix 8).

PART 1: PCR SURVEY OF *CYC*-LIKE GENES IN LEGUMINOSAE

2.2 MATERIALS AND METHODS

2.2.1 Taxon sampling and DNA extraction

Samples were chosen to represent the taxonomic range of the Leguminosae, with multiple representatives of the three subfamilies Caesalpinioideae, Mimosoideae and Papilionoideae (taxa listed in table 2-1). Particular emphasis was placed on sampling representatives from all major papilionoid groups defined by current molecular phylogenetic evidence (Doyle *et al.*, 1997; Hu *et al.*, 2000; Kajita *et al.*, 2001; Pennington *et al.*, 2001; summarised in Wojciechowski, 2003; figure 2-1). Genomic DNA was extracted from fresh or silica dried leaf material following a small-scale 2X CTAB procedure modified from Doyle and Doyle (1987) (details of protocol given in appendix 1A). Previously extracted DNA was available for *Dialium guianense* (R.T. Pennington, Royal Botanic Garden Edinburgh (RBGE)), *Inga nobilis* (J. Richardson, RBGE) and *Pisum sativum* (J. Hofer, John Innes Centre (JIC) Norwich). DNA quality was tested by PCR of the chloroplast gene *trn*L which is known to amplify in the taxa examined using universal primers (Pennington *et al.*, 2001).

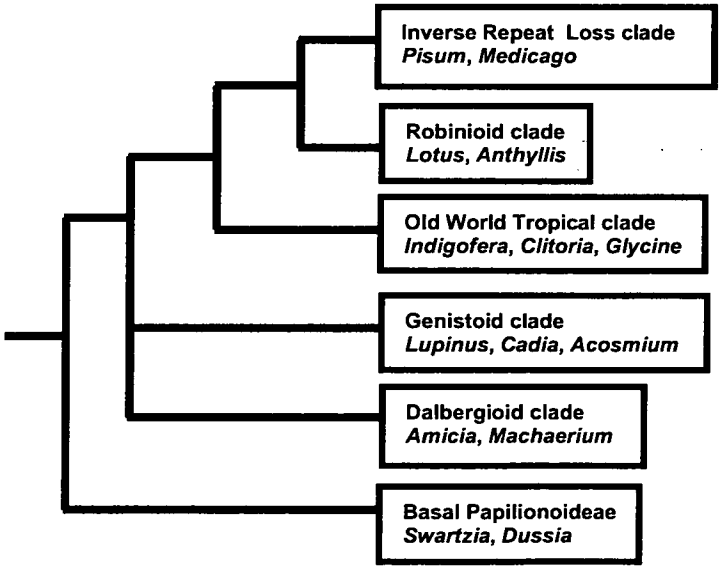


Figure 2-1.
Schematic representation of the relationship of some of the major groups in the Papilionoideae as defined by current molecular evidence (Doyle *et al.*, 1997; Hu *et al.*, 2000; Kajita *et al.*, 2001; Pennington *et al.*, 2001), with representative taxa used in the legume *CYC* sequence analyses.

Table 2-1. List of taxa included in the PCR survey of *CYC*-like genes using primers LEGCYC_F1 and LEGCYC_R1. Taxa are listed according to subfamily (Caesalpinioideae, Mimosoideae, Papilionoideae). Within the Papilionoideae, the major clades are shown (* names follow the nomenclature of Pennington *et al.*, 2001) based on recent phylogenetic evidence (Doyle *et al.*, 1997; Hu *et al.*, 2000, Kajita *et al.*, 2001; Pennington *et al.*, 2001). ** Source number refers to either RBGE living collection number (*e.g.* 1996 0942A) or collector's voucher number from wild collections (*e.g.* R.T. Pennington 473), with the exception of *Pisum sativum* DNA from genetic line 399 grown at the JIC. All herbarium vouchers at RBGE.

SUBFAMILY – clade*	Taxon	Source **	Location
CAESALPINIOIDEAE	*Ceratonia oroethauma* (Hillc.) Lewis & Verdc.	1996 0942A	Oman
	Sclerolobium paniculatum Vogel	R.T. Pennington 473	Goiás, Brazil
	Diptychandra aurantica (Mart.) Tul.	R.T. Pennington 484	Goiás, Brazil
	Dimorphandra mollis Benth.	R.T. Pennington 472	Goiás, Brazil
	Dialium guianense (Aubl.) Sandw.	R.T. Pennington 639	Napo, Ecuador
	Hymenaea courbaril L.	R.T. Pennington 843	La Paz, Bolivia
	Cercis griffithii Boiss.	1969 1039	Afghanistan
	Chamaecrista glandulosa L.	R.T. Pennington 828	La Paz, Bolivia
MIMOSOIDEAE	*Inga nobilis* Willd.	T.D. Pennington 16480	Peru
	Acacia farnesiana (L.) Willd.	1997 0065A	Costa Rica
	Enterolobium contortisiliquum (Vell.) Morong	1998 0256	Brazil
	Dichrostachys cinerea (L.) Wight & Arn.	1997 0193A	Yemen
	Pithecellobium dulce (Roxb.) Benth.	1999 1147	Honduras
	Hesperalibizia occidentalis (Brandegee) Barneby & J.M. Grimes	1999 1145	Mexico
	Samanea saman (Jacq.) Merr.	1999 1148	Honduras
	Zapoteca tetragona (Willd.) H.M. Hernandez	1999 1149	Guatemala
PAPILIONOIDEAE			
Inverse Repeat Loss clade*	*Pisum sativum* L.	Line 399	UK: JIC Norwich, cultivated
	Lathyrus grandiflorus Sibth. & Sim.	1944 0032A	UK: RBGE, cultivated
Robinioid clade*	*Anthyllis hermanniae* L.	1975 1501	Mediterranean
	Lotus berthelotii Masf.	1978 0702B	Canary Islands
	Coursetia maraniona M. Lavin	R.T. Pennington 958	Amazonas, Peru
Old World Tropical clade*	*Indigofera pendula* Franch.	1991 0547A	China
	Clitoria sp.	R.T. Pennington 990	San Martín, Peru
	Desmodium sp.	R.T. Pennington 965	San Martín, Peru
	Lonchocarpus atropurpureus Benth.	R.T. Pennington 799	Amazonas, Peru

Genistoid clade*	*Cadia purpurea* (Picc.) Aiton	1994 2001A	Yemen
	Acosmium subelegans (Mohl.) Yakovlev	S. Bridgewater 358	Mato Grosso do Sul, Brazil
	Ormosia amazonica Ducke	R.T. Pennington 645	Napo, Ecuador
	Bowdichia virgilioides Kunth	R.T. Pennington 477	Goiás, Brazil
	Lupinus sp.	R.T. Pennington 815	Piura, Peru
	Lupinus nanus Doug. Ex Benth.	-	UK: Sutton's Seeds, cultivated
Dalbergioid clade*	*Machaerium scleroxylon* Tul.	1999 0888A	Brazil
	Aeschynomene sp.	R.T. Pennington 656	Loja, Ecuador
	Amicia glandulosa Kunth	R.T. Pennington 654	Loja, Ecuador
	Platymiscium sp.	R.T. Pennington 692	Antioquia, Colombia
Basal Papilionoideae*	*Dussia macroprophyllata* Harms	R.T. Pennington 597	Heredia, Costa Rica
	Ateleia guaraya Herzog	R.T. Pennington 904	Santa Cruz, Bolivia
	Swartzia jorori Harms	R.T. Pennington 938	Santa Cruz, Bolivia

2.2.2 Primer design

To amplify *CYC*-like genes in members of the Leguminosae, primers were designed to match the most conserved regions of the TCP and R domains, the defining elements of *CYC*-like genes (figure 2-2). These regions were identified by comparison of one sequence from the model legumes *Lotus japonicus* (Lotus japonicus 2) and *Glycine max* (Soya 1) provided by D. Luo, and sequences from *Arabidopsis TCP12* and *TCP1* (nomenclature of Cubas *et al.*, 1999a), *Antirrhinum majus CYC* and *DICH*, and maize *TB1* (Genbank accession numbers given in appendix 3). Sequences of primers LEGCYC_F1 (5'-TCA GGG SYT GAG GGA CCG -3') and LEGCYC_R1 (5'- TCC CTT GCT CTT GCT CTT GC -3') matched exactly the sequence of this region in *L. japonicus* and *G. max*.

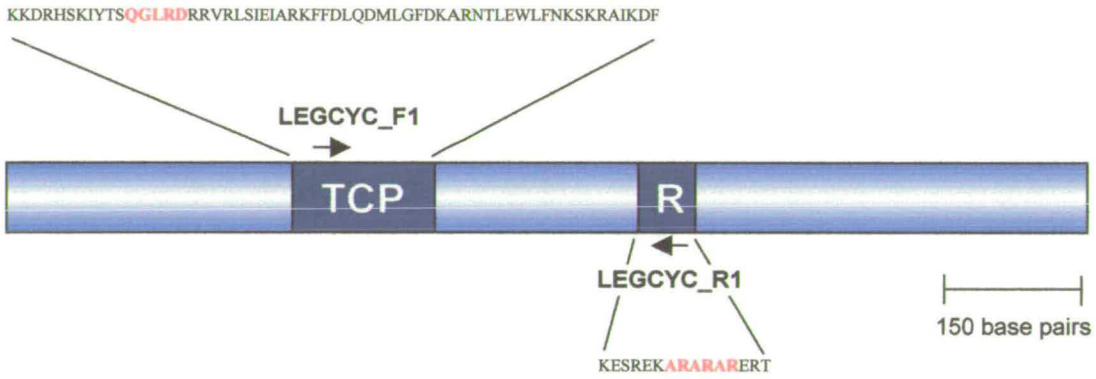


Figure 2-2. Representation of the legume *CYC*-like open reading frame (ORF), based on a cDNA sequence from *Lotus japonicus* (Lotus japonicus 2; D. Luo, pers. comm.), with sequences of the conserved TCP and R domains shown. Sequence in red/bold mark the priming location of the legume *CYC* primers LEGCYC_F1 and LEGCYC_R1.

2.2.3 PCR conditions

50µl PCR mix comprised sterile distilled water, X10 NH_4 polymerase buffer, $MgCl_2$ (2.5mM), dNTPs (20µM), primers LEGCYC_F1 and LEGCYC_R1 (0.5µM each), 1 unit *Taq* polymerase, and 20 – 30 ng genomic DNA. PCR amplifications were carried out using Bioline

Taq and reagents (Bioline, London NW2, UK). PCR conditions consisted of an initial denaturation step at 94°C (3 minutes), followed by 30 cycles of: denaturation at 94°C (1 minute), annealing at 50-55°C (30 seconds) and extension at 72°C (30 seconds), followed by a final extension step 72°C (5 minutes). PCR products (3 µl) were separated by electrophoresis on a 2% agarose gel for 2 ½ hours at 80V.

2.2.4 Cloning and sequencing

Nucleotide sequences from cloned PCR products amplified with primers LEGCYC_F1 and LEGCYC_R1 were obtained from a subset of the taxa listed in table 2-1, including three caesalpinioid, one mimosoid and 13 papilionoid species. PCR products from the following taxa were cloned using TOPO TA Cloning Kit for Sequencing (Invitrogen Ltd, Paisley, UK).

- Caesalpinioideae: *Ceratonia oreothauma, Dialium guianense, Cercis griffithii*
- Mimosoideae: *Zapoteca tetragona*
- Papilionoideae: *Dussia macroprophyllata, Swartzia jorori* (basal papilionoid), *Amicia glandulosa, Machaerium scleroxylon* (dalbergioid), *Cadia purpurea, Acosmium subelegans, Lupinus sp.* and *Lupinus nanus* (genistoid), *Clitoria* sp., *Indigofera pendula* (old world tropical), *Anthyllis hermanniae, Lotus berthelotii* (robinoid), *Pisum sativum* (inverse repeat loss clade); see figure 2-1 for relationships of the major clades in the Papilionoideae.

Prior to cloning, PCR products were purified using Qiaquick PCR Purification kit (Qiagen Ltd, Dorking, Surrey, UK) to remove primer-dimers from the reaction. After selection of clones containing the desired insert by PCR, plasmid DNA was extracted and purified using Qiagen Spin Miniprep kit (Qiagen Ltd, Dorking, Surrey, UK) and sequenced using the universal

M13 primers (Invitrogen Ltd, Paisley, UK). To ensure that all products amplified by primers LEGCYC_F1-LEGCYC_R1 were isolated from the two main taxa of interest, *C. purpurea* and *L. nanus*, 36 and 40 cloned PCR products from each reaction were sequenced respectively. Dye-terminator cycle sequencing was carried out using Thermosequenase II (Amersham Pharmacia, Buckinghamshire, UK). Samples were analysed on an ABI model 377 Prism Automatic DNA sequencer.

2.2.5 Confirmation and expansion of results

2.2.5a Degenerate primers

Degenerate primers were designed in an attempt to isolate other *CYC*-like sequences that may not have been amplified with the general primers LEGCYC_F1 and LEGCYC_R1. These were based on a combination of amino acid sequences of the TCP and R domain and apparent codon bias to reduce degeneracy. Two forward primers were designed to bind to regions within the TCP domain: LEGCYC_F2: 5'- GCI MGI AAG TTC TTY CTI CAR GAT G -3', LEGCYC_F4: 5'- CTT YGA TCT HCA RGA CAT GYT RGG RTT YGA YAA -3', and one reverse primer binding to the R domain: LEGCYC_R2: 5'- GTY CKY TCC CTS GCY CKY GCT CTY GC -3' (location of primers shown in appendix 2). These primers were tried on genomic DNA from *C. purpurea*, *Lupinus* sp., and *P. sativum*. The PCR mix was as above (section 2.2.3) with the exception of the final primer concentration, which was increased tenfold (5μM). PCR conditions were optimised to increase the likelihood of primers binding to an array of templates with 5 initial cycles with a low annealing temperature of 45°C for 30 seconds, followed by 30 additional cycles with the annealing temperature raised to 55°C. To allow larger products to be amplified, extension time was increased to 1 minute 30 seconds. These products were cloned and sequenced as described above (section 2.2.4).

2.2.5b Survey of *CYC*-genes in a basal caesalpinioid legume: *Cercis griffithii*

In addition to using primers LEGCYC_F1 and LEGCYC_R1 on *C. griffithii* genomic DNA, other primer combinations were tested to increase the chance of amplifying *CYC*-like genes in this species. Two other forward primers binding to the TCP domain, a general non-degenerate forward primer LEGCYC_F3 (5'- CAA GAC ATG YTA GGG TTT GAC -3') and the degenerate forward primer LEGCYC_F4 (described in section 2.2.5a), were used in combination with the reverse primer LEGCYC_R1. Products from LEGCYC_F3-LEGCYC_R1 amplifications were cloned and sequenced. Sequences were compared with those isolated in *C. griffithii* using primers LEGCYC_F1-LEGCYC_R1.

2.3 RESULTS

2.3.1 PCR survey

Results of the PCR survey using primers LEGCYC_F1 and LEGCYC_R1 are summarised in table 2-2. Examples of amplification results are shown in figure 2-3. Primers worked best in taxa from the Papilionoideae, amplifying multiple products in most taxa surveyed from this subfamily, ranging from members of the basal-most clades of the Papilionoideae (*e.g. Swartzia jorori*) to those from more derived clades. In the Caesalpinioideae and Mimosoideae, the primers either failed to amplify any product, or usually amplified only a single product, with no correlation with systematic relationships or floral morphology. Amplification in some mimosoid taxa suggests that *CYC*-like genes are present in this subfamily, whose members have actinomorphic flowers. The discrepancy in the number of visible PCR products between papilionoids and the other two subfamilies may reflect a bias in primer design, which was based on sequences from model legumes (*Lotus japonicus*, *Glycine max*) that are derived elements of the Papilionoideae.

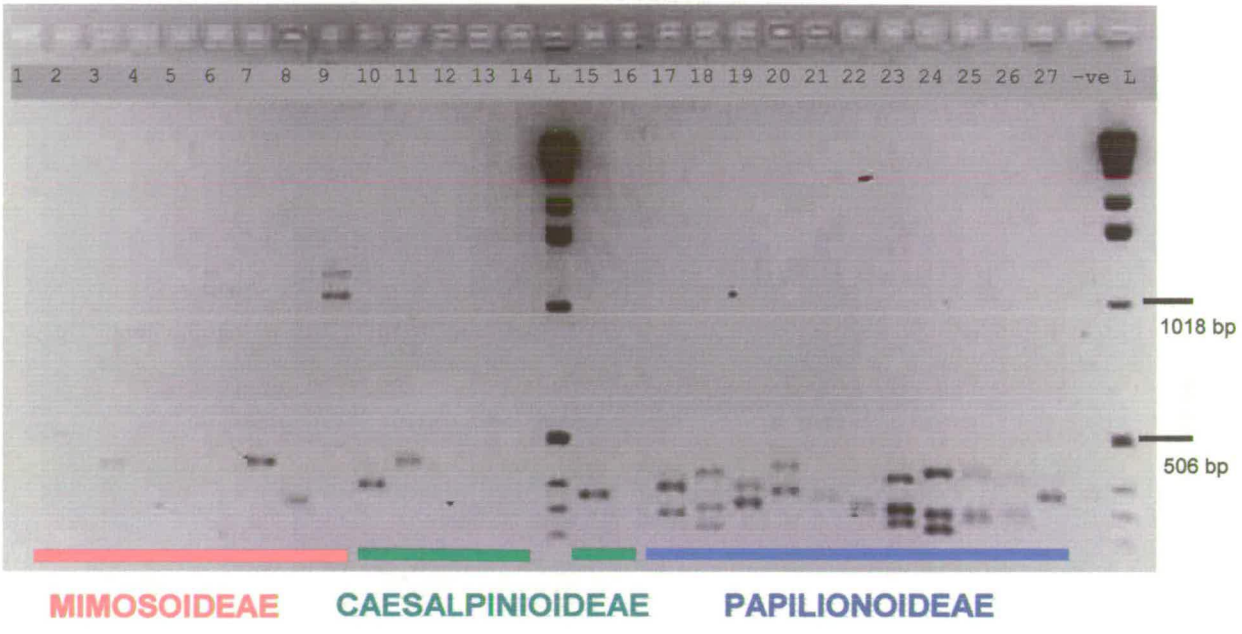


Figure 2-3. PCR products (3µl load), amplified using primers LEGCYC_F1 and LEGCYC_R1 in a range of legumes, separated on a 2% agarose gel for 2 ½ hours at 80V. Products are run against a 1 kb ladder (L). –ve: negative control. Taxa corresponding to each lane are given in table 2-2.

Subfamily - clade	Lane	Taxon	Number of bands
CAESALPINIOIDEAE	10	*Ceratonia oroethauma*	1
	12	*Sclerolobium paniculatum*	0
	13	*Diptychandra aurantica*	0
	14	*Dimorphandra mollis*	0
	11	*Dialium guianense*	1
	15	*Hymenaea courbaril*	1
	16	*Chamaecrista glandulosa*	0
	ns	*Cercis griffithii*	2
MIMOSOIDEAE	1	*Calliandra haematocephala*	0
	2	*Acacia farnesiana*	0
	3	*Enterolobium contortisiliquum*	1
	4	*Dichrostachys cinerea*	0
	5	*Pithecellobium dulce*	0
	6	*Hesperalibizia occidentalis*	0
	7	*Samanea saman*	1
	8	*Zapoteca tetragona*	1
	9	*Inga nobilis*	(2 , >1kb)
PAPILIONOIDEAE			
Inverse Repeat Loss Clade	ns	*Pisum sativum*	2
	ns	*Lathyrus grandiflorus*	(1, >1kb)
Robinioid	ns	*Anthyllis hermanniae*	2
	ns	*Lotus berthelotii*	2
Old World Tropical	ns	*Indigofera pendula*	2
	ns	*Clitoria* sp.	3
	ns	*Desmodium* sp.	1
	ns	*Lonchocarpus atropurpureus*	1
	ns	*Coursetia maraniona*	2
Genistoid	23	*Cadia purpurea*	3
	24	*Acosmium subelegans*	3
	26	*Ormosia amazonica*	2
	25	*Bowdichia virgilioides*	2
	27	*Lupinus* sp.	1
	ns	*Lupinus nanus*	1
Dalbergioid	19	*Machaerium scleroxylon*	2
	22	*Aeschynomene* sp.	2
	20	*Amicia glandulosa*	2
	21	*Platymiscium* sp.	1
Basal	17	*Dussia macroprophyllata*	2
	ns	*Ateleia guaraya*	1
	18	*Swartzia jorori*	3

Table 2-2. Results of the PCR survey using primers LEGCYC_F1 and LEGCYC_R1 on an array on taxa from the three subfamilies of the Leguminosae. The presence and number of bands visible on a 2% agarose gel run for 2 ½ hours at 80V is given for each taxa. The lane number refers to figure 2-3, some products are not shown (ns). Products much larger than 500 bp are given in parentheses.

2.3.2 Sequence data

2.3.2a Sequence survey using LEGCYC_F1-LEGCYC R1

Thirty nine distinct sequences with a TCP and R domain were obtained from cloned products amplified using primers LEGCYC_F1-LEGCYC_R1 in 17 different taxa. Sequences obtained from the same genomic DNA with no more than four nucleotide mismatches were considered to represent allelic variation or PCR error. The number of sequence types per taxon ranged from one to four, with only one sequence type isolated from non-papilionoid taxa, with the exception of *Cercis griffithii*. However, basal papilionoid taxa, such as *Swartzia jorori* and *Dussia macroprophyllata*, had multiple copies comparable in number with more derived papilionoid species (see table 2-3 for summary and GenBank accession numbers).

Fragment length was highly variable and ranged from 274 base pairs (bp) (Pisum 1) to 427 bp (Clitoria 1), with a mean length of 334.15 bp (± 40.2). These fragments were also highly variable in sequence at the amino acid and nucleotide level, with numerous substitutions and insertion-deletion (indel) events in the region between the TCP and R domains.

Taxon	No. of *CYC*-like sequence types amplified by primers F1-R1	Sequence name	GenBank accession number
Ceratonia oreothauma	1	Ceratonia 1	AY225810
Dialium guianense	1	Dialium 1	AY225811
Cercis griffithii	2	Cercis 1	-
		Cercis 2	-
Zapoteca tetragona	1	Zapoteca1	AY225812
Dussia macroprophyllata	3	Dussia 1	AY225845
		Dussia 2	AY225846
		Dussia 3	AY225847
Swartzia jorori	3	Swartzia 1	AY225848
		Swartzia 2	AY225849
		Swartzia 3	AY225850
Amicia glandulosa	2	Amicia 1	AY225843
		Amicia 2	AY225844
Machaerium scleroxylon	2	Machaerium1	AY225841
		Machaerium2	AY225842
Cadia purpurea	4	Cadia 1	AY225825
		Cadia 2	AY225826
		Cadia 3	AY225827
		Cadia 4	AY225828
Acosmium subelegans	3	Acosmium 1	AY225829
		Acosmium 2	AY225830
		Acosmium 3	AY225831
Lupinus sp.	3	Lupinus sp. 2	AY225833
		Lupinus sp. 3	AY225834
		Lupinus sp. 4	AY225835
Lupinus nanus	2	Lupinus nanus 2	AY225837
		Lupinus nanus 3	AY225838
Clitoria sp.	3	Clitoria 1	AY225822
		Clitoria 2	AY225823
		Clitoria 3	AY225824
Indigofera pendula	3	Indigofera 1	AY225819
		Indigofera 2	AY225820
		Indigofera 3	AY225821
Anthyllis hermanniae	3	Anthyllis 1	AY225814
		Anthyllis 2	AY225815
		Anthyllis 3	AY225816
Lotus berthelotii	2	Lotus berthelotii 1	AY225817
		Lotus berthelotii 2	AY225818
Pisum sativum	1	Pisum 1	AY225813

Table 2-3. Number of sequence types with a TCP and R domain obtained from cloned PCR products amplified using primers LEGCYC_F1 and LEGCYC_R1. GenBank accession numbers corresponding to the partial gene nucleotide sequence are given. Two other *CYC*-like sequences were obtained with different primers from *Lupinus* species and included in the phylogenetic analyses (part 3, this chapter): Lupinus sp. 1 (AY225832) amplified with primers LEGCYC_F2 and LEGCYC_R2 (described in section 2.2.3a), and L. nanus 1 (AY225836) amplified with locus specific primers (see chapter 3).

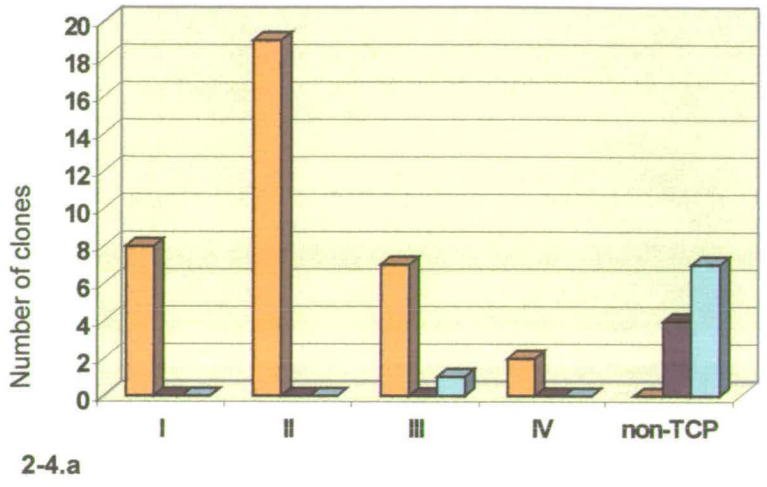
2.3.2b Saturation cloning

Among the forty sequences of cloned PCR products amplified by LEGCYC_F1-LEGCYC_R1 in *L. nanus*, fourteen did not have a TCP and R domain, and the remainder belonged to only two distinct *CYC*-like sequence types. In *C. purpurea*, four *CYC*-like sequence types were found, one of which (Cadia 4) occurred in lesser abundance in the pool of PCR products. Sequence variation between clones of the same "type" never exceeded four nucleotides in number and was therefore not considered to represent different loci. This low level of variation may be allelic in nature or may be an artefact caused by errors in the replication process during PCR. These results are summarised in figure 2-4, along with results obtained using degenerate primers (section 2.3.2c).

2.3.2c Degenerate primers

Results using highly degenerate primers showed that this approach can be problematic. Although many PCR products within the expected size range were amplified, many of these products did not have a TCP or R domain.

BLAST searches of these sequences revealed that one possible reason for this problem was that the reverse primer, designed on the arginine – guanine repeats characteristic of the 45 bp-long R domain, shared similarities with a motif found in the chloroplast *atp*B gene as well as the actin genes. Nevertheless, some TCP genes were isolated using degenerate primers in *C. purpurea* and *Lupinus* sp, including a product in *Lupinus* sp. (Lupinus sp. 1, GenBank accession number AY225832), which was not amplified by primers LEGCYC_F1 and LEGCYC_R1 (see figure 2-4).



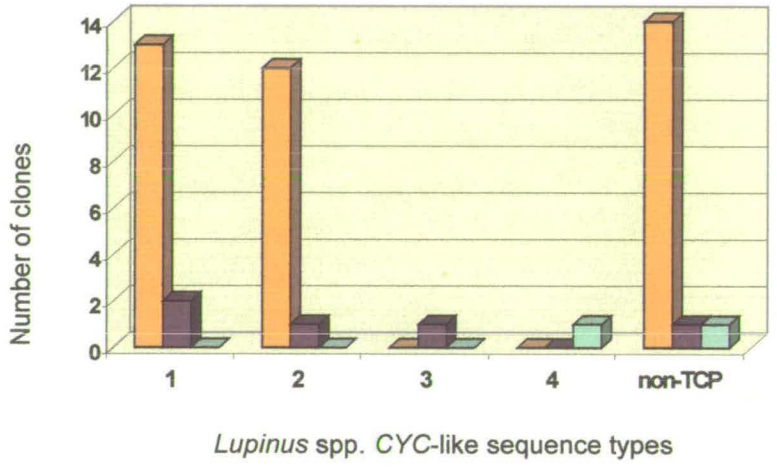


Figure 2-4. Number of clones sequenced from *Cadia purpurea*, *Lupinus* sp. and *Lupinus nanus* from PCR reactions using different primer combinations, including the highly degenerate primers F2, F4 and R2. Cloned PCR products have been grouped into different *CYC*-like sequence types (*i.e.* with a TCP and R domain), plus those which are not TCP genes. Numbers referring to sequence type do not imply homology between *C. purpurea* and *Lupinus* sequences. **2-4a.** ☐ *C. purpurea* F1-R1, ■ *C. purpurea* F2-R2, ☐ *C. purpurea* F4-R2. Sequence type I = Cadia 1, sequence type II = Cadia 2, sequence type III = Cadia 3, sequence type IV = Cadia 4. **2-4b.** ☐ *L. nanus* F1-R1, ■ *Lupinus* sp. F1-R1, ☐ *Lupinus* sp. F2-R2. Sequence type 1 = Lupinus nanus 2, Lupinus sp. 2 ; sequence type 2 = Lupinus nanus 3, Lupinus sp. 3 ; sequence type 3 = Lupinus sp. 4 ; sequence type 4 = Lupinus sp. 1 (sequence names listed in table 2-3). Degenerate primers were not found to amplify *CYC*-like genes specifically in *C. purpurea*, but did amplify a *CYC*-like gene in *Lupinus* sp. that was not amplified by LEGCYC_F1-LEGCYC_R1.

2.3.3 Sequencing of *CYC*-like genes in *Cercis griffithii*

The same two *CYC*-like genes (Cercis 1 and Cercis 2, table 2-3) were isolated in *Cercis griffithii* using two different forward primers in the TCP domain, LEGCYC_F1 and LEGCYC_F3, in combination with LEGCYC_R1. The sequence Cercis 1 showed high similarity to a sequence from another Caesalpinoid taxon *Ceratonia oreothauma* (Ceratonia 1) with 82.7 % protein similarity and requiring the insertion of six gaps of one to five amino acids, whereas Cercis 2 showed high similarity to Cadia 4 (71.7 % protein similarity, with the insertion of two gaps of two amino acids) (figure 2-5).

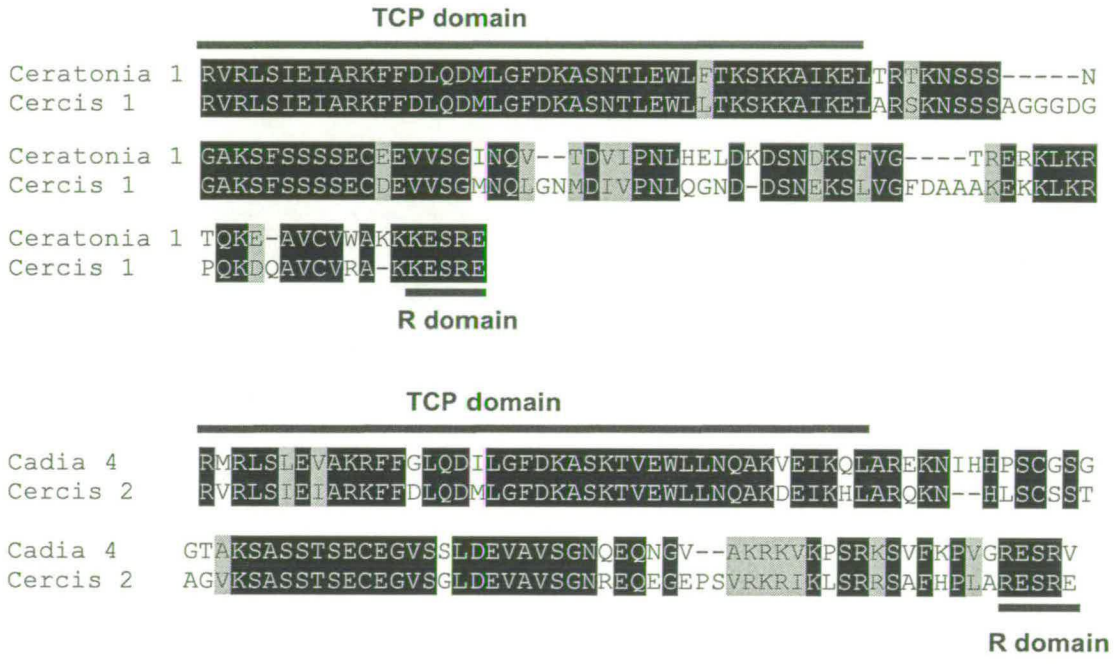


Figure 2-5. An alignment of the predicted amino acid sequence from Ceratonia 1 and Cercis 1, and Cadia 4 and Cercis 2. Identical amino acids are in black boxes, while amino acids with similar charge or hydrophobicity are in grey. The partial TCP and R domains are shown for both sequence pairs.

PART 2: LEGUME *CYC* GENES WITHIN THE TCP GENE FAMILY

2.4 MATERIALS AND METHODS

2.4.1 TCP sequence sampling

Seven TCP domain sequences from two species critical in this study for investigating the function of *CYC*-like genes, *Lotus japonicus* and *Cadia purpurea* (Lotus japonicus 1 and Lotus japonicus 2 (D. Luo, pers. comm.), and Cadia 1 – 4 described in part 1 of this chapter), were placed in the context of the TCP gene family. Comparison with the other legume *CYC*-like sequences described in section 2.3.2 suggested that these seven sequences represented the diversity of legume *CYC*-like TCP sequences.

Sampling of TCP sequences was similar to that of Cubas (2002). In this analysis, however, certain *Arabidopsis* TCP genes belonging to the PCF group (Cubas, 2002), which is not the focus of this study, were excluded (*TCP7*, *TCP8*, *TCP14*, *TCP15*, *TCP20*, *TCP21*, and *TCP22* following the nomenclature of Cubas (2002)), whereas other sequences of particular interest were added: *Gossypium hirsutum* AUXIN, *Lupinus albus* 'TCP1', *Linaria vulgaris LCYC*, and *Antirrhinum majus DICH* (Genbank accession numbers given in appendix 3).

2.4.2 Phylogenetic analyses

Phylogenetic analysis of TCP genes was carried out using an amino acid matrix of the conserved TCP domain, the only region that could be aligned unambiguously across all sequences. Manual alignment of the 58 amino acids of the TCP domain was straightforward. The matrix of 31 sequences (appendix 3) was analysed using not only protein distance methods similar to those of Cubas (2002), but also maximum parsimony, maximum likelihood (ML), and Bayesian methods, which operate directly on discrete character data rather than on a matrix of

pairwise distances. Protein distance analysis was carried out using programs from the PHYLIP package (Felsenstein, 1993). One hundred half-deletion jackknife data sets were obtained with SEQBOOT, distance matrices were calculated with PROTDIST using maximum likelihood estimates based on the PAM-Dayhoff model of amino acid substitution, neighbour-joining trees were obtained with NEIGHBOR, and a consensus tree was produced by CONSENSE. Branches with < 50% support were collapsed. The most parsimonious trees were calculated with PROTPARS (Felsenstein, 1993) with support values obtained by 100 half-deletion jackknife replicates as described above. A majority rule consensus tree was obtained with CONSENSE, collapsing branches with <50% jackknife support. Protein ML analysis was carried out using TREEPUZZLE v.5.0 (Schmidt *et al.*, 2002) with the BLOSUM 62 model of substitution (Henikoff & Henikoff, 1992), which is better for distantly related proteins, and allowing for two rates of heterogeneity (invariable sites plus gamma distributed rates) estimated from the data. As support values cannot easily be obtained for ML analyses of large datasets, these were obtained by Bayesian analysis using MrBayes v.2.01 (Huelsenbeck & Ronquist, 2001). Bayesian analysis, like ML, is based on the likelihood function and can take into account complex models of sequence evolution, but instead of searching for the optimal tree as in ML or parsimony, trees are sampled repeatedly according to their posterior probability. The consensus of the sampled trees can be considered an approximation of branch support (Huelsenbeck *et al.*, 2001). In this analysis, one million Markov Chain Monte Carlo (MCMC) generations, sampled every 100 generations, were run. The first 100,000 generations (the "burn-in", before the chain reaches its equilibrium) were discarded.

2.5 RESULTS

Analysis of the TCP domain peptide matrix using protein distance, parsimony,

maximum likelihood (ML), and Bayesian methods resulted in congruent trees with strong support values for the major groups. Figure 2-6 shows the protein ML unrooted phylogram, with support values obtained by Bayesian analysis of the data. The 50% majority rule (MR) protein distance and maximum parsimony trees (figures 2-7a and 2-7b respectively), are shown for comparison. All analyses strongly suggest that the TCP gene family can be divided into three main groups. The PCF group contains the rice *PROLIFERATING CELL FACTORS*, characterised by amino acid deletions at positions 8-10 and 13 from the start of the TCP domain protein sequence (see appendix 3). A second group contains *CYC/TB1*, and three *Arabidopsis* genes (*TCP1*, *TCP12*, *TCP18*) with an R domain. These results confirm the conclusions of Cubas (2002), but with greater sampling and more comprehensive phylogenetic analyses. A third well-supported clade in all analyses contains proteins that are related to the leaf development gene *CINCINNATA* (*CIN*) in *Antirrhinum* (sequence not included here) (Nath *et al.*, 2003). Some of the *Arabidopsis* genes in the *CIN* group (*TCP2*, *TCP3*, *TCP4*, *TCP10*, *TCP24*) are also believed to be involved in leaf morphogenesis (Palatnik *et al.*, 2003) (see figure 2-5).

All analyses suggest that the *CYC*-like sequences from *C. purpurea* and *L. japonicus*, with the exception of Cadia 4, form a strongly supported group found in 92% of Bayesian trees. This monophyletic group, LEGCYC, is sister to the *CYC-TCP1* clade in the ML, Bayesian (figure 2-6) and distance (figure 2-7a) trees. Although it is difficult to infer relationships from unrooted trees, these trees strongly suggest that the LEGCYC genes are putative orthologues of *CYC* and *TCP1*. Cadia 4 is recovered in ML (figure 2-6) and distance (figure 2-7a) analyses in the clade containing *TB1*, *TCP12* and *TCP18*. The parsimony analysis is not informative because the relationship between the LEGCYC clade, Cadia 4, the *CYC/LCYC/DICH* clade, *TCP1*, *TCP12*, *TCP18*, and *TB1* collapses in a 50% MR consensus tree.

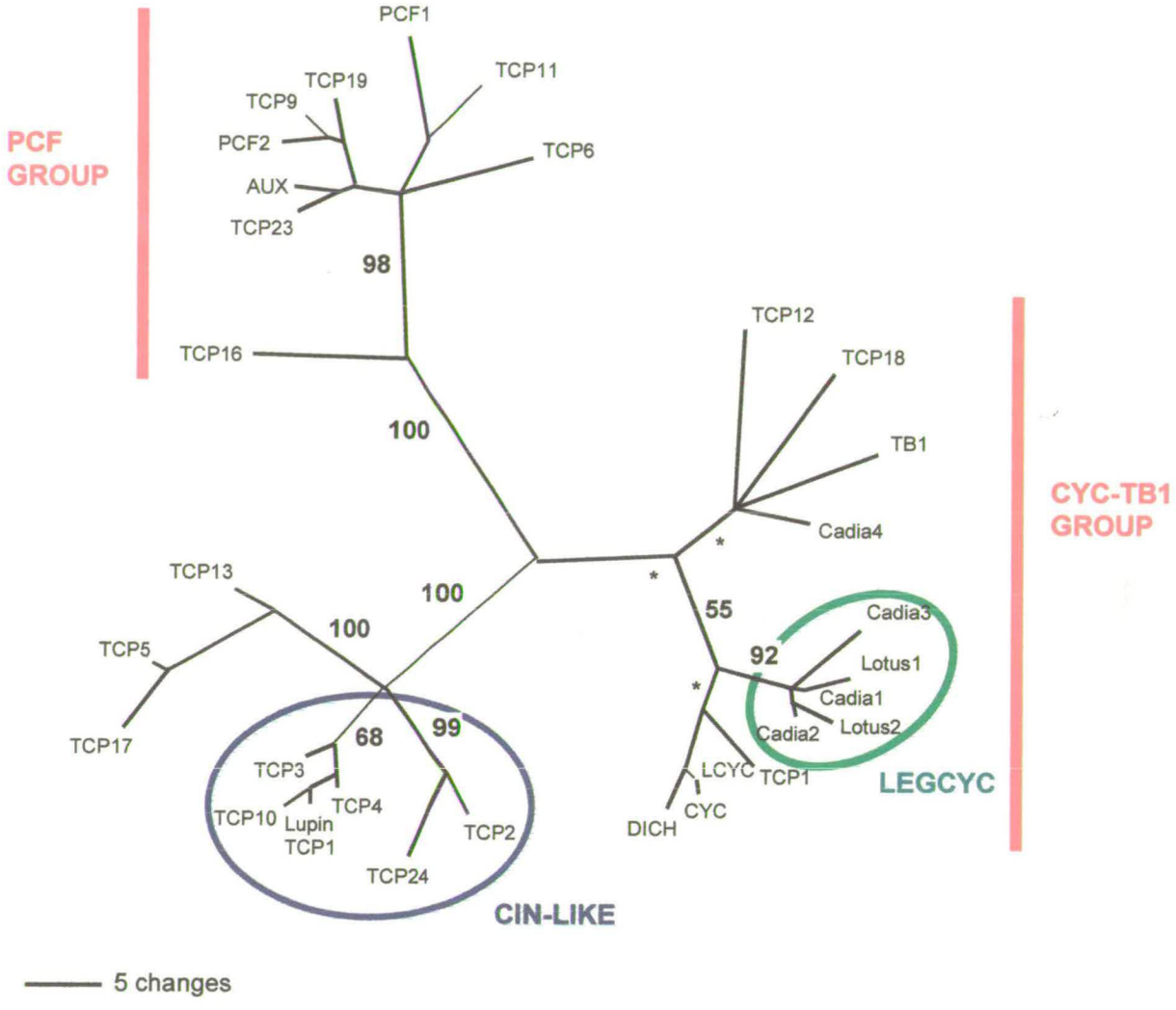


Figure 2-6. Unrooted phylogram of protein ML analysis using TREEPUZZLE v5.0 (Schmidt *et al.*, 2000) of the TCP domain data set including representative legume sequences. The *CYC-TB1* and PCF groups described in Cubas (2002) are recovered here, as well as a group containing *CIN*-like genes (Palatnik *et al.*, 2003). Support values were obtained using MrBayes (Hulsenbeck and Ronquist, 2001); asterisks * indicate that a clade was recovered in < 50% of Bayesian trees. Results support a LEGCYC clade (highlighted in green, excluding Cadia 4) as sister to the *CYC/TCP1* clade. All TCP genes, unless otherwise indicated, are from *Arabidopsis*; *PCF* from rice; *TB1* from maize; *LCYC* from *Linaria vulgaris*, *CYC* and *DICH* from *Antirrhinum*; *AUX* from cotton (accession numbers in appendix 3).

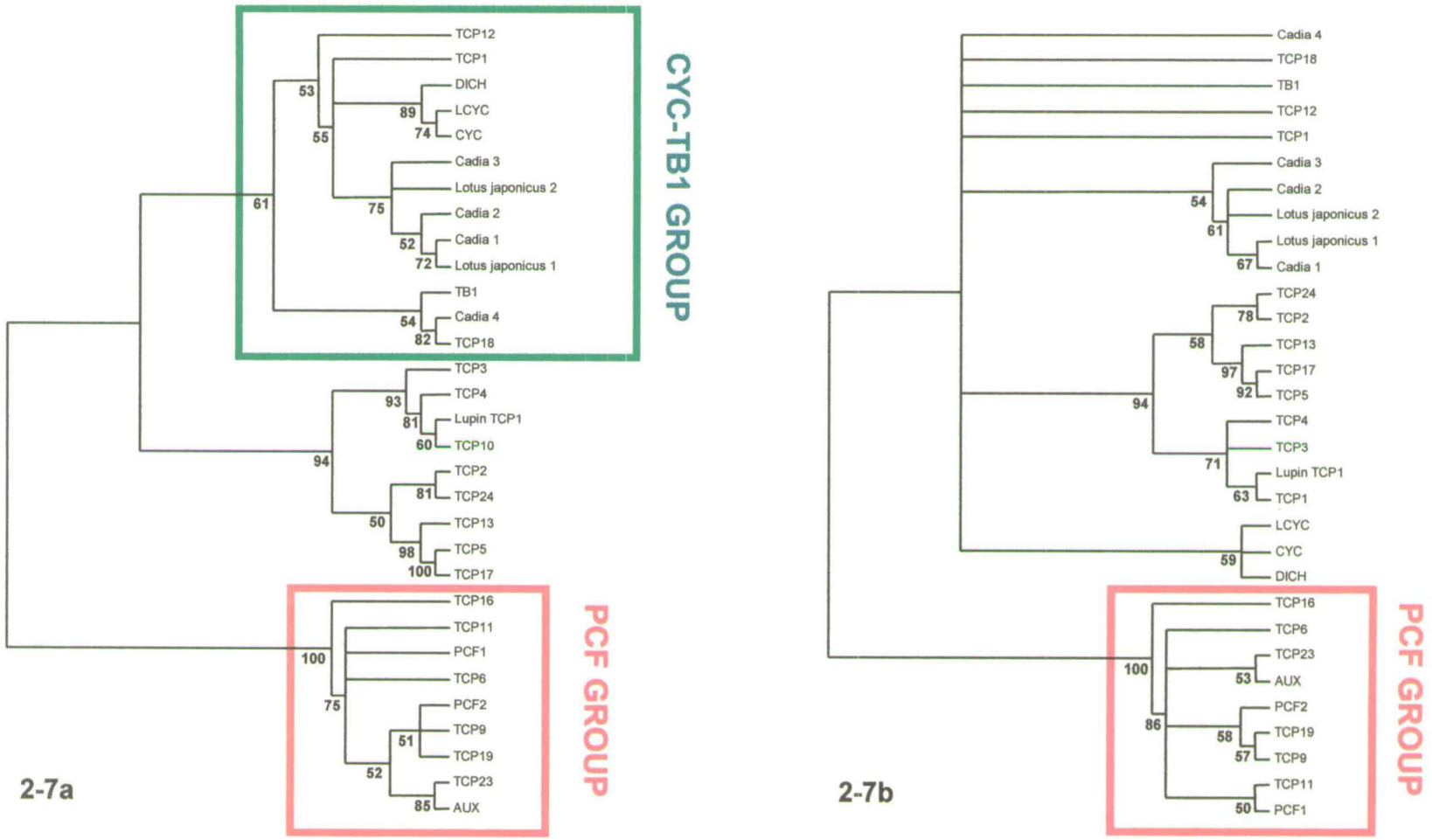


Figure 2-7. 2-7a. Fifty percent Majority Rule (MR) consensus tree of the protein distance analysis using the PAM-Dayhoff model of protein substitution (PROTDIST; Felsenstein, 1993) of the TCP domain. Values > 50% of the 100 jackknife replicates are given at branch nodes. Taxa as in legend to figure 2-6. **2-7b.** Fifty percent MR consensus tree of protein maximum parsimony analysis (PROTPARS; Felsenstein, 1993) of the TCP domain. Support values above 50% from the 100 jackknife replicates are shown. Maximum parsimony fails to resolve groups recovered in protein ML, Bayesian and distance analyses. Although it does not contradict any of the results from other methods, it offers no support for a *CYC-TB1* clade, and only weak support (54%) for a LEGCYC clade.

PART 3: LEGUME *CYC* GENE PHYLOGENY

2.6 MATERIALS AND METHODS

2.6.1 Sequence sampling and alignment

After identification of the putative orthologues in legumes of *Antirrhinum CYC* in the context of the TCP gene family (this chapter, part 2), phylogenetic analysis of *CYC*-like sequences within the Leguminosae was carried out. Based on primary homology assessment, all sequences listed in table 2-3 were included, with the exception of Cadia 4 and Cercis 2 which were not found to belong to the LEGCYC clade (see results section 2.3.3 and 2.5). Legume *CYC* sequences from separate studies on model legumes were included in these analyses: *Lotus japonicus* (Lotus japonicus 1, Lotus japonicus 2), *Glycine max* (Soya 1), *Pisum sativum* (Pisum CYC1, Pisum CYC2) (D. Luo, pers. comm.), and *Medicago truncatula* (Medicago 1, BG455508). *CYC*-like sequences obtained during the course of this project with primers other than the ones described in this chapter were also included: *Lupinus angustifolius* cv Merrit (Lupinus angustifolius 1, AY225839; Lupinus angustifolius 2, AY225840; described in chapter 5), and *Lupinus nanus* (Lupinus nanus 1, AY225836; described in chapter 3). Results from the TCP gene family analyses (section 2.5) suggested that *Antirrhinum CYC*, *DICH* and *Arabidopsis TCP1* be used as outgroups for the legume *CYC* gene phylogeny.

Unambiguous alignment of all LEGCYC sequences from the 25 taxa was only possible in the TCP and R domains and reduced the matrix to 145 nucleotide characters. Although the region between the TCP and R domains could not be aligned between all legume sequences, it was believed to contain characters that may be phylogenetically informative. It was possible to align certain parts of the variable region for a subset of legume sequences, excluding a total of 300 ambiguous characters. Protein sequences were aligned using Clustalx (Thompson *et al.*, 1997), followed by manual adjustments taking both amino acids and nucleotides into

consideration. Analyses of the variable region were unrooted as outgroup sequences from *Antirrhinum* or *Arabidopsis* were not alignable with legume sequences.

2.6.2 Legume *CYC* phylogenetic analyses

Maximum parsimony and model-based methods of phylogeny reconstruction were used for analysing partial LEGCYC nucleotide sequences.

Maximum parsimony analysis was carried out using PAUP* 4.0b10 (Phylogenetic Analysis Using Parsimony and other methods, version 4.0b10, Sinauer Associates, Sunderland, MA; Swofford, 2001). Heuristic searches with 1,000 random addition replicates, to avoid local optima of globally suboptimal trees, and tree bisection reconnection (TBR) branch swapping were conducted with steepest descent and multrees options selected. A maximum of 10 minimal trees were retained per replicate, and a further heuristic search by TBR was carried out on the shortest trees. Branch support values were calculated by 1,000 boostrap replicates with simple sequence addition and a maximum of 10 minimal trees retained per replicate. This search method was carried out both for the TCP and R nucleotide matrices, as well as the matrix incorporating certain variable regions. As the parsimony analysis of the TCP and R region provided no resolution within the LEGCYC clade, certain sequences identified using RadCon v 1.1.5 (Thorley and Page, 2000) with a low "leaf stability" value (a measure of the certainty of the position of a sequence, or "leaf", in a set of bootstrap trees) were then removed from the matrix. The reduced dataset was analysed as above.

ML analyses were carried out for the reduced TCP plus R dataset and the matrix incorporating the more variable regions. The best-fit model of nucleotide substitution was selected for each data set by the Akaike Information Criterion, which imposes a penalty for unnecessary parameters, using Modeltest v3.06 (Posada and Crandall, 1998). For the reduced TCP plus R dataset, the TIM + I + G model was selected. This is a transitional model (TIM)

where a proportion of sites can be invariable (I) and among-site variation of substitution rate follows a gamma distribution (G). This parameter-rich model estimates empirical substitution rates for transitions while equal rates are assumed for transversions (Rmat = A-C:1.0000 A-G:2.2829 A-T:0.4622 C-G:0.4622 C-T:3.5964). Base frequencies were estimated empirically (Lset Base = A:0.3558 C:0.2362 G:0.2106), as were the proportion of invariable sites (Pinvar = 0.4259). The shape of the gamma distribution was $\alpha = 1.0094$, where $1/\alpha$ describes the variance in substitution rate. The GTR + I + G model was selected for the matrix incorporating more variable regions. This is a general time reversible model where nucleotide frequencies can be unequal and the six possible transitions between nucleotide states can occur at different rates (Rmat = 1.9079 2.8427 0.9545 1.2000 4.1774), with estimated base frequencies (Lset Base = 0.3348 0.1814 0.2567), among-site rate variation distributed according to a gamma-distribution ($\alpha = 1.1731$) and proportion of invariable sites (Pinvar = 0.175). A heuristic ML analysis with TBR branch swapping was carried out using PAUP* v4.0b10 with the parameters defined above.

Bayesian phylogenetic analyses of the reduced TCP plus R dataset and the matrix incorporating the more variable regions were carried out using MrBayes v2.01 (Huelsenbeck and Ronquist, 2001) using a general time reversible (GTR) model and site-specific rates partitioned by codon. Chains were run for 600,000 and 1,000,000 generations (burn-in of 100,000 generations) for each data set respectively, sampled every 100 generations. Resultant trees were used to generate a 50% majority rule consensus tree in PAUP* v4.0b10.

2.7 RESULTS

2.7.1 Evolution of LEGCYC genes: partial TCP and R nucleotide analyses

Parsimony analysis of all LEGCYC partial TCP and R nucleotide sequences resulted in 194 most parsimonious trees of 486 steps, with a low consistency index (CI) of 0.321, and a low

retention index (RI) of 0.567, indicating high homoplasy (parallel evolution) in the data. The strict consensus of the most parsimonious trees, rooted on *Antirrhinum CYC* and *DICH*, resolved very few relationships with little bootstrap support within the LEGCYC clade (figure 2-8). A summary of descriptive values of this data matrix and parsimony analysis, as well as the other two nucleotide parsimony analyses (see below), are given in table 2-4.

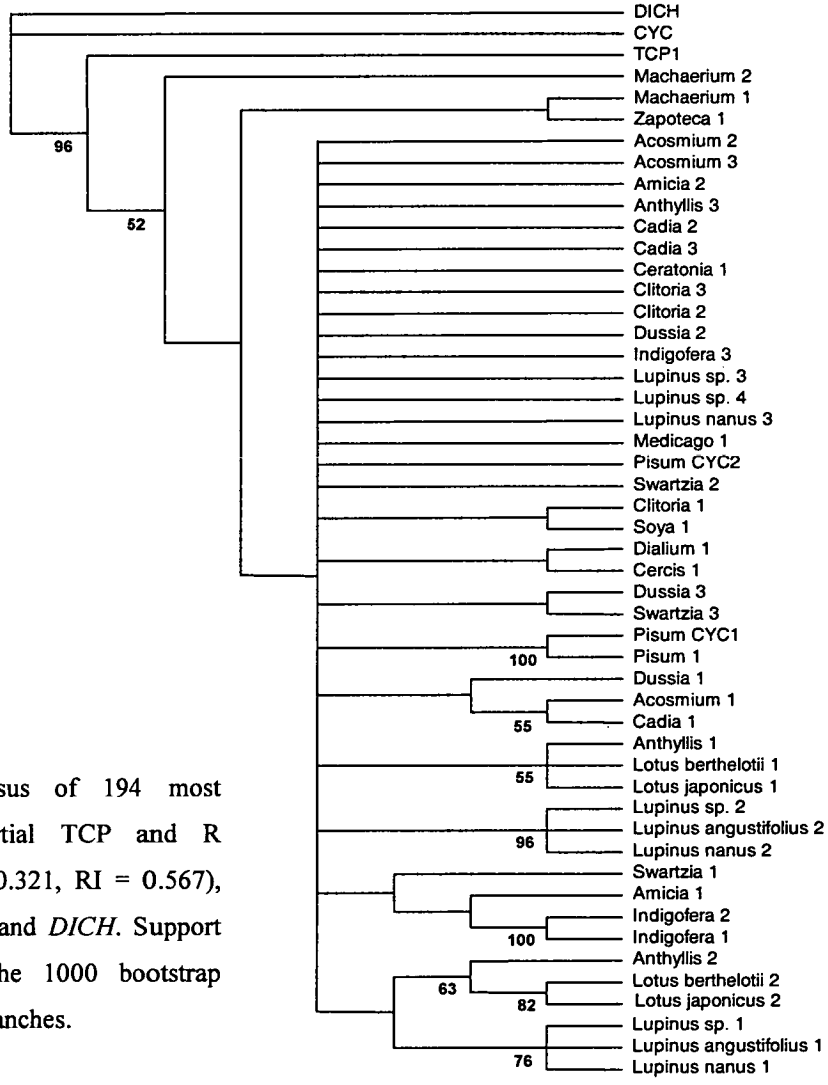


Figure 2-8. Strict consensus of 194 most parsimonious trees of partial TCP and R nucleotide sequences (CI = 0.321, RI = 0.567), rooted on *Antirrhinum CYC* and *DICH*. Support values above 50% from the 1000 bootstrap replicates are shown below branches.

To attempt to recover major groups within the LEGCYC genes, a reduced matrix of 29 legume partial TCP and R sequences was analysed, excluding caesalpinoid and mimosoid sequences and certain papilionoid sequences with a low leaf-stability index (Thorley and Page, 2000). Trees were rooted on *Antirrhinum CYC* and *DICH*. Parsimony analysis of the 67 parsimony informative sites out of 145 characters, produced 168 trees with a minimal length of 278 steps, with CI = 0.424 and RI = 0.636 (see table 2-4). Despite the high level of homoplasy, the strict consensus tree of the most parsimonious trees resolved one large clade within the ingroup corresponding to group II (defined below) (figure 2-9a). Bootstrap support for this clade was 67%. Within this clade, only the relationship between sequences from different species of the same genus (*e.g. Lupinus* spp.) or related genera (*e.g. Anthyllis hermanniae* and *Lotus* spp.) were supported in this analysis.

Model-based methods, such as maximum likelihood and Bayesian inference, are explicitly designed to deal with superimposed substitutions and may therefore be better for analysing homoplastic data (Lewis, 2001; Holder & Lewis, 2003). Bayesian analysis of the reduced TCP and R dataset recovered two groups of legume sequences referred to as group I and group II (figure 2-9b). Support values are defined here as the percentage of trees among those sampled by Bayesian analysis recovering a particular group. Group II has a very high Bayesian support (97%), whereas group I has weak support (52%). Both groups include species from basal as well as more derived papilionoids and would appear to represent an early duplication event. However, relationships between sequences other than from closely related species or genera were difficult to interpret. For comparison, one of three ML trees, which have identical topology but differing branch lengths, is shown (figure 2-10). Although group II is nested within a grade of LEGCYC sequences, the short branch lengths, representing the amount of change over time, within the LEGCYC clade further illustrate why analysing of TCP and R domain nucleotide sequences from legumes is so problematic.

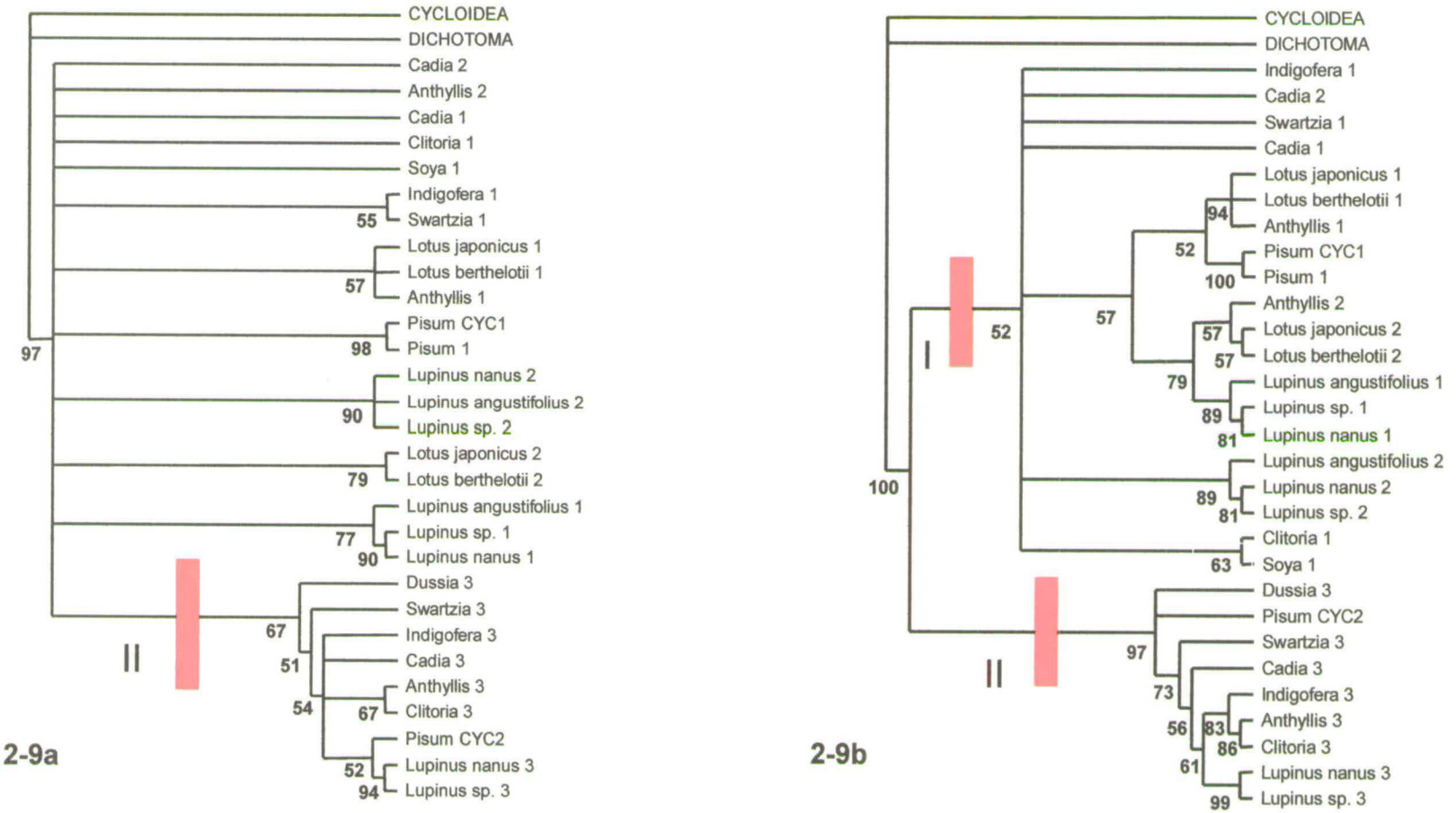


Figure 2-9. Analyses of 29 partial legume TCP and R domain nucleotide sequences. **2-9a**. Strict consensus of 168 most parsimonious trees (CI = 0.424, RI = 0.636), with bootstrap values shown below branches. **2-9b**. Bayesian analysis 50% MR tree of the legume TCP and R nucleotide sequences allowing for codon specific nucleotide substitution. Major clades I and II within LEGCYC are indicated with high Bayesian support. Both consensus trees are rooted on *Antirrhinum CYC* and *DICH*.

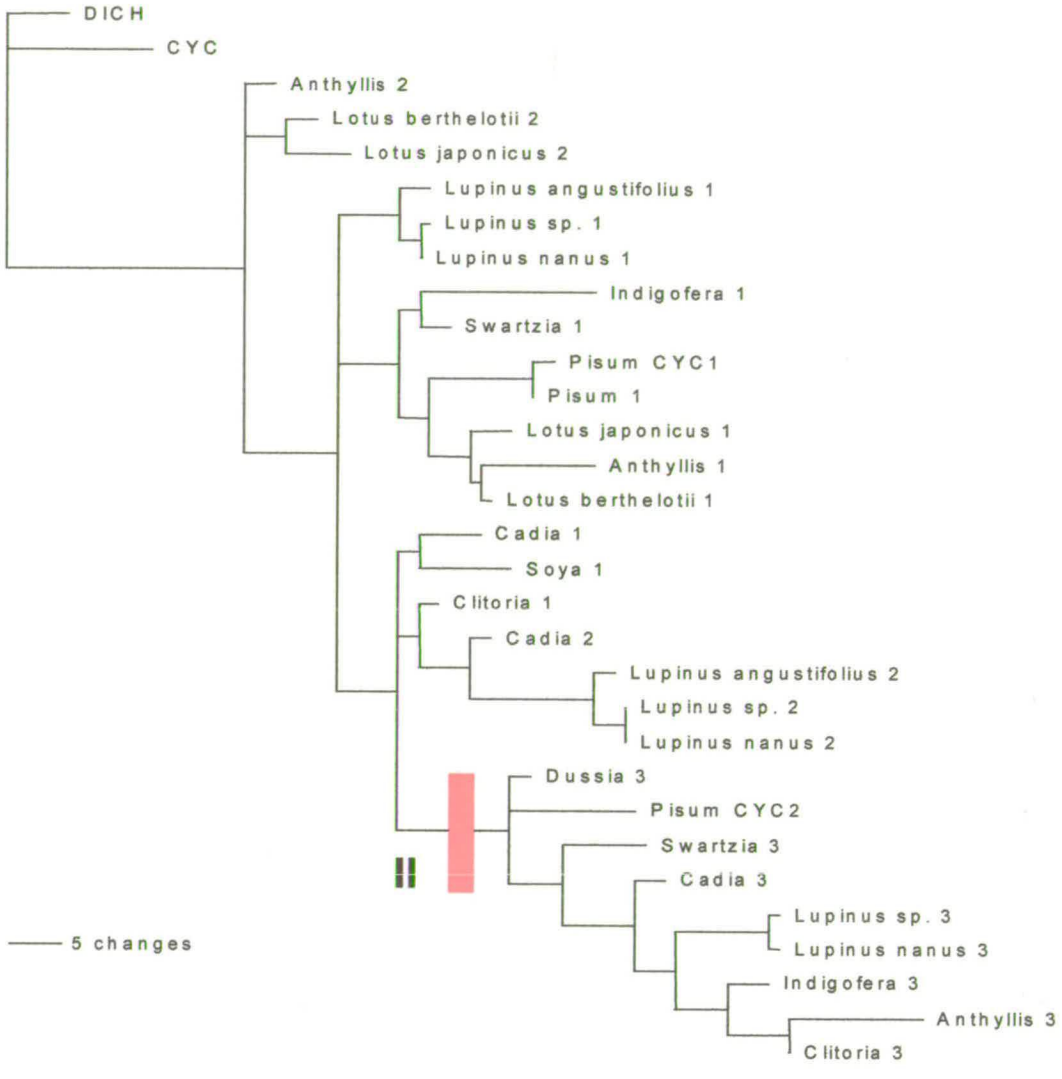


Figure 2-10. One of three most likely trees of the TCP plus R data set, analysed with the parameters of the best-fit model TIM + I + G selected by the Akaike Information Criterion. All trees have an identical topology, but differ in branch lengths. Group II (marked by the red bar), also recovered by maximum parsimony and Bayesian analysis of the same data, is nested here within a grade of LEGCYC sequences.

In conclusion, although parsimony analysis of the reduced data set did not resolve relationships well between LEGCYC genes, Bayesian analysis gave a more fully resolved tree. The poor performance of parsimony analysis was probably due to high homoplasy in the data

coupled with a low number of informative characters (also highlighted in the ML tree) with consequent low phylogenetic signal.

2.7.2 Evolution of LEGCYC genes: inclusion of sequence data between the TCP and R domains

The data set from 38 LEGCYC sequences incorporating nucleotides between the TCP and R domains consisted of 292 aligned characters, requiring the insertion of 29 gaps of one to 18 base pair triplets (see appendix 4 for alignment).

Parsimony analysis of the 153 parsimony informative characters resulted in a single most parsimonious tree of 748 steps, with CI = 0.452 and RI = 0.601 (see table 2-4). The tree suggested two clades corresponding to groups I and II from the previous analyses with a bootstrap value of 65% (figure 2-11a). Sequence relationship within these groups had little bootstrap support with the exception of sequences from closely related taxa. The topology of the ML tree and the 50% MR consensus tree from the Bayesian analysis was identical with only three nodes collapsing in the Bayesian consensus tree (figure 2-11b). The topology of those trees was also similar to the tree from the parsimony analysis, but the level of support for the nodes was much higher in the model-based analysis (estimated by Bayesian inference). For instance, groups I and II were recovered in 100% of trees sampled in the Bayesian analysis.

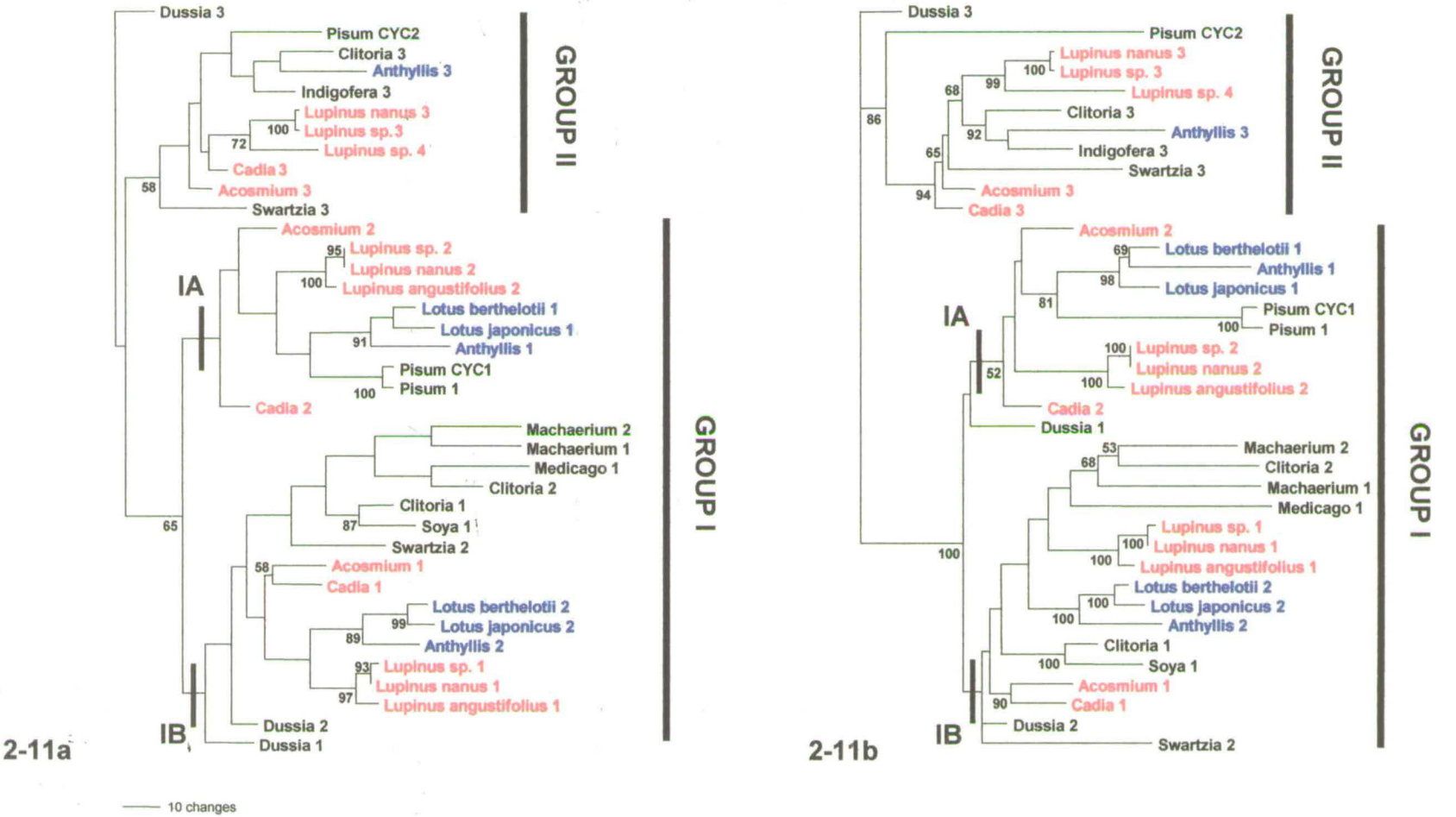


Figure 2-11. Maximum parsimony and ML analyses of 38 partial legume *CYC*-like sequences including some sequence data from the hypervariable region. Major groups recovered from the previous analyses (group I and group II) are shown, as well as one putative duplication event in group I is marked by IA and IB. Clades containing genistoid (in red) and robinioid (in blue) sequences are highlighted suggesting these putative duplication events. **2-11a**. Unrooted phylogram of the single most parsimonious tree (748 steps, CI = 0.452, RI = 0.601). Bootstrap values are given for branches with > 50% support. **2-11b**. Unrooted phylogram of the ML analysis using the GTR + I + G model of nucleotide substitution. Support values at each node were obtained by Bayesian analysis of the data set and represent the frequency of each node in the MR consensus tree.

Comparison of the partial TCP domain amino acid sequences from group I and II showed that they could be distinguished by five synapomorphies, suggesting these clades are genuine (figure 2-12). These groupings were also supported by considerable differences in the variable region between the TCP and R domain, for instance in the presence or absence of motifs such as the EVV amino acid motif characteristic of group I sequences (see chapter 3, figure 3-2;), which could not be included in the analysis.

```
                  A                       C           DE
                W S  D       N          R   D   T  NR MED
GROUP I     RVRLSIEIARKFFDLQDMLGFDKASNTLEWLFNKSKKAIKEL
                 *                 *   *           **
GROUP II    RVRLSSEIARKFFDLQDMLEFDKPSNTLEWLFTKSENAIKEL
                  NDV         E  DV           LA  DT
                   Q              QY           N   S
```

Figure 2-12. Comparison of the partial TCP domain amino acid sequence from group I and II *CYC*-like sequences in legumes. Asterisk highlights group-specific changes; above and below bold sequences are amino acid differences found less frequently in these groups.

Within group I, two sequences from most taxa were found. These segregated into two putative clades referred here as 1A and 1B (see figure 2-11), which for the most part contained one sequence per taxon, with a few exceptions, for example Machaerium 1 and 2, and Clitoria 1 and 2. Clade 1A contained one LEGCYC sequence from representatives from both genistoid (*Lupinus* spp., *C. purpurea*, *Acosmium subelegans*) and robinioid (*Lotus* spp., *Anthyllis hermanniae*) clades. Clade 1B contained another LEGCYC sequence from these taxa. Although these clades have no bootstrap support in the parsimony analysis, they were found in the ML tree and in most Bayesian trees. This suggests a putative orthology relationship between sequences within these clades, and a further conserved duplication of LEGCYC sequences of possible functional significance.

Descriptive statistics	Total TCP + R[1]	Reduced TCP + R[2]	TCP + R + variable[3]
total no. of sequences	51	31	38
aligned sequence length	145	145	292
no. of excluded sites	-	-	300
no. of indels	-	-	29
size of indels (bp)	-	-	3-54
proportion of variable sites	0.593	0.490	0.692
proportion of uninformative sites	0.131	0.069	0.168
proportion of parsimony informative sites	0.462	0.351	0.524
transition/transversion ratio	1.386	1.436	1.285
% steps at 1st codon position	15.3	14.0	20.1
% steps at 2nd codon position	8.6	8.5	17.8
% steps at 3rd codon position	76.1	77.5	62.1
average no. steps per character	2.476	1.628	2.562
number of MP trees	194	67	1
length of MP trees	486	278	748
CI	0.321	0.424	0.452
RI	0.567	0.636	0.601

Table 2-4. Descriptive values of the maximum parsimony analyses carried out with different nucleotide data sets: 1: all LEGCYC, *Antirrhinum CYC*, *DICH*, and *Arabidopsis TCP1* partial TCP and R nucleotide data (strict consensus tree: figure 2-8); 2: partial TCP and R nucleotide data of a subset of LEGCYC sequences (strict consensus tree: figure 2-9a); 3: inclusion of the hypervariable region between the TCP and R domain, aligned against a subset of LEGCYC sequences (single most parsimonious tree: figure 2-11a). MP trees: most parsimonious trees, CI: consistency index, RI: retention index.

2.8 DISCUSSION

2.8.1 Presence of *CYC/TCP1* orthologues in the Leguminosae

In the TCP gene family analyses, evidence from sequence similarity (PROTDIST) and evolution (ML and Bayesian analyses) strongly suggests that the legume *CYC*-like sequences examined here are homologous to the floral symmetry genes *CYC* and *DICH* in *Antirrhinum*, and to the adaxially expressed floral gene *TCP1* in *Arabidopsis*. Within this legume clade, a lower estimate of three *CYC*-like copies were found within the Papilionoideae, in species ranging from the basal-most clade (*e.g. Swartzia jorori*) to higher papilionoids (*e.g.* the robinioid *Anthyllis hermanniae*). In the basal caesalpinioid legume *Cercis griffithii*, only one *CYC* orthologue was found along with a putative *TB1* orthologue. This suggests that duplication of LEGCYC genes occurred during the evolution of the Leguminosae, possibly at the onset of papilionoid evolution. A more detailed examination of *CYC*-like genes in the Caesalpinioideae and Mimosoideae, as well as in the putative sister groups of the Leguminosae, some of which like Polygalaceae have flowers superficially like papilionoid legumes (Doyle & Luckow, 2003), are required to confirm this finding.

These results are in agreement with parallel studies of *CYC*-like genes in legumes. Three *CYC*-like genes were isolated from a *Lotus japonicus* floral cDNA library (D. Luo, pers. comm.), and these are similar to the three genes found here in *Anthyllis hermanniae*, a member of the sister genus to *Lotus*. Fukuda, Yokoyama and Maki (2003) have also isolated multiple copies of genes with a TCP and R domain in four papilionoid species. The three *CYC*-like genes they have isolated in *Cytisus racemosus* (AB076986, AB076987, AB076988) are orthologous to the *Lupinus nanus* sequences 1-3, whereas other sequences (*Sophora flavescens* SfCYC2 AB076994, *Wisteria floribunda* WfCYC3 AB076997, *Pueraria montana* var. *lobata* PmCYC3 AB076991) are putative orthologues of Cadia 4 (analyses not shown).

Because of their apparent orthology with *Antirrhinum CYC*, these LEGCYC copies are candidate floral developmental genes in the Leguminosae. The expression pattern of these was investigated and contrasted in closely related species with different floral symmetry, *Lupinus nanus* and *Cadia purpurea* (see chapter 4). However, these phylogenetic analyses, many of which lead to poorly resolved trees, highlight some of the difficulties in making detailed orthology statements within gene families and the rapidly evolving *CYC*-like genes in particular.

2.8.2 Problematic reconstruction of legume *CYC*-like gene evolution

No simple pattern of gene evolution tracking organismal phylogeny within the legume *CYC* family was recovered in the phylogenetic analyses. Confounding factors such as intermediate levels of concerted evolution, variation in the rate of sequence evolution, and independent gene loss and duplication events which render the interpretation of gene trees difficult (Doyle, 1994) cannot be ruled out here. Because the analysis also includes clades that may be functionally differentiated, particular amino acid positions may be subject to different selection pressure in different parts of the tree. This within-site rate variation, or heterotachy (Lopez *et al.*, 2002), is also likely to make phylogenetic reconstruction more difficult.

Different levels of variation in different parts of these *CYC*-like genes also made analysis difficult. The highly conserved TCP and R domains were alignable, making character definition simple, but contained few phylogenetically informative characters. By contrast, the region between the two domains was variable but difficult to align, making character definition ambiguous. Futhermore, the variation in the TCP and R domains was mainly at the synonymous third codon position and showed a high degree of homoplasy (accounting for two-thirds of the steps required in the parsimony analyses). High levels of homoplasy, resulting in artificial groupings, is also suggested by the low support values for the most parsimonious trees of the TCP and R legume sequences and the collapse of many nodes in the strict consensus trees.

For this type of problematic data, theoretical considerations regarding how primary homologies are treated and simulation studies suggest that model-based approaches such as maximum likelihood and Bayesian inference perform better for phylogeny reconstruction than parsimony (*e.g.* Hillis, 1996; Alfaro *et al.*, 2003). In this study, Bayesian inference provided better resolution and support for putative major LEGCYC clades (groups I-A, I-B and II) than parsimony. However, branch support values obtained by posterior probability from Bayesian inference are thought to be an over-estimation (Suzuki *et al.*, 2002; Erixon *et al.*, 2003). Despite limitations associated with various methods of phylogeny reconstruction, and the problematic nature of the data, certain patterns did emerge from the analyses.

2.8.3 Evidence for multiple duplication events within the Papilionoideae

Results of the rooted Bayesian analysis suggests that LEGCYC genes can be divided into two main groups (I = LEGCYC1, and II = LEGCYC2), which are characterised by different amino acid signatures in the TCP domain. The results of the analyses of the extended data set are also consistent with the two-group hypothesis; these groups, although only moderately supported by the maximum parsimony, are strongly supported by Bayesian inference. Taxa, ranging from the basal-most papilionoids to highly derived species from the "inverse repeat loss clade" such as *Pisum*, have both groups of genes suggesting that these genes probably diverged after a duplication event that occurred before the evolution of the Papilionoideae. In addition to the putative amino acid synapomorphies in the TCP domain, these groups are also distinguished by specific motifs in the otherwise variable region between the TCP and R domains.

Within LEGCYC1, one other major duplication event appears to have occurred, giving rise to two subgroups LEGCYC1A and LEGCYC1B. Genes belonging to both clades were recovered in a wide range of species sampled, implying that this duplication also occurred prior to the diversification of the papilionoids.

However, the relationships between sequences within these groups appear complex and require further investigation. Even though the sampling here is extensive compared to many studies of developmental gene phylogeny, increasing it may help resolve relationships within and between gene copies. Nevertheless, these results are in agreement with a trend of independent duplications, and possible losses, with rapid gene evolution outside of the conserved TCP and R domains, previously documented in *CYC*-like genes families from other plant groups (*e.g.* Antirrhineae: Hileman & Baum, 2003, Gübitz *et al.*, 2003; Gesneriaceae: Citerne *et al*, 2000; Solanaceae: K. Coenen, unpublished).

2.8.4 The limitations and potential of *CYC*-like gene phylogenetics

The rapid rate of evolution of *CYC*-like genes, outside of the conserved TCP and R domains, do not make them suitable for phylogenetic analysis across the legume subfamilies. Reconstructing the history of LEGCYC evolution may nevertheless be improved by increasing taxon sampling. However, between closely related taxa, these LEGCYC genes are a potential source of phylogenetic information (further discussed in chapters 3 and 5). In New World *Lupinus* species, both LEGCYC1A and LEGCYC1B phylogenies improved relationship estimates from those obtained from sequences of the ribosomal internal transcribed spacers (ITS) within this recently diverged clade (Ree *et al.*, 2004).

The recognition of a major legume *CYC*-like group (LEGCYC) in this study does suggest likely candidate genes for functional equivalents of *Antirrhinum CYC* and *Arabidopsis TCP1*. Furthermore, within this group of legume *CYC*-like genes, further subgroups are recognised (LEGCYC1A, LEGCYC1B, LEGCYC2), inviting investigation of possible functional differences between these. Thus, even where phylogenetic analyses are difficult, partial resolution may still enable hypotheses based on sequence homology to be generated.

CHAPTER 3: CHARACTERISATION OF *CYC*-LIKE GENE SEQUENCES IN *CADIA PURPUREA* AND *LUPINUS NANUS*

3.1 INTRODUCTION

Prior to studying the expression pattern of a gene of interest, it is valuable to characterise its full-length open reading frame (ORF). In particular, knowledge of the 5'-end sequence of a gene is desirable for RNA *in situ* hybridisation, as probes from regions around the start of the ORF have been found to produce better hybridisation signals (E. Coen, pers. comm.). Different PCR-based approaches can be used to isolate upstream and downstream regions of a known fragment. For instance, inverse PCR works by amplifying circularised fragments of digested genomic DNA using primers which face outward of the known sequence (Ochman *et al.*, 1988; Triglia *et al.*, 1988). Another genome walking method requires digested DNA fragments that are not circularised but ligated to double-stranded adaptors. These adaptors have a blunt-ended strand to which the adaptor-specific primer binds and a complementary strand with a recessed 3' terminus blocked by an amine group to prevent adaptor primer extension in the same direction as the gene specific primer (Siebert *et al.*, 1995). These approaches have been used to sequence the entire ORF of the two orthologues of the putative floral symmetry genes in *Lotus japonicus* LEGCYC1A and LEGCYC1B (Lotus japonicus 1, Lotus japonicus 2), as determined by phylogenetic analysis (see chapter 2), in two closely related genistoid species *Cadia purpurea* and *Lupinus nanus* that differ in their floral symmetry.

3.2 MATERIALS AND METHODS

3.2.1 Specific amplification of *CYC*-like loci in *Cadia purpurea* and *Lupinus nanus*

Locus specific primers were designed for the two orthologous gene pairs in *Cadia purpurea* and *Lupinus nanus*, which are primary candidates for the expression study: a forward primer located in the TCP domain binding to both loci (LEGCYC_F3: 5'- CAA GAC ATG YTA GGG TTT GAC -3') and two locus specific reverse primers situated before the start of the R domain. The latter were LEGCYC_R4 (5'- CTA CYA CTA CCC CTT CTG G -3') amplifying Cadia 2/ Lupinus nanus 2 (LEGCYC1A) and LEGCYC_R3 (5'- CAA GCS GGT TCC TTY TGT T -3') amplifying Cadia 1/ Lupinus nanus 1 (LEGCYC1B) (see appendix 2 for primer location). PCR mix and cycling conditions were as described in chapter 2, section 2.2.3. The annealing temperature of the PCR cycle was optimised to yield a single product for each locus and taxon. Products were purified with Qiaquick PCR purification kit (Qiagen Ltd, Dorking, Surrey, UK) and sequenced directly.

3.2.2 Isolation of regions upstream and downstream of the initial LEGCYC1A and LEGCYC1B fragments in *C. purpurea* and *L. nanus* using different PCR based approaches

3.2.2a Inverse PCR

Approximately 200ng of genomic DNA were digested for 3½ hours in a 25µl reaction with 1 unit of the restriction endonuclease *RsaI*, which leaves a 4 bp overhang and does not cut the known fragment (New England Biolabs, Herts, UK)). To make sure the DNA was fully digested, fragments (10µl aliquot) were visualised by electrophoresis on a 1% agarose gel run for 1 hour at 80V. Fragments were then self-ligated overnight at 16°C in a 50µl reaction comprising 15µl digested genomic DNA, 1 unit of T4 DNA Ligase (Bioline, London NW2, UK), ligase buffer, and sterile distilled water. The dilution of digested fragments in this reaction

ensured that intra-molecular ligation was favoured over ligation to other fragments in the pool. The reaction was terminated by heating at 70°C for 5 minutes. Ligated fragments were then purified with Qiagen mini-elute purification kit (Qiagen Ltd, Dorking, Surrey, UK).

Two sets of primers facing outwards from the known sequence were designed to amplify both loci specifically. These were the inverse of the locus specific primers LEGCYC_R3 (LEGCYC_iR3: 5'- CAC ARA AGG AAC CWG CTT G -3') and LEGCYC_R4 (LEGCYC_iR4: 5'- CCA GAA GGG GTA GTR GTA G -3'), and the inverse of general primer in the TCP domain LEGCYC_F3 (LEGCYC_iF3: 5'- GTC AAA CCC TAR CAT GTC TTG -3') (see appendix 2). Internal primers for nested PCR were modified from the general primers LEGCYC_F1 and LEGCYC_R1 described in chapter 2, section 2.2.2: LEGCYC_iF1: 5'- TCA CCC TSC GGT CCC TCA -3' and LEGCYC_iR1: 5'- AAA GCA AGA GCA AGA GCA AGG -3' (see appendix 2). A summary of PCR conditions is given in table 3-1. Products were purified using Qiagen PCR purification kit (Qiagen Ltd, Dorking, Surrey, UK) and sequenced directly.

3.2.2b Standard PCR

To confirm results of inverse PCR in the case of *C. purpurea*, and to amplify most of the ORF in *L. nanus*, primers were designed near the start (LEGCYC_F5: 5'- CTT TCY TTA ACC CTG AAA ATG CTT C -3') and end (LEGCYC_R5: 5'- YAT TSG CAT CCC AAT TTG GAG -3'; LEGCYC_R8: 5'- CAC TCY TCC CAR GAY TTT CC -3') of the ORF (see appendix 2). These were used in combination of with locus specific primers LEGCYC_R3/R4 and LEGCYC_iR3/iR4 respectively. PCR conditions are summarised in table 3-1.

3.2.2c Genome walking

A genome walking protocol modified from Siebert *et al.* (1995) (G. Ingram, University of Edinburgh, pers. comm.) was followed to further sequence the flanking regions of the

LEGCYC1A and LEGCYC1B fragments in *C. purpurea* and *L. nanus*. 2.5μg of genomic DNA were digested overnight with 5 units of a 6 bp blunt-end cutter (*EcoRV*, *HpaI*, *SmaI*, *ScaI*) in a 100μl reaction. The digest was purified using phenol-chloroform and eluted in the final step in 20μl distilled water. These fragments were then ligated to 2.4μl adaptor solution (25μM) (G. Ingram, pers. comm.) using T4 DNA ligase (New England Biolabs, Herts, UK) in a 10μl reaction overnight at 16°C. The reaction was terminated at 70°C for 5 minutes, then made up with distilled water to 100μl final reaction volume.

To amplify specific DNA fragments, a nested hot-start PCR protocol, with "step-down" conditions similar to that of Zhang and Gurr (2000), was followed (see table 3-1 for PCR conditions). 'Hot start', *i.e.* heating the reaction mix at 94°C for 2 min prior to the addition of 1 unit of *Taq* polymerase to minimise non-specific priming, was carried out for both the first and nested PCRs. Gene-specific primers designed to amplify upstream (LEGCYCI-GW1: 5'- AAC CCT ARC ATG TGT TGW AGA TCR AAG AAC -3', LEGCYCIA-GW2: 5'- CMG GTT TGT TWG YAA GAA AAT TGG AG -3', LEGCYCIB-GW2: 5'- GTC TTG TTT SGG CAT TGW AGC AG -3') and downstream (LEGCYCI-RGW1: 5'- GGA ATG CAT TGT GAT MAR GAG AAA RTT GAA GC -3', LEGCYCI-RGW2: 5'- CAG CAT GAA TCT MTC WAC AGG TAT -3') of the known fragment were used in combination with nested adaptor-specific primers (AP1 5'- GGA TCC TAA TAC GAC TCA CTA TAG GGC-3', AP2 5'- AAT AGG GCT CGA GCG GC - 3' (G. Ingram, pers. comm.)). Location of the LEGCYC specific genome walking primers are given in appendix 2. Products were gel extracted using Qiaquick gel extraction kit (Qiagen Ltd, Dorking, Surrey, UK), and sequenced directly using the nested gene specific primers, or cloned into the pCR4-TOPO vector (Invitrogen Ltd, Paisley, UK).

PCR method	Template	Template amount in PCR	Primers	PCR cycling conditions
Inverse PCR	Self-ligated DNA fragments	3µl of purified circular DNA fragments	1st PCR: iR3-F3, iR4-F3	94°C 3 min 94°C 1 min 55°C 30 s 72°C 2.5 min } x 35 72°C 5 min
		1µl of a $1/10^3$ dilution of 1st PCR	nested PCR: iF1-iR1	as above
'Standard' PCR	Genomic DNA	20 – 30 ng genomic DNA	5': F5-R3, F5-R4 3': iR3-R5, iR4-R5 iR3-R8, iR4-R8	94°C 3 min 94°C 1 min 55°C 30 s 72°C 1min } x 30 72°C 5 min
Genome walking	Adaptor-ligated DNA fragments	1µl of ligated DNA	1st PCR: LEGCYC1_GW1-AP1 (5' end) LEGCYC1_RGW1-AP1 (3' end)	94°C 2 min (hot start) 94°C 3 s 68°C* 3 min } x 8 94°C 3 s 61°C 3 min } x 24 61°C 10 min
		1µl of a $1/10^2$ dilution of 1st PCR	nested PCR: LEGCYC1A_GW2-AP2 (5' end) LEGCYC1B_GW2-AP2 (5' end) LEGCYC1_RGW2-AP2 (3' end)	94°C 2 min (hot start) 94°C 3 s 65°C* 3 min } x 8 94°C 3 s 58°C 3 min } x 24 58°C 10 min

Table 3-1. Summary of the different PCR approaches used to isolate regions flanking known fragments of two *CYC*-like genes, LEGCYC1A and LEGCYC1B, in *Lupinus nanus* and *Cadia purpurea*. Details of template preparation for inverse PCR and genome walking are given in sections 3.2.2a and 3.2.2c respectively. Primer sequences and location are given in appendix 2. PCR mix was as follows in all reactions: sterile distilled water, polymerase buffer, $MgCl_2$ (2.5mM), dNTP's (20µM), primers F1 and R1 (0.5µM each), 1 unit *Taq* polymerase (Bioline Ltd., London NW2, UK). * The annealing/extension temperature is decreased by 1°C per cycle for the first eight cycles of the genome walking PCRs.

3.2.3 Sequence compilation and comparison

Sequence fragments obtained from these various PCR methods were assembled using AutoAssembler (PE Applied Biosystems, Foster City, CA). Overlapping region identity strongly suggested that the different fragments belonged to the same locus. Predicted protein sequences of the ORF were aligned using CLUSTAL X (Thompson *et al.*, 1997), followed by manual adjustments. The start and end of the ORF were identified by comparison with *Lotus japonicus* and *Glycine max* sequences (D. Luo, pers. comm.). Pairwise sequence divergence was calculated using PAUP* v4.0b10 (Swofford, 2001).

3.2.4 Characterisation of intron and splice site

Translation of nucleotide sequences into amino acids suggested that both LEGCYC1A and LEGCYC1B in *C. purpurea* and *L. nanus* have a putative intron close to the end of the ORF. cDNA of both genes in both taxa was sequenced to characterise these introns.

Total RNA extraction from young flower buds of *C. purpurea* and *L. nanus* was carried out using QIAGEN Rneasy mini kit (Qiagen Ltd, Dorking, Surrey, UK). Complementary DNA (cDNA) was synthesised with QIAGEN Omniscript RT kit (Qiagen Ltd, Dorking, Surrey, UK), with added RNasin RNase inhibitor (Promega Ltd, Southampton, UK), using an oligo-T primer (18 bp). LEGCYC1A and LEGCYC1B were amplified using locus specific primers (LEGCYC_iR4, LEGCYC_iR3) in combination with the general primer LEGCYC_R8 located downstream of the putative intron region (described section 3.2.2b). Products were either cloned into pCR4-TOPO (Invitrogen Ltd, Paisley, UK) or sequenced directly.

As the location of the splice site was ambiguous, it was predicted using a programme available on the NetPlantGene server (http://www.cbs.dtu.dk/services/NetPGene/), which uses a method combining global and local sequence information designed for predicting intron splice sites in *Arabidopsis thaliana* (Hebsgaard *et al.*, 1996).

3.2.5 Characterisation of the 3'-end of other LEGCYC genes in *C. purpurea* and *L. nanus*, with particular reference to LEGCYC2

To test the range of the reverse primers LEGCYC_R5 and LEGCYC_R8, PCR was carried out using the forward primer LEGCYC_F3 in the TCP domain, which based on sequence data, binds to LEGCYC1A, LEGCYC1B and LEGCYC2 in *C. purpurea* and *L. nanus*. The reverse primer LEGCYC_R1 in the R domain, known to bind to at least LEGCYC1A, LEGCYC1B and LEGCYC2 in those two species, was used with primer LEGCYC_F3 as a control (figure 3-1). Products amplified using primers LEGCYC_F3-LEGCYC_R8 were cloned into pCR4-TOPO (Invitrogen Ltd, Paisley, UK), then sequenced.

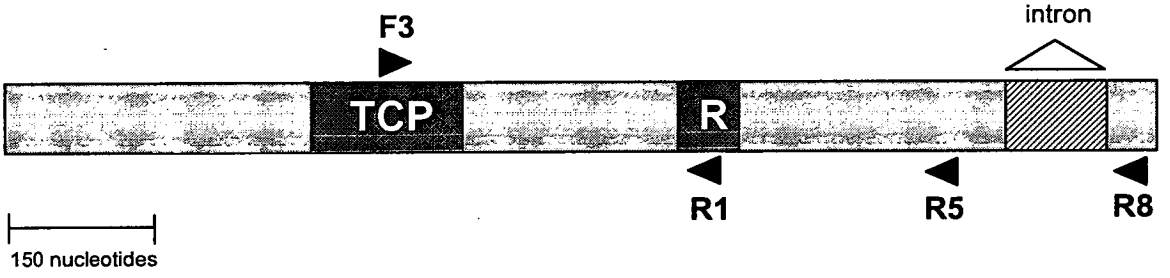


Figure 3-1. Schematic representation of the LEGCYC open reading frame (ORF), showing the TCP and R domains, and the short intron. The binding sites of general primers LEGCYC_F3, LEGCYC_R1, LEGCYC_R5 and LEGCYC_R8 are shown.

Locus specific forward primers for LEGCYC2 were designed to bind to the known region between the TCP and R domains in *C. purpurea* (Cadia 3) and *L. nanus* (Lupinus nanus 3). Primer LEGCYC_F10: 5'- SAW CRA CAC RTC AAA TGA G -3', was designed to bind to LEGCYC2 of both *C. purpurea* and *L. nanus*, and is slightly degenerate, whereas LEGCYC_F12: 5'- GAG AAA GTA GCA TCA TTG - 3', is specific to *L. nanus* LEGCYC2 only and has no degenerate bases. These were used in combination with the reverse primer LEGCYC_R8. In addition, a new reverse primer LEGCYC_R9: 5'- TTC CAA AGA TTT GAA

GCT -3', also downstream of the intron, was designed using the *C. purpurea* LEGCYC2 sequence (see appendix 2 for primer location).

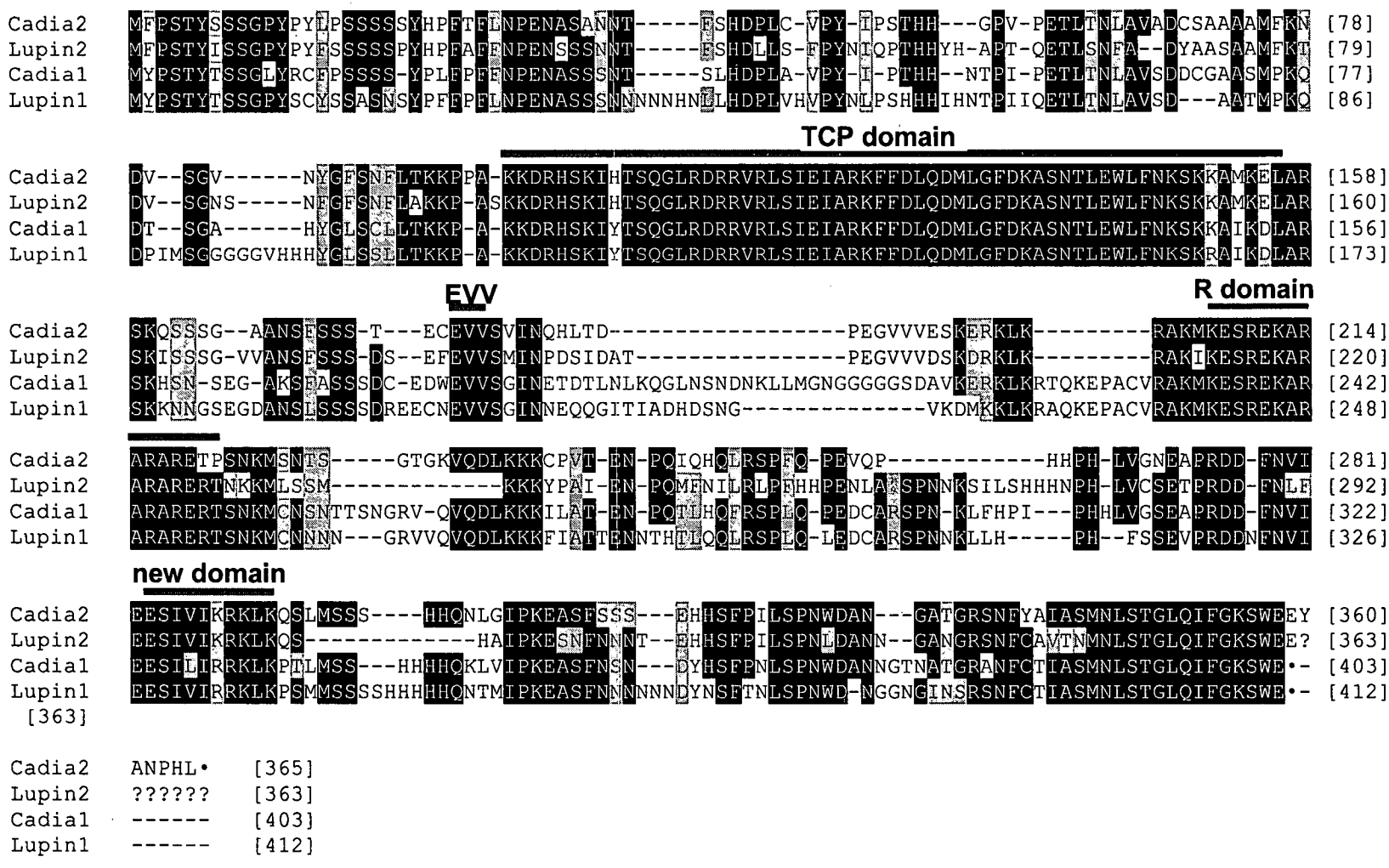
3.3 RESULTS

3.3.1 Characterisation of LEGCYC1A and LEGCYC1B in *Cadia purpurea* and *Lupinus nanus*

Compiled sequences of LEGCYC1A and LEGCYC1B from *Cadia purpurea* and *Lupinus nanus* are given in appendix 5. Results from cDNA sequencing and splice site predictions suggest that all four genes have an intron located in the same region. Intron size ranged from 80 bp (Lupinus nanus 2; LEGCYC1A) to 103 bp (Cadia 1; LEGCYC1B). Predicted protein sequence length ranged from 365 (Cadia 2; LEGCYC1A) to 410 (Lupinus nanus 1; LEGCYC1B) amino acids. The predicted protein sequences of *C. purpurea* LEGCYC1A and LEGCYC1B did not contain any frame-shift or premature stop codons. An amino acid alignment is given in figure 3-2.

In addition to the TCP and R domains, another domain downstream of the R domain, known in *Antirrhinum DICH* (sequence ESIMIKRKL) but absent in *CYC*, was identified in all LEGCYC copies, including LEGCYC2 ("new domain", figure 3-2). Protein secondary structure prediction, using NNPREDICT (Kneller *et al.*, 1990), suggests this region has a helix structure. The EVV domain, between the TCP and R domains (figure 3-2), mentioned in chapter 2 as apparently characteristically absent in LEGCYC2 genes, is also found in *DICH*.

Figure 3-2. An alignment of the predicted amino acid sequence of the complete open reading frame of Cadia 1, Lupinus nanus 1 (LEGCYC1B), Cadia 2, Lupinus nanus 2 (LEGCYC1A). Identical amino acids are in black boxes, while amino acids with similar charge or hydrophobicity are in grey. The TCP and R domains are shown, as well as the EVV motif and another putative helix domain ("new domain") which are both found in *Antirrhinum DICH.*



Sequence analysis over the entire reading frame confirmed that the genes are evolving rapidly by substitutions and insertions/deletions in the regions flanking the conserved TCP and R domains. Nucleotide pairwise distances were greater between LEGCYC1A (82.43% overall sequence similarity) than LEGCYC1B (86.72% sequence similarity) orthologues in *C. purpurea* and *L. nanus*. However, more gaps were required for alignment between *C. purpurea* and *L. nanus* LEGCYC1A than between LEGCYC1B orthologues (15 gaps of 3-36 bp and 26 gaps of 3-45 bp respectively). In addition, different regions within the two loci exhibit different levels of variation, with regions outside the TCP and R domains showing greater sequence divergence than the conserved domains (figure 3-3).

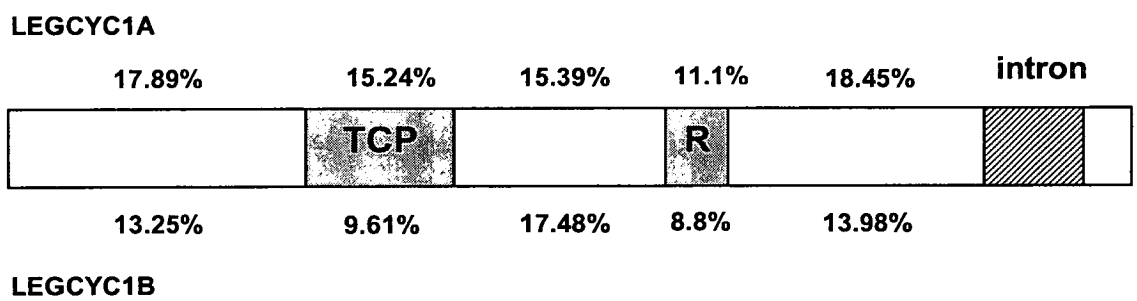


Figure 3-3. Pairwise distances of nucleotide sequences (excluding the intron: hatched region) between *Lupinus nanus* and *Cadia purpurea* LEGCYC1A and LEGCYC1B orthologues respectively. Loci are divided into five regions: three hypervariable regions and the TCP and R domains (in grey).

3.3.2 Investigation of other LEGCYC genes including LEGCYC2

Separate PCRs using the forward primer LEGCYC_F3 in combination with the reverse primers LEGCYC_R1, LEGCYC_R5 and LEGCYC_ R8 all amplified three distinctive bands in *C. purpurea* and *L. nanus* (figure 3-4).

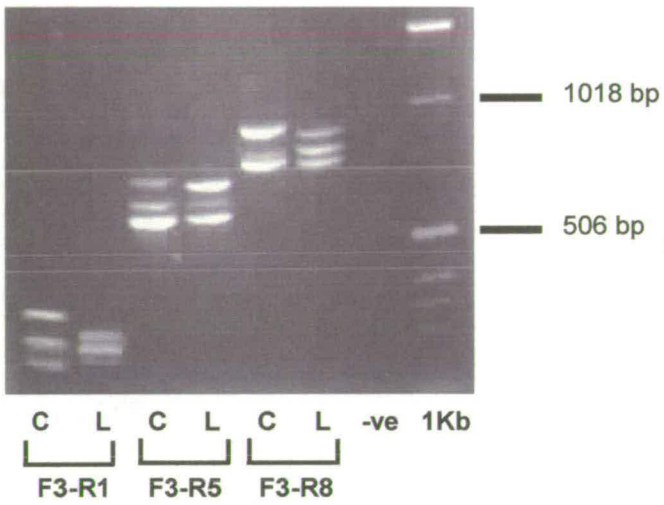


Figure 3-4. PCR products (3µl load) amplified in *Cadia purpurea* and *Lupinus nanus* using the forward primer in the TCP domain LEGCYC_F3 in combination with LEGCYC_R1 (in the R domain), LEGCYC_R5 and LEGCYC_R8 (3' of the intron). All primer combinations amplify three distinct bands in both taxa. C: *Cadia purpurea*, L: *Lupinus nanus*, –ve: negative control (no DNA in sample); 1Kb: 1Kb ladder (Bioline Ltd., London NW2, U.K.).

Cloned products amplified with LEGCYC_F3 and LEGCYC_R8, corresponding to two of the three PCR fragments of distinct size, were identified as being either LEGCYC1A or LEGCYC1B using gene specific primers in the PCR screen. In *C. purpurea*, sequences from multiple clones of the third band were found to be identical, in the region of overlap, to the LEGCYC2 fragment Cadia 3 (see appendix 6 for sequence). Sequence analysis suggested that, as with LEGCYC1A and LEGCYC1B, LEGCYC2 also has an intron at the 3'-end of the gene.

In *L. nanus*, however, clones that were neither LEGCYC1A nor LEGCYC1B were found to be a new *CYC*-like sequence that was similar to LEGCYC1A (79.72% nucleotide sequence similarity) (see appendix 6 for sequence). The level of divergence, and the putative insertions and deletions between LEGCYC1A and this new sequence (7 gaps of 3-30 bp) strongly suggest the latter to be an additional locus. This new copy LEGYC1A* may be the result of a further duplication event in *L. nanus* (further analysed in chapter 5).

Specific amplification of the 3'end of LEGCYC2 was straightforward in *C. purpurea*, using the locus-specific primer LEGCYC_F10 in combination with LEGCYC_R8. The resulting single band was sequenced directly and found to be identical to the LEGCYC2 cloned sequences described above. However, attempts to amplify the 3'-end of LEGCYC2 in *L. nanus* were not successful. No product was visible for *L. nanus* using LEGCYC_F10 with either LEGCYC_R8 or the new reverse primer LEGCYC_R9. Primer LEGCYC_F12, despite exactly matching a region between the TCP and R domain of *L. nanus* LEGCYC2, did not amplify well, or at all, in combination with LEGCYC_R1, LEGCYC_R5 or LEGCYC_R9. Amplification using LEGCYC_F12 and LEGCYC_R8 resulted in a single band, but sequencing of this ~ 500 bp fragment revealed this was a portion of chloroplast DNA.

3.4 DISCUSSION

Complete sequence characterisation can reveal certain aspects of gene function and evolution. For instance, the absence of frame shifts or premature stop codons in the predicted protein sequences of *Cadia purpurea* LEGCYC1A and LEGCYC1B suggests that both copies are theoretically functional, and does not support the hypothesis that the radial symmetry of *C. purpurea* flowers evolved by complete loss-of-function of these *CYC*-like genes. In addition, the presence of cDNA transcripts from both copies in young flower buds of *C. purpurea* indicates these genes are florally expressed, which is also the case for *Lupinus nanus*. A more in depth examination of gene expression is described in chapter 5.

As reported in chapter 2, LEGCYC genes in the Papilionoideae are evolving rapidly by nucleotide substitution as well as by insertions and deletions. It is apparent here that this is the case not only in the region between the conserved TCP and R domains, but also upstream and downstream of these regions. Different levels of nucleotide sequence variation between the TCP and R domains and the other regions suggest that different portions of the gene may be evolving under different modes of molecular evolution. Similar patterns of rapid gene evolution have been observed *CYC* orthologues in the Antirrhineae (Gübitz *et al.*, 2003; Hileman & Baum, 2003) and the Gesneriaceae (Citerne *et al.*, 2000). Within this general pattern of rapid evolution, variation in the rate of nucleotide substitutions and the number of insertion and deletion events were apparent between LEGCYC1A and LEGCYC1B. These patterns of molecular evolution are examined further in chapter 4.

CHAPTER 4: EXPRESSION PATTERNS OF *CYC*-LIKE GENES IN *LUPINUS NANUS* AND *CADIA PURPUREA*

4.1 INTRODUCTION

4.1.1 Comparative gene expression in closely related taxa

Comparative genetic studies between closely related species that differ in a particular trait of interest have been advocated by Baum *et al.* (2002) as a strategy for understanding the genetic basis of morphological change. Two *CYC*-like genes, LEGCYC1A and LEGCYC1B, have been identified by phylogenetic analyses as candidate genes for the control of floral symmetry in two closely related legume taxa that differ in their floral symmetry, *Cadia* (*C. purpurea*) and *Lupinus* (*L. nanus*) from the genistoid clade of papilionoid legumes (Pennington *et al.*, 2001) (chapter 2). Their expression pattern is investigated here by RNA *in situ* hybridisation and reverse transcription (RT)-PCR.

4.1.2 Expression of *CYCLOIDEA* (*CYC*) and *DICHOTOMA* (*DICH*)

In *Antirrhinum majus*, the floral symmetry genes *CYC* and *DICH* have overlapping expression in the adaxial region of the developing flower. *CYC* and *DICH* transcripts are detected in the floral meristem prior to any sign of asymmetry at the junction between the flower and inflorescence meristem, with *DICH* activated slightly before *CYC* (Luo *et al.*, 1996; Luo *et al.*, 1999). In the early phases of floral development, *CYC* is detected in the dorsal sepal and adjacent adaxial part of floral dome. In later stages, *CYC* expression becomes more concentrated in the dorsal petals and dorsal staminode (figure 4-1; Luo *et al.*, 1996). *DICH* expression is similar to *CYC* at the early stages of floral development, but at later stages becomes restricted to the dorsal half of each dorsal petal (Luo *et al.*, 1999).

The aim of this study is to see whether a pattern of expression similar to that of *Antirrhinum CYC* is found in *CYC* homologues in papilionoid legumes with typical strongly zygomorphic papilionaceous flowers, and if these patterns differ between closely related papilionoid species with actinomorphic and zygomorphic flowers.

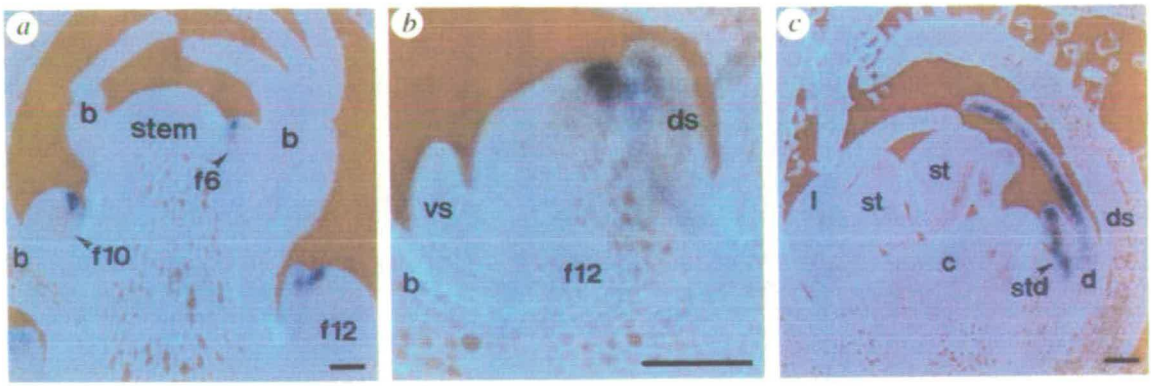


Figure 4-1. RNA *in situ* hybridisation of longitudinal sections of wild type *Antirrhinum* inflorescence (a) and flowers (b, c) probed with *CYC*. A signal can be detected in the adaxial region of the floral meristem prior to organogenesis through to organ differentiation. At early stages, the signal can be detected in the adaxial sepal primordia and the dorsal region of the floral dome (b). At later stages, the signal is detected in the dorsal petal and staminode (c). b: bract, ds: dorsal sepal, vs: ventral sepal, d: dorsal petal, l: lateral petal, st: stamen: std: staminode, c: carpel. Scale bar 100 μm. Reproduced from Luo *et al.*, 1996.

4.2 MATERIALS AND METHODS

4.2.1 RNA *in situ* hybridisation

4.2.1a Tissue fixation

Individual *Cadia purpurea* flower buds were collected at different stages of development, ranging from 2 to 5 mm in length. The hard bracts enclosing the flower were removed prior to overnight fixation in FAA (2% formaldehyde, 5% HOAc, 60% ethanol). Bracts were removed from whole *Lupinus nanus* inflorescences, which were then fixed overnight in either FAA or 4% paraformaldehyde (PFA). A vacuum was applied to the samples for 10 minutes, repeated at least three times, to ensure that the fixative infiltrated the tissue. Material was then dehydrated through an ethanol series and embedded in Paraplast X-tra (Structure Probe Inc./SPI Supplies, West Chester, PA, USA). Details of tissue fixation and embedding protocols are given in appendix 1B. 7-10 μm longitudinal (*L. nanus*) and transverse (*C. purpurea*) sections were fixed onto pre-coated Polysine microscope slides (BDH, Poole, UK).

4.2.1b Probe synthesis

DNA segments from *L. nanus* and *C. purpurea* LEGCYC1A and LEGCYC1B, located near the start of the ORF, were used as gene specific templates for *in situ* hybridisation. These were amplified using primers LEGCYC_F5-LEGCYC_R4 and LEGCYC_F5-LEGCYC_R3 respectively (described in chapter 3 and appendix 2). In addition, a histone gene from *C. purpurea*, homologous to *Sesbania rostrata* histone 4 locus 1 (GenBank accession no. Z79637) and amplified using primers 5'- AAC CAT GTC TGG AAG AGG -3' (forward) and 5'- TAT CTA ACC GCC RAA WCC -3' (reverse), was used as a positive control for *C. purpurea* samples (sequence given in appendix 6). Digoxigenin-labelled sense (*i.e.* negative control) and antisense RNA probes were generated using either T3 or T7 polymerases from linearized

templates cloned into pCR4 plasmids (Invitrogen Ltd, Paisley, UK). Details of protocols are given in appendix 1C.

4.2.1c RNA hybridisation

Two separate RNA *in situ* hybridisation experiments were carried out on *L. nanus* inflorescences, at the John Innes Centre (JIC), Norwich, and at the Institute of Cell and Molecular Biology (ICMB), University of Edinburgh, on tissue fixed in either FAA (JIC) or PFA (ICMB). RNA *in situ* hybridisation of *C. purpurea* material, fixed in FAA, was carried out at ICMB. The protocol followed at the JIC was similar to that of Bradley *et al.*, 1993. The protocol followed at ICMB was similar to that from the Barton laboratory (http://www-ciwdpb.standford. edu/ research/barton/in_situ_protocol.html), and was similar to the one followed at JIC (protocols given in appendix 1D).

4.2.2 Reverse transcription (RT)-PCR

4.2.2a RNA extraction and cDNA synthesis

Total RNA was extracted from a range of tissue from *L. nanus* and *C. purpurea*, including young flowers (< 2mm diameter), dissected older flowers, and vegetative leaves, using QIAGEN Rneasy mini kit (Qiagen Ltd, Dorking, Surrey, UK). Complementary DNA (cDNA) was synthesised using QIAGEN Omniscript RT kit (Qiagen Ltd, Dorking, Surrey, UK), with added RNasin RNase inhibitor (Promega Ltd, Southampton, UK) and using an oligo-T primer (18 bp). Dissected flowers from *L. nanus* and *C. purpurea* were at a comparable stage in development, their size approximately half that of mature flowers where individual organs could be easily removed to prevent cross-tissue contamination. To increase yield, tissue from three or four flowers from *L. nanus* at the same developmental stage was combined for each extraction. In *C. purpurea*, RNA was extracted from tissue from a single flower. This was carried out to

prevent combining tissues from different domains, as it can be difficult to determine the orientation of these radial flowers in bud. RNA was extracted from the four floral whorls in both *L. nanus* and *C. purpurea*. Dissections of mature flowers of *Ulex europaeus* L., a close relative of *Lupinus* within the tribe Genisteae *sensu stricto* with similarly typical papilionoid flowers, and *C. purpurea*, are shown (figure 4-2).

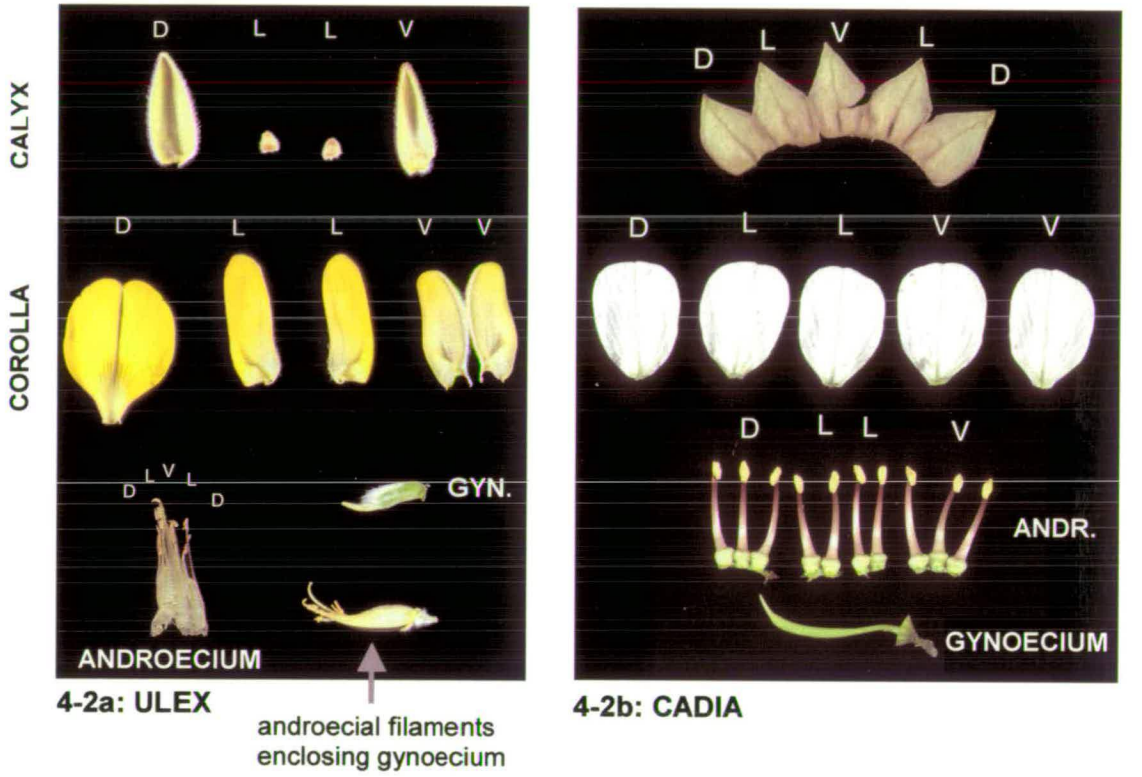


Figure 4-2. Dissected mature flowers of *Ulex europaeus* (4-2a), a close relative of *Lupinus* with similar typical papilionoid flowers, and *Cadia purpurea* (4-2b). Organs in the three outer whorls are divided into dorsal (D), lateral (L) and ventral (V) domains. Strong differentiation in the calyx, corolla and androecium (ANDR) is found in typical papilionoid flowers such as those of *Ulex*, whereas no differentiation is observed in these whorls in *C. purpurea*. The gynoecium (GYN) in both taxa is typical of the Papilionoideae.

In *L. nanus*, RNA was extracted from the ventral and reduced lateral sepals combined, while RNA from the dorsal sepals was extracted separately. RNA from the standard, wing and keel petals was extracted separately. The androecium of *L. nanus* is monadelphous, *i.e.* all ten filaments form a partially fused cylinder. The adaxial three filaments, separated from the rest of the androecial tube by a groove on either side, were excised and RNA from these was extracted separately from the remaining lateral and ventral seven stamens. In *C. purpurea*, floral orientation was determined by the curvature of the gynoecium, which is the only floral organ in this species with clear dorso-ventral asymmetry (see figure 4-2). The gynoecium in *C. purpurea* is like that of typical papilionoid legumes throughout development, with fused carpel margins on the adaxial side and pronounced dorso-ventral curvature apparent during organ elongation (Tucker, 2002; Tucker 2003). Nevertheless, to prevent any mis-identification of the dorsal region, RNA was extracted from each petal separately. The androecium of *C. purpurea* was divided into three parts, consisting of the top three (dorsal), the bottom three (ventral) and the remaining four (lateral) stamens. The calyx was also divided into three parts, with the two dorsal sepals and the two lateral sepals combined respectively. RNA was also extracted from the gynoecium in both taxa. RT-PCR was carried out using RNA from two (*L. nanus*) to four (*C. purpurea*) separate extractions as described above, to ensure that results could be replicated

4.2.2b RT-PCR

The amount of RNA in each sample was normalised by comparing the band intensity on a 1% agarose gel of the housekeeping gene actin amplified by reverse transcription (RT) PCR. To ensure that the amount of amplified products was visualised prior to PCR saturation, aliquots were taken after 20, 25 and 30 cycles. Actin products are either shown here after 25 cycles (amplification from *C. purpurea* petals, androecium and gynoecium) or 30 cycles (amplification from *C. purpurea* sepals, young flower and leaves, and all tissues from *L. nanus*), whereas

LEGCYC products are shown here after 30 cycles. PCR cycling conditions were as follows: an initial denaturation step at 95°C (3 minutes), followed by 10 cycles of denaturation at 94°C (1 minute), annealing at 55°C (1 minute) and extension at 72°C (1 minute), followed by 20 cycles of denaturation at 94°C (1 minute), annealing at 55°C (45 seconds) and extension at 72°C (45 seconds), and a final extension step 72°C (7 minutes). Actin was amplified using the primers 5'-GCG ATA ATG GAA CTG GAA TGG - 3' (forward) and 5'- GAC CTC ACT GAC TAC CTT ATG -3' (reverse) (K. Coenen, ICMB, pers. comm.). To confirm that the primers were actin specific, cDNA products amplified with these were sequenced directly in both *L. nanus* and *C. purpurea* (sequences given in appendix 6). LEGCYC genes were amplified using locus specific primers LEGCYC_iR3 (LEGCYC1B), LEGCYC_iR4 (LEGCYC1A), and the reverse primer LEGCYC_R8 (described in chapter 3 and appendix 2). Both actin and LEGCYC primers span an intron region that distinguish cDNA from genomic DNA. LEGCYC product identity was confirmed by sequencing of RT-PCR products in *L. nanus* and *C. purpurea* (see chapter 3).

Despite not being able to amplify the 3'end of LEGCYC2 in *L. nanus* (chapter 3), RT-PCR was carried out for *C. purpurea* LEGCYC2 using the locus-specific forward primer LEGCYC_F10 (described in chapter 3 and appendix 2) and the reverse primer LEGCYC_R8 as described above.

4.3 RESULTS

4.3.1 RNA *in situ* hybridisation

LEGCYC1A and LEGCYC1B RNA was detected in floral tissue of *L. nanus* (figures 4-3: whole inflorescence, figures 4-4 and 4-5: details of individual developing flowers), in a pattern similar to *Antirrhinum CYC* (Luo *et al.*, 1996). Both genes were detected in floral meristems prior to organogenesis, on the adaxial side of the meristem (figures 4-4a, 4-4c and 4-5a, 4-5c). At more advanced developmental stages, both genes were detected in the corolla (figures 4-4d, 4-4h and 4-5b, 4-5d). Similar to *CYC*, expression of LEGCYC1B in the dorsal petal was found in the inner cell layers at the site where cell division was repressed early in organogenesis (figures 4-1 and 4-4h). Although the expression domains of LEGCYC1A and LEGCYC1B are largely overlapping, suggesting functional redundancy, LEGCYC1A appears to have a reduced expression domain relative to LEGCYC1B. In the developing dorsal petal, for instance, it appears that LEGCYC1A is restricted to the upper part of the petal, whereas LEGCYC1B is expressed throughout the entire length of the petal (figures 4-5b and 4-5d). These results have been replicated at ICMB (figures 4-3 and 4-4) and JIC (figure 4-5).

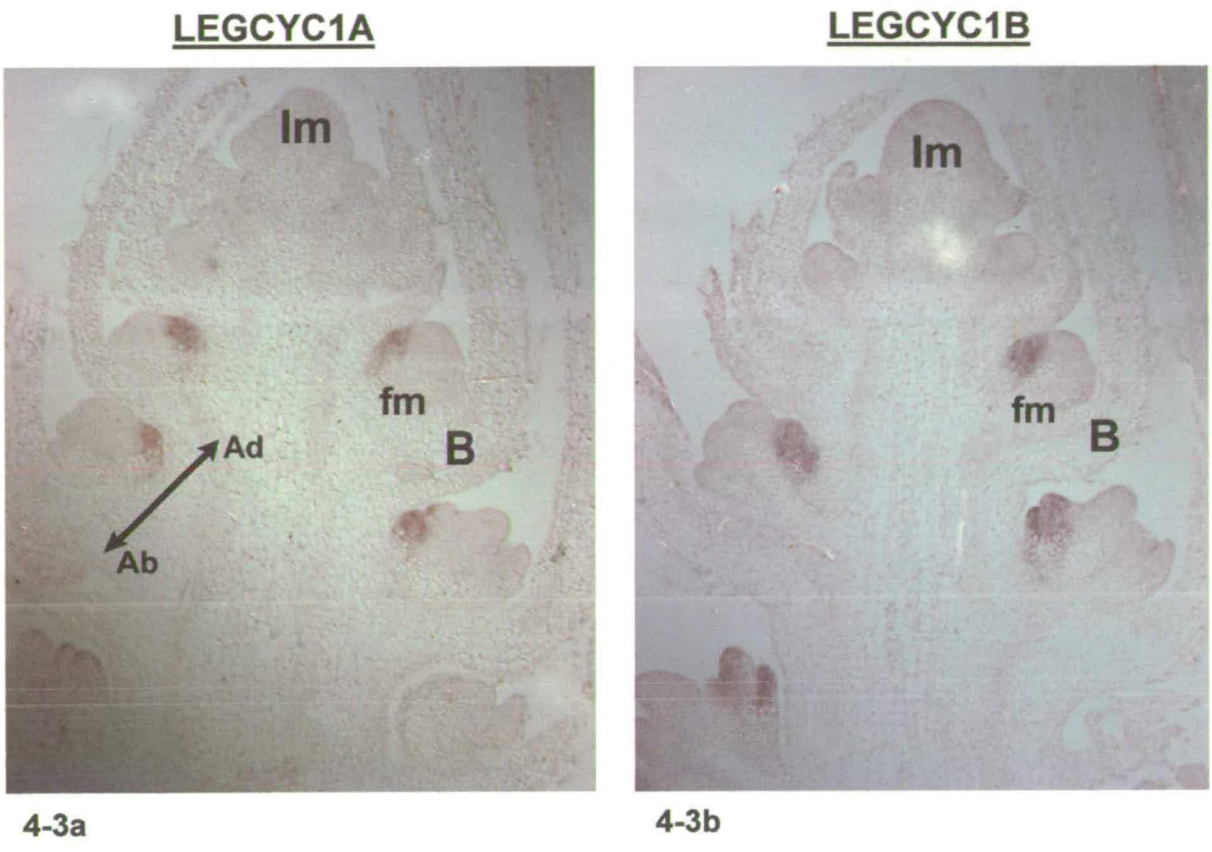


Figure 4-3 Expression pattern of LEGCYC1A (4-3a) and LEGCYC1B (4-3b) in *Lupinus nanus* inflorescences fixed in PFA (hybridisation carried out at ICMB; appendix 1B-D). Longitudinal sections of *L. nanus* inflorescences show floral meristems (fm) in the axil of bracts (B). The adaxial (Ad) and abaxial (Ab) regions are shown in relation to one floral meristem (4-3a). The early stages of organogenesis can be seen in more developmentally advanced flowers at the base of the inflorescence. RNA from LEGCYC1A and LEGCYC1B is detected in the adaxial part of floral meristems prior to organogenesis, as well as during floral organ development. Negative control (sense probe) shown in figure 4-3c.

4-3c

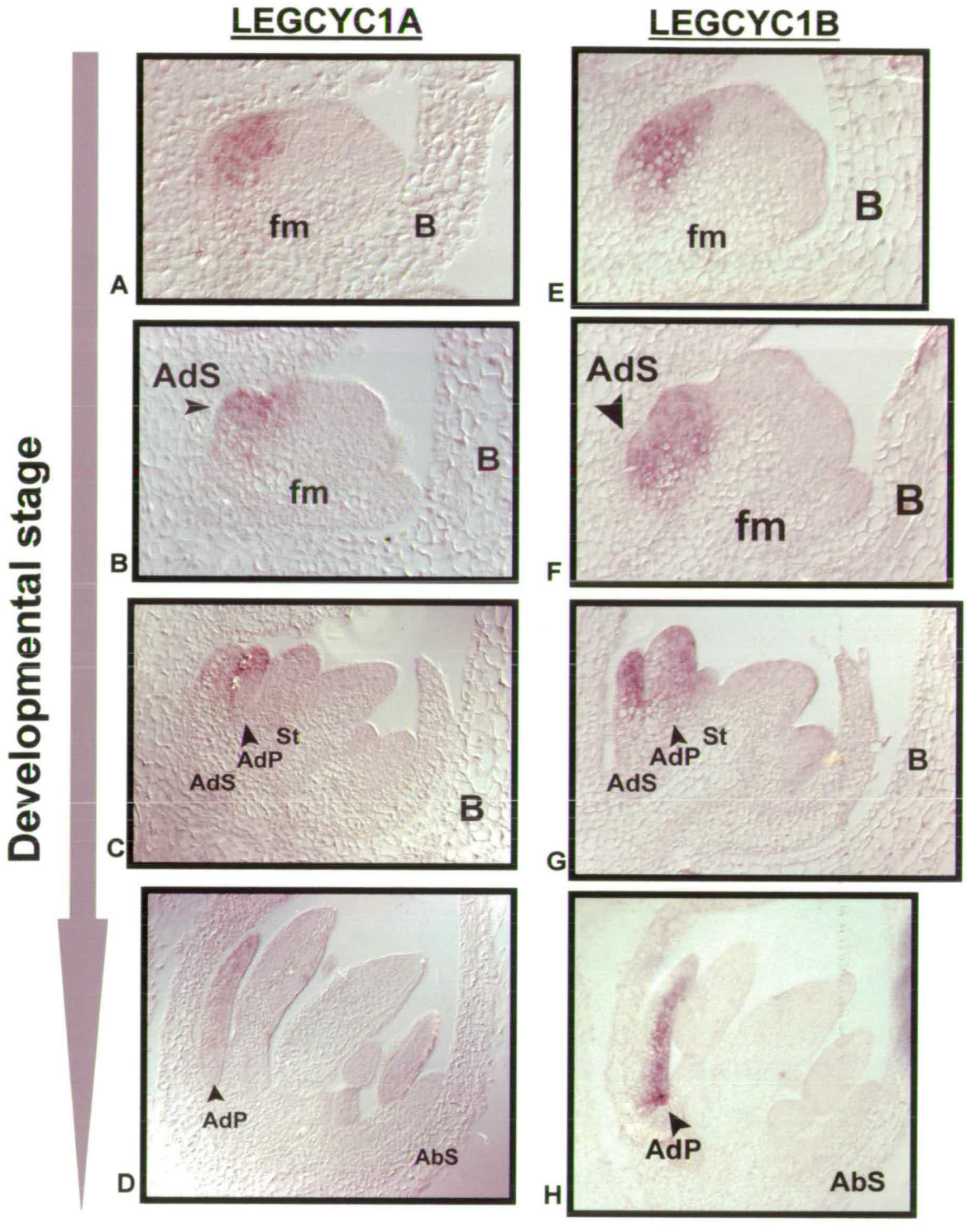


Figure 4-4. RNA *in situ* hybridisation of LEGCYC1A (A-D) and LEGCYC1B (E-H) in the developing flowers of *Lupinus nanus* (hybridisation carried out at ICMB). The flowers are subtended by bracts (B) on the abaxial (ventral) side. Both genes are expressed in the flower meristem (fm) prior to organogenesis (figures A, E), and in the adaxial sepal (AdS) as it develops (figures B, F). In more advanced developmental stages (figures C-D, G-H), expression is found in the adaxial petal (AdP). Although both copies have a similar expression pattern, LEGCYC1B has a wider expression domain than LEGCYC1A, particularly in later developmental stages. St: stamen, AbS: abaxial sepal.

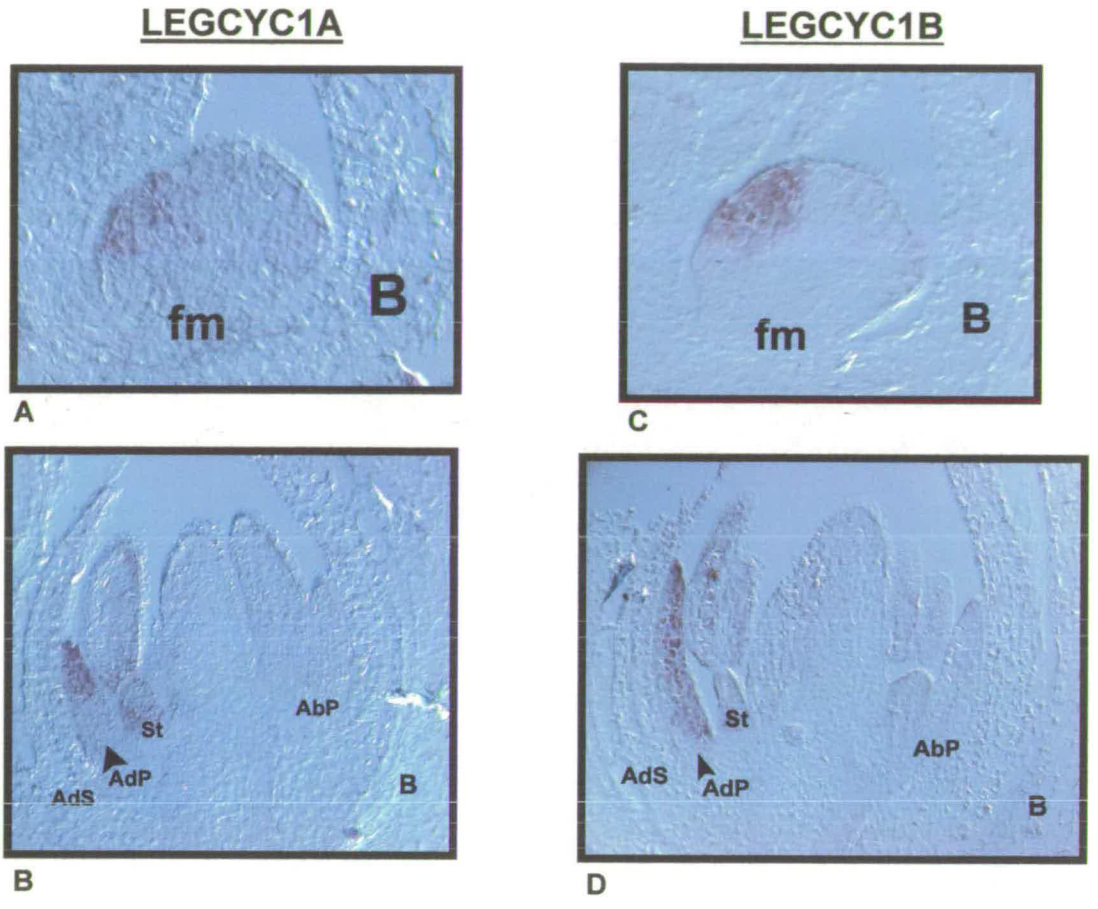
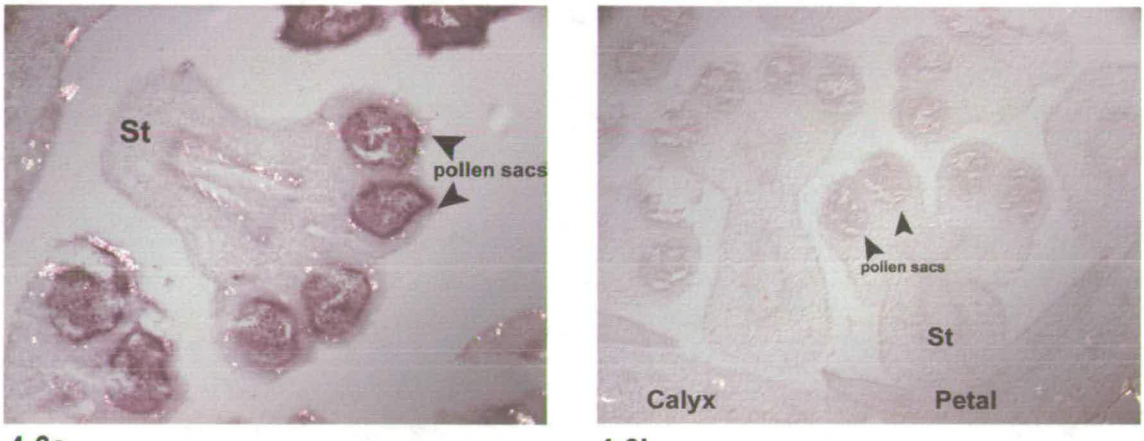


Figure 4-5. RNA *in situ* hybridisation of LEGCYC1A and LEGCYC1B in *Lupinus nanus* flowers fixed in FAA (hybridisation carried out at JIC). Patterns of expression are in agreement with *in situ* hybridisation of LEGCYC1A and LEGCYC1B in inflorescence material fixed in 4% PFA (figures 4-3 and 4-4). As in figures 4-3 and 4-4, LEGCYC1B was found to have a larger expression domain compared to LEGCYC1A, particularly at later stages (B and D). fm= floral meristem, B= bract (subtending the flower on the abaxial side), AdS = adaxial sepal, AdP= adaxial petal, AbP= abaxial petal, St= stamen.

RNA *in situ* hybridisation in *C. purpurea* flower material failed to detect any signal for either LEGCYC gene. However, the histone positive control appeared to have hybridised with the anther locules, an area of intense cell cycling, whereas the negative control, using a sense LEGCYC1B probe, did not produce such a pattern (figure 4-6). Nevertheless, it cannot be ruled out that this signal is the result of background hybridisation.



4-6a 4-6b

Figure 4-6. RNA *in situ* hybridisation in *Cadia purpurea* flower material. Although no hybridisation was detected using either LEGCYC1A or LEGCYC1B antisense probes (not shown), a histone probe used as a positive control (4-6a) may be showing hybridisation in a region of intense cell division, the pollen sacs in the stamens (St), compared to the negative control (using a LEGCYC1B sense probe) (4-6b).

4.3.2 RT-PCR

Size differences predicted by the intron distinguished cDNA and genomic DNA LEGCYC products. The size difference was also unambiguous between cDNA and genomic DNA of the control housekeeping gene actin. Although actin is commonly used as a positive control for RT-PCR, the usefulness of actin as a quantitative RNA marker is complicated by the fact that it belongs to a large gene family of similar proteins (Moniz de Sa & Drouin, 1996). The primers used here, designed for members of the Lamiales (K. Coenen, pers. comm.) amplified two products differentially in *C. purpurea* and *L. nanus* genomic and cDNA. Direct sequencing of actin cDNA revealed that the copies in both taxa were similar in sequence but had numerous double peaks (corresponding to 4.61% of the sequence fragment in *C. purpurea*) suggesting that multiple loci may have been amplified (see appendix 6 for sequences). Although this made comparison and quantification between species problematic, it was possible to compare samples between individuals from the same species.

Comparison of RT-PCR LEGCYC products from young floral (< 2mm diameter) and vegetative (leaf) tissue in *L. nanus* and *C. purpurea* suggests that although both LEGCYC1A and LEGCYC1B are transcribed in immature flower buds, only one locus, LEGCYC1A, appears to be transcribed in developing leaves in both species (figure 4-7). Both genes are transcribed in the flowers of *C. purpurea* at this early developmental stage, refuting the hypothesis that transcription of *CYC*-like genes may have been lost in this actinomorphic species. The difference in the level cDNA amplification between the two paralogues in young *C. purpurea* flowers suggests that LEGCYC1A may be more strongly expressed than LEGCYC1B during the early stages of floral development. On *C. purpurea* genomic DNA, the LEGCYC1B primers appear to work slightly better than those specific to LEGCYC1A (figure 4-7), reinforcing the hypothesis that LEGCYC1B may be less highly expressed than LEGCYC1A during early floral development.

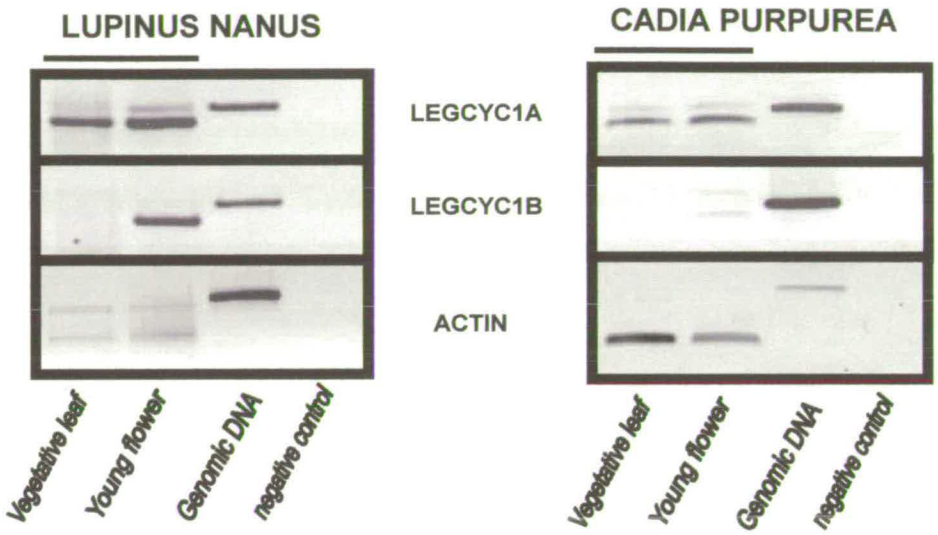


Figure 4-7. RT-PCR analysis of LEGCYC1A and LEGCYC1B expression in developing vegetative (leaf) and floral tissue in *Cadia purpurea* and *Lupinus nanus*, with amplification of actin cDNA used as a control. Lanes with cDNA amplification are marked by a line. Results in *L. nanus* confirm that both LEGCYC1A and LEGCYC1B are florally expressed genes, however LEGCYC1A is also transcribed in vegetative leaf tissue. Results in *C. purpurea* suggests that both LEGCYC copies are expressed florally, with LEGCYC1A also expressed in leaf tissue as in *L. nanus*.

The expression pattern of the LEGCYC genes can be compared in greater detail in dissected flowers. RT-PCR results in *L. nanus* suggest that, in agreement with the findings *in situ*, both LEGCYC1A and LEGCYC1B are expressed only in the adaxial part of the developing flower (figure 4-8). They also suggest that both copies are expressed not only in young flowers, as shown from *in situ* hybridisation, but also at more advanced developmental stages. In addition, both copies are transcribed at this stage not only in the standard (dorsal) petal, but also in the dorsal anthers and sepals (figure 4-8). By contrast, LEGCYC transcripts were detected *in*

situ in all three whorls early in organogenesis, but not in the calyx or androecium at more advanced developmental stages. This difference may reflect the greater sensitivity of RT-PCR compared with RNA *in situ* hybiridisation in detecting transcripts present in lesser abundance.

In *C. purpurea*, RT-PCR from individual floral organs revealed that LEGCYC1A and LEGCYC1B have a very different expression pattern from each other, and from their *L. nanus* orthologues, at this advanced developmental stage. In the corolla of *C. purpurea*, LEGCYC1A is only expressed in the dorsal petal, and its level of expression appears moderate to weak (figure 4-8). LEGCYC1B, however, is expressed in all petals (figure 4-8), suggesting an expansion of the expression domain of this gene which correlates with the radial phenotype of the corolla. These results has been replicated in separate extractions of the corolla from four individual flowers, and therefore seem unlikely to be false positives. In addition, unlike in *L. nanus*, neither LEGCYC1A nor LEGCYC1B appear to be expressed in the androecium of *C. purpurea* (figure 4-8). In the calyx, LEGCYC1A expression, but not LEGCYC1B, was detected in the dorsal and lateral sepals. As in *L. nanus*, no LEGCYC transcripts were detected in the gynoecium.

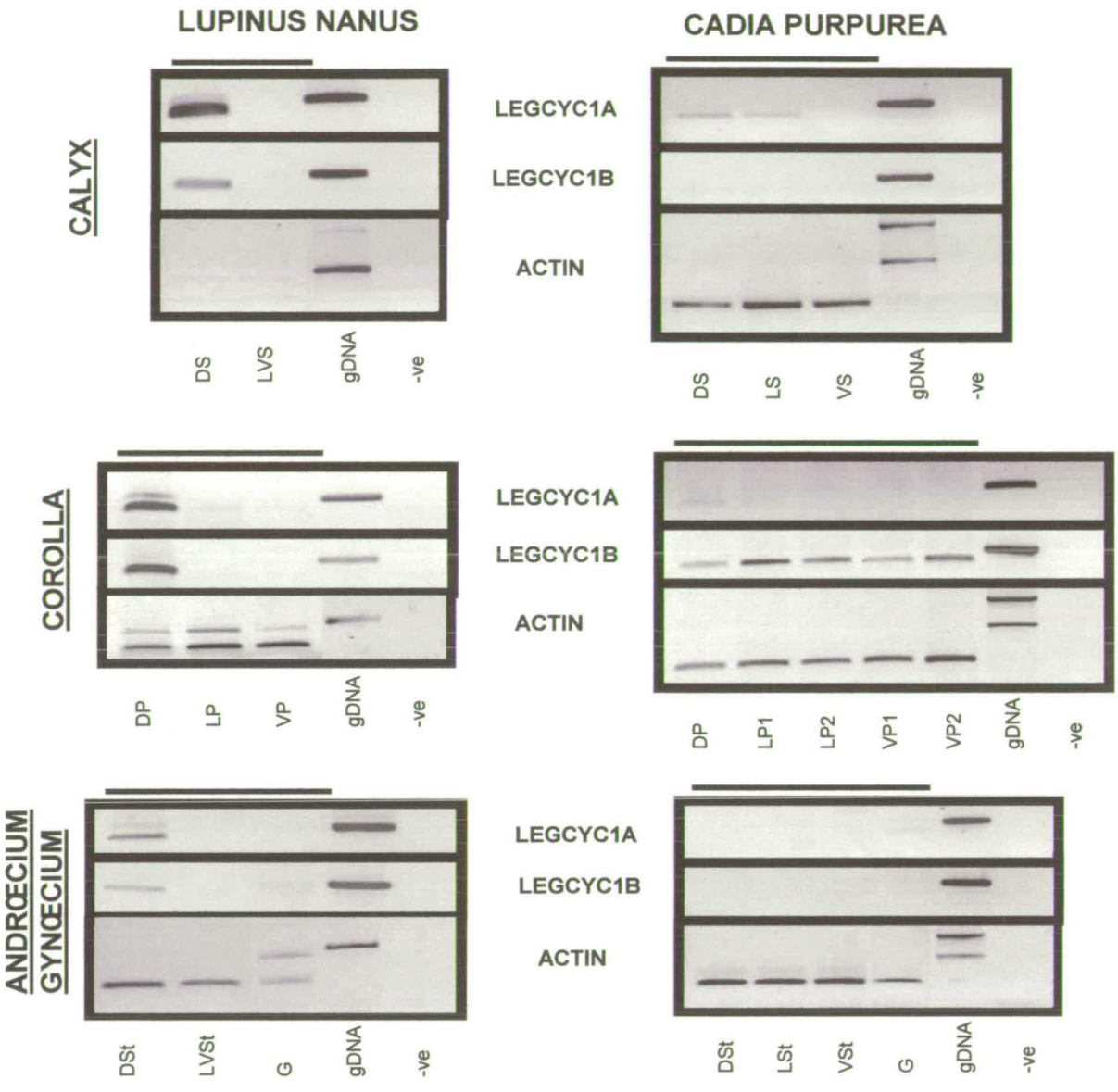


Figure 4-8. RT-PCR analysis of LEGCYC1A and LEGCYC1B expression in the different whorls of the developing flower of *Cadia purpurea* and *Lupinus nanus*, with amplification of actin cDNA used as a control. Results in *L. nanus* are in agreement with the *in situ* hybridisation pattern, with both LEGCYC1A and LEGCYC1B transcribed in the dorsal region. Results in *C. purpurea* suggest that whereas LEGCYC1A is weakly expressed in the dorsal petal, LEGCYC1B is expressed in all petals, and correlates with the lack of differentiation within the corolla. Neither LEGCYC1A nor LEGCYC1B seem to be transcribed in the androecium or gynoecium, whereas LEGCYC1A appears to be transcribed in the dorsal and lateral region of the calyx. DS = dorsal sepal, LS = lateral sepals, VS = ventral sepals, DP = dorsal petal, LP = lateral petal, VP = ventral petal, DSt = dorsal stamens, LSt = lateral stamens, VSt = ventral stamens, G = gynoecium, gDNA = genomic DNA, -ve = negative control. Lanes with PCR products amplified from cDNA are marked by a line.

Although no expression data could be obtained for *L. nanus* LEGCYC2, as attempts to amplify the region spanning the intron of this locus in this species were not successful (see chapter 3), RT-PCR of LEGCYC2 in *C. purpurea* suggests this gene is also florally expressed, albeit weakly (figure 4-9). The size difference between cDNA and genomic DNA corresponds to the predicted intron size (~ 89 bp, see appendix 6). In dissected flowers, LEGCYC2 cDNA was amplified in all domains (dorsal, lateral and ventral) of the calyx and corolla, although here no amplification was detected in one of the ventral petals (figure 4-10). Products were not detected in either the androecium or gynoecium (figure 4-10).

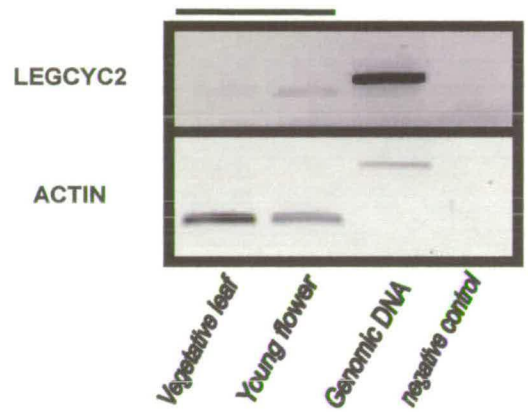


Figure 4-9. RT-PCR analysis of LEGCYC2 expression in developing vegetative (leaf) and floral tissue in *Cadia purpurea*, with amplification of actin cDNA used as a control. Lanes with cDNA amplification are marked by a line. An apparently low level of LEGCYC2 transcripts was detected in floral tissue.

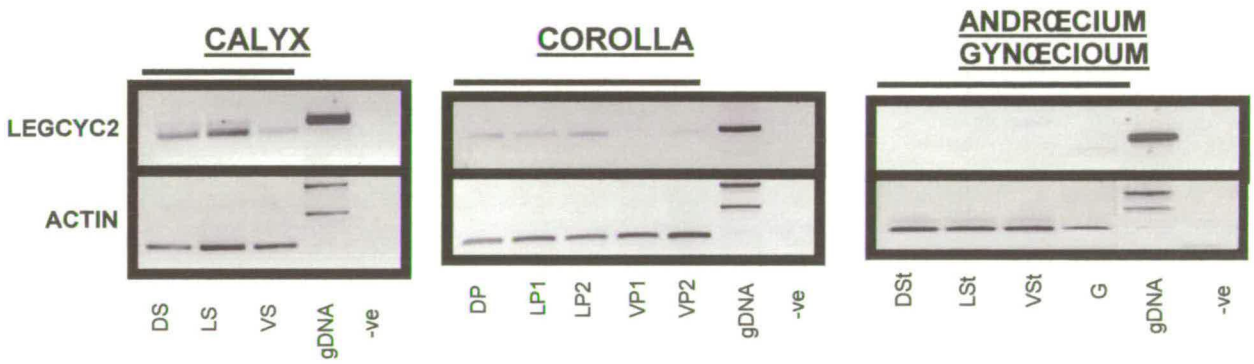


Figure 4-10. RT-PCR analysis of LEGCYC2 expression in the dissected calyx, corolla, androecium and gynoecium of *Cadia purpurea*, with amplification of actin cDNA used as a control. Lanes with cDNA amplification are marked by a line. LEGCYC2 transcripts were detected in the calyx and corolla, with no apparent asymmetry, but not in the androecium or gynoecium. D = dorsal, L = lateral, V = ventral, S = sepal, P = petal, St = stamen, G = gynoecium.

4.4 DISCUSSION

4.4.1 Expression of LEGCYC genes in a typical papilionoid legume *Lupinus nanus*

The two candidate *CYC*-like genes, LEGCYC1A and LEGCYC1B, have been found to be expressed in the dorsal region of the developing flower of *Lupinus nanus* in a pattern highly similar to that *Antirrhinum CYCLOIDEA*, and are therefore strong candidates for the control of floral symmetry in legumes. These results suggest that similar genes could have been recruited more than once for the control of a trait that has evolved independently in distantly related lineages.

The orthologue of *CYC* in *Arabidopsis*, *TCP1*, is also expressed on the adaxial side of the developing flower, as well as in the axillary shoot meristems (Cubas *et al.*, 2001). Unlike *CYC*-like genes in *Antirrhinum* and *L. nanus*, however, the expression of *TCP1* in flowers is transient and only found during the early stages of floral development, and this may account in part for the lack of dorsoventral asymmetry in *Arabidopsis* (Cubas *et al.*, 2001). *Arabidopsis* and *Antirrhinum* belong to two different major clades of eudicots, the Rosidae and Asteridae respectively (Soltis *et al.*, 1999; APG, 2003; see figure 4-11). The occurrence of adaxial expression in axillary meristems of *CYC*-like genes in these two model organisms has led Cubas *et al.* (2001) to suggest that this pattern may pre-date the divergence of the rosid/asterid clade. Adaxial expression of *CYC* othologues in the Leguminosae supports this hypothesis. This asymmetrical "pre-pattern", occurring in the common ancestor of rosids and asterids which presumably had radially symmetric flowers, may therefore have been modified repeatedly to lead to the evolution of complex zygomorphic flowers in such distantly related lineages as Lamiales and Leguminosae. Genetic modifications resulting in the evolution of zygomorphic flowers may have included changes in the timing of gene expression, by extending the length of time the gene is expressed, and interactions with target genes such as floral organ identity genes,

which have been shown in *Antirrhinum* to modulate the specific effects *CYC* has on organ development (Clark & Coen., 2002).

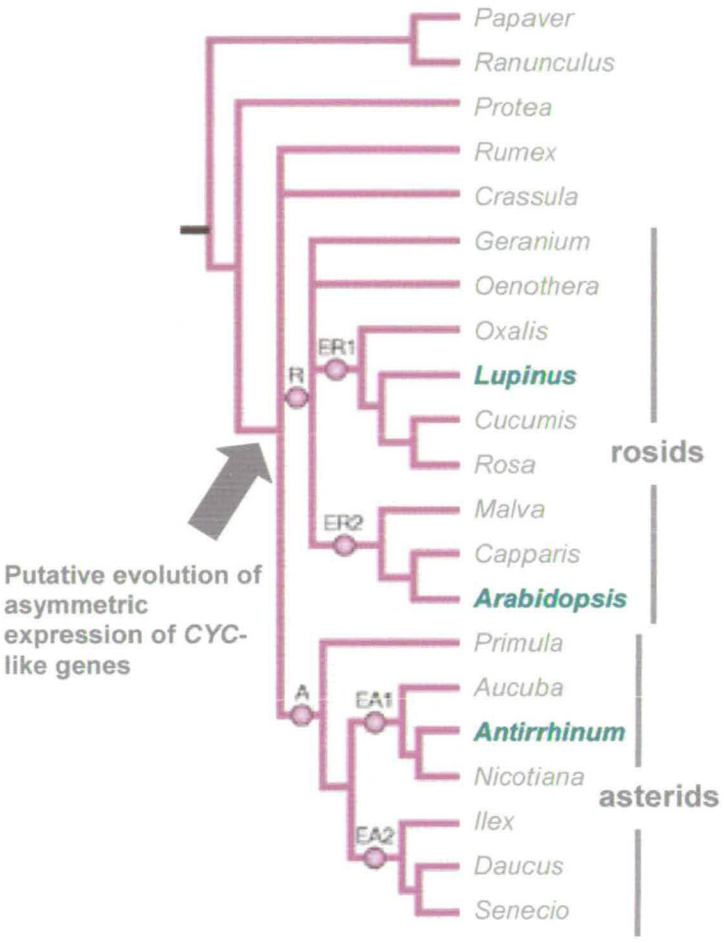


Figure 4-11. Summary of eudicot phylogeny (based on results from Soltis *et al.*, 1999). Representative taxa with known asymmetric expression of *CYC*-like gene in axillary meristems are shown in green. The occurrence of this adaxial expression pattern in distantly related species may suggest that it facilitated the evolution of zygomorphy in distantly related lineages, through modifications of *CYC*-like gene regulation. Phylogeny reproduced from Cronk (2001). R = rosid, ER1 = eurosid 1, ER2 = eurosid 2, A = asterid, EA1 = euasterid 1, EA2 = euasterid 2.

CYC-like genes have been found to evolve rapidly and to have undergone independent duplication events in angiosperm clades such as Antirrhineae (Hileman & Baum, 2003; Gubitz *et al.*, 2003), Gesneriaceae (Citerne *et al.*, 2000), Solanaceae (K. Coenen, unpublished) and the Papilionoideae (Citerne *et al.*, 2003; see chapter 2). In this study, it was found that two

LEGCYC paralogues had largely overlapping expression patterns in developing flowers, and were probably functionally redundant. However, one copy, LEGCYC1A, has a reduced expression domain compared to LEGCYC1B. This partial redundancy is also observed in *Antirrhinum* between *CYC* and *DICH*, where *CYC* has the largest expression domain and greatest effect on phenotype (Luo *et al.*, 1996; Luo *et al.*, 1999). *DICH* has been implicated in the control of petal shape (Luo *et al.*, 1999), and along with *CYC*, contributes to the complex zygomorphic phenotype of wild-type *Antirrhinum* flowers. In addition *CYC*, but not *DICH*, appears to act non-autonomously with a gene involved in lateral identity, *RADIALIS*, promoting the differentiation between ventral and lateral floral organs (Almeida *et al.*, 1997; Luo *et al.*, 1999) The specialised papilionoid flowers, with strongly differentiated standard, wing and keel petals, may also require the expression of the two LEGCYC genes, which may have subtly different effects on phenotype. The effects of LEGCYC1A and LEGCYC1B on development are further investigated by gene silencing (chapter 6).

A study of the molecular evolution of the *CYC/DICH* paralogues in the Antirrhineae (Hileman & Baum, 2003), suggested that both copies have been maintained by complementary sub-functionalisation, *sensu* Lynch and Force (2000), where duplicated genes experience degenerative mutations that reduce their activity so that both copies are required for development. This may also be the case in papilionoid legumes, where long-term maintenance of paralogues without functional divergence has occurred, and could therefore explain in part why duplicated *CYC*-like genes are maintained in the genome. Another possibility is that the two genes have different pleitropic effects. For instance, LEGCYC1A is expressed in vegetative shoots, but not LEGCYC1B. Expression of *Antirrhinum CYC* has also been observed in shoots (Clark & Coen, 2002), however loss of *CYC* function does not have any visible effect on vegetative phenotype.

4.4.2 Radial symmetry in *Cadia* as an evolutionary innovation

Loss of function of *CYC*-like genes results in radial symmetry in *Antirrhinum* and its close relative *Linaria* (Luo *et al.*, 1996; Cubas *et al.*, 1999b). In the Papilionoideae, a number of unrelated genera also appear to have evolved radial symmetry from a zygomorphic ancestral state (Pennington *et al.*, 2000). This study has shown that *CYC*-like genes, based on their expression pattern, are likely to control floral symmetry in this subfamily. Therefore, have these unusual radially symmetric phenotypes in the Papilionoideae evolved by loss of function or by changes in expression of *CYC*-like genes?

Results from *Cadia purpurea* suggest that, although only LEGCYC1A appears expressed in developing leaves, both LEGCYC1A and LEGCYC1B are transcribed in developing flowers. Taking ontogeny into account, this is not surprising as early development of *C. purpurea* flowers is similar to that of most papilionoid species with zygomorphic flowers (Tucker, 2002b). As in *Lupinus affinis* (Tucker, 1984), the sepals, petals, and stamens in *C. purpurea* are initiated unidirectionally, starting on the abaxial side (Tucker, 2002b). Although organogenesis is asymmetric, a phase of uniform organ growth precedes zygomorphic development in papilionoid legumes (Tucker, 2003). Organ differentiation therefore occurs at an advanced stage of floral ontogeny (Tucker, 2003). Considering the development of typical papilionoid flowers, Tucker (2002b) interpreted the phenotype of *C. purpurea* as "neotonous", that is retaining the characteristics of early flower development (*i.e.*, uniform growth) and not undergoing the differentiation phase. In genetic terms, if organ differentiation in typical papilionoid flowers is caused by *CYC* expression during the later stages of floral development, then radial symmetry could be caused by the absence of late *CYC* expression. Molecular data, however, suggest a different interpretation. Rather than failing to develop *CYC* expression during the late stages of flower development, it was found that one gene, LEGCYC1B, is expressed in all five petals of *C. purpurea*. The other copy, LEGCYC1A, is expressed adaxially

but may be down-regulated. The expression of LEGCYC1B in the corolla is reminiscent of the *backpetals* mutation in *Antirrhinum* (Luo *et al.*, 1999). This mutant has ectopic expression of *CYC* in the lateral and ventral petals. A transposon insertion in an AT-rich site ~ 4.2 Kb upstream of start codon is believed to affect a *cis*-acting region that normally suppresses *CYC* transcription during the later stages of development in wild type *Antirrhinum* flowers (Luo *et al.*, 1999). It may be that a change in *cis*-regulation has also led to the expansion of the expression domain of LEGCYC1B in *C. purpurea.*

The occurrence of a putative ancestral state such as radial symmetry within a clade that has a derived character (zygomorphy) is frequently referred to as an "evolutionary reversal" (*e.g.* Endress, 1997). RT-PCR results suggest that from a genetic point of view, however, the radial symmetry of *Cadia* is an evolutionary innovation caused in part by the expansion of the expression domain of a *CYC*-like gene. This change can be considered homeotic as the lateral and ventral petals of *Cadia* have assumed a dorsal phenotype, *CYC* being a marker for dorsal identity (figure 4-12). This interpretation is supported by morphology. In *Cadia*, the five petals are large and bilaterally symmetric, features that are typical of the papilionoid standard petal. By contrast, wing and keel petals in typical papilionoid flowers are asymmetric and small relative to the standard (figure 4-12). Such homeotic-like transformations may play an important role in establishing morphological diversity. In *Mohavea concertiflora*, stamen number is reduced from four to two compared to its close relative *Antirrhinum majus* by expansion of *CYC* and *DICH* expression from the adaxial to the lateral region (Hileman *et al.*, 2003).

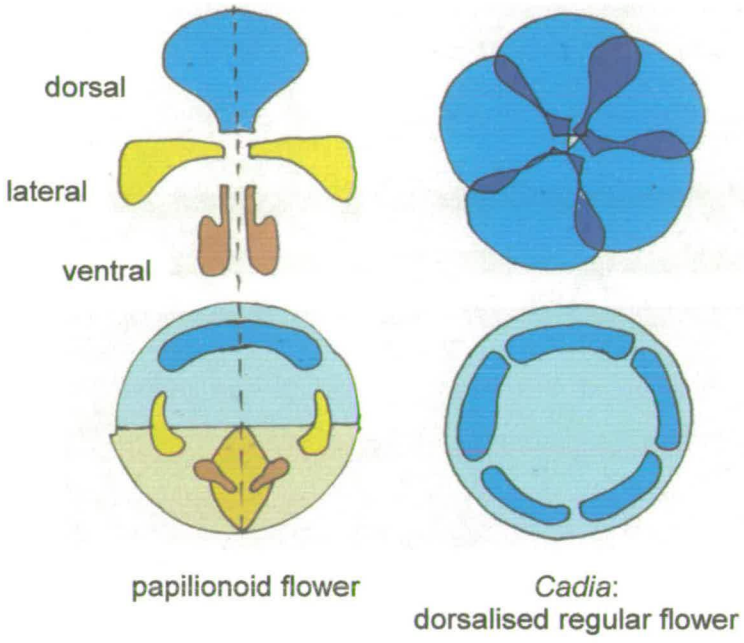


Figure 4-12. Simplified model of the control of symmetry of the corolla in papilionoid legumes. A typical papilionoid flower (left, with only petals shown) can be divided into dorsal, lateral and ventral domains, where LEGCYC is a marker for dorsal identity. The evolution of radial symmetry in the corolla of *Cadia* appears to have resulted from the expansion of the expression domain of one LEGCYC gene, so that all petals have dorsal identity (right).

4.4.3 A complex expression pattern of LEGCYC genes in *C. purpurea*

This simple pattern of either wild type (adaxial) or uniform expression in all organs within a whorl of LEGCYC genes does not hold for the either the calyx or androecium of *C. purpurea*. No transcripts of either gene were detected in the stamens of *C. purpurea*, whereas in *L. nanus*, both are expressed in the adaxial stamen(s). This suggests that unlike in the corolla, the ten free stamens of *C. purpurea* may have developed equally as a result of loss of *CYC* expression. In the calyx, LEGCYC1B was not detected, whereas LEGCYC1A was detected in the dorsal and lateral sepals. This is harder to relate phenotypically, as the sepals are sub-equal.

Investigation of LEGCYC2 expression in *C. purpurea* suggests that this gene may also have a role in flower development, particularly in the calyx and corolla. Without knowledge of LEGCYC2 expression in a typical papilionoid legume such as *L. nanus*, however, it is difficult to speculate what this may be.

4.4.4 Further work

It is clear that confirmation of the expression pattern of LEGCYC1A and LEGCYC1B is required in *C. purpurea.* Attempts at *in situ* hybridisation in *C. purpurea* flowers were not successful due to the nature of the material and also possibly the low level of LEGCYC gene expression. Flowers of *C. purpurea* have small solitary buds (< 0.4 mm after complete organogenesis (Tucker, 2002b)), their sepals covered in trichomes, and contain crystallised material that makes fixative penetration and sectioning particularly difficult without compromising RNA quality. It was found that better sections were obtained from older flower bud material (> 2mm diameter).

Although RT-PCR is a sensitive method of detection of gene expression, it is also prone to false positive results. In addition, comparison of the level of gene expression between species and loci is difficult using a PCR approach as primers may have different binding properties in each case. Nevertheless, a technique such as real-time quantitative RT-PCR would provide a more detailed cDNA amplification profile. Ideally, RNA hybridisation, *in situ* or by Northern analysis, could provide strong evidence for gene expression patterns, although this may be difficult if little template is present as a result of low gene expression or due to the nature of the material, as discussed above.

Initially, this study was focused on the candidate genes LEGCYC1A and LEGCYC1B, as these were known to be expressed asymmetrically in *Lotus japonicus* (D. Luo, unpublished). However, the expression pattern of LEGCYC2 in *C. purpurea* suggests this gene is also florally

expressed and may be involved in the control of floral symmetry. Complete characterisation of this gene in *L. nanus* and subsequent analysis of its expression pattern would allow this hypothesis to be tested.

CHAPTER 5: MOLECULAR EVOLUTION OF LEGCYC GENES IN THE GENISTOID CLADE

5.1 INTRODUCTION

The evolution of morphological differences between species has been related to changes in the function of regulatory loci (*e.g.* Doebley & Lukens, 1998; Lamb & Irish, 2003). One way such changes can come about is through modifications in gene regulation, altering the spatial and/or temporal pattern of expression. This appears to have occurred in *CYC*-like genes on numerous occasions (*e.g. Mohavea*; Hileman *et al.*, 2003), including in the Leguminosae (chapter 4). Another way is through changes in protein function, such that interactions with new targets such as DNA binding sites or proteins may have evolved. Analysis of regulatory gene sequence evolution by identifying the selection pressures acting on genes, in particular positive selection, may therefore provide insights into the origins of morphological diversity.

Detecting adaptive molecular evolution in protein-coding genes usually involves the comparison of synonymous (silent, d_S) and non-synonymous (amino acid changing, d_N) substitution rates, which are the number of synonymous and non-synonymous nucleotide substitutions per site. The ratio of these two rates, $\omega = d_N/d_S$ measures selection pressure at the protein level (Goldman & Yang, 1994; Muse & Gaut, 1994). Under neutral selection, where selection has no effect on fitness, non-synonymous mutations will be fixed at the same as rate synonymous ones and $\omega = 1$. Under purifying selection, where non-synonymous substitutions are deleterious, $d_N < d_S$ and $\omega < 1$. Under positive, or directional selection, where non-synonymous substitutions are fixed at a higher rate than synonymous substitutions, $d_N > d_S$ and $\omega > 1$.

With increasingly sensitive methods of detection, using sequence evolution models in a maximum likelihood framework, numerous cases of directional evolution have now been identified in a variety of genes (see Yang & Bielawski (2000) for selected examples). However, evidence of positive selection in regulatory genes associated with morphological evolution has been mixed. In *Arabidopsis thaliana*, naturally occurring alleles of the MADS-box *CAULIFLOWER* gene appear to possess an excess of non-synonymous substitutions, and this variation is associated with effects on floral morphology (Purugganan & Suddith, 1998). In the Hawaiian silversword alliance, which has undergone rapid morphological diversification, directional selection was detected in homologues of the *Arabidopsis* floral regulatory genes *APETALA1* and *APETALA3* (Barrier *et al.*, 2001), but not in the coding region of putative growth regulator genes from the DELLA subfamily (Remington & Purugganan, 2002).

Several studies of molecular evolution have been carried out in members of the TCP gene family, with variable results. Analysis of d_N/d_S ratio in orthologues of the maize architecture gene *TEOSINTE BRANCHED 1* (*TB1*) in the morphologically diverse grass tribe Andropogoneae did not suggest instances of positive selection (Lukens & Doebley, 2001). Equally, no evidence of directional selection was observed in the duplicated genes *CYCLOIDEA* (*CYC*) and *DICHOTOMA* (*DICH*) in the Antirrhineae (Hileman *et al.*, 2003), which includes *Antirrhinum majus* for which these genes have been functionally characterised (Luo *et al.*, 1996; Luo *et al.*, 1999). By contrast, an extension of the work on legume *CYC* described here examining the molecular evolution of LEGCYC1A and LEGCYC1B in diverse *Lupinus* species suggested a correlation between morphological change and positive selection at certain codon sites in the LEGCYC1B locus (Ree *et al.*, 2004). Unlike the *TB1* and *CYC/DICH* studies, however, Ree *et al.* (2004) used a "branch-site" model that accounts for both lineage and site specific variation and has been found to be more sensitive in detecting signatures of positive selection than models that account for either lineage or site variation separately (Yang &

Neilsen, 2002). Previous models allowing the d_N/d_S ratio to vary among sites but not along lineages (Nielsen & Yang, 1998; Yang *et al.*, 2000), or among lineages but not across sites (Yang, 1998) have been found to sometimes lack power in detecting positive selection. Functional proteins may have strong structural constraints, and many amino acids sites may be largely invariable, with ω close to 0. If adaptive evolution affects only a few amino acids in certain lineages, for instance after gene duplication, then a "branch-site" model will be more powerful in detecting positive selection (Yang & Nielsen, 2002).

This study examines the molecular evolution of LEGCYC1A and LEGCYC1B paralogues in the genistoid clade *sensu* Wojciechowski (2003). This work will also establish to what extent the locus-specific LEGCYC primers, described in chapter 3, may be useful for phylogenetic analysis within this group. This large clade is defined from recent molecular phylogenetic studies and comprises ~ 1,300 species from seven different tribes, some of which were previously thought to be unrelated (Wojciechowski, 2003) (figure 5-1). Members of this clade have typical papilionoid flowers, with some notable exceptions, including *Cadia purpurea*. The expression pattern of LEGCYC1A and LEGCYC1B in *C. purpurea* was found to differ from that of another genistoid species with typical papilionoid flowers, *Lupinus nanus*, where these genes are expressed exclusively in the adaxial region of the developing flower (see chapter 4). In particular, LEGCYC1B in *C. purpurea* was found to be expressed homeotically in all five petals. To test whether the morphological shift from zygomorphy to actinomorphy, as occurred in the *Cadia* lineage characterised by bell-shaped radially symmetrical flowers and represented here by *C. purpurea*, is associated with episodes of directional selection in *CYC*-like genes, models of codon evolution were evaluated in phylogenies of LEGCYC1A and LEGCYC1B from members of the genistoid clade.

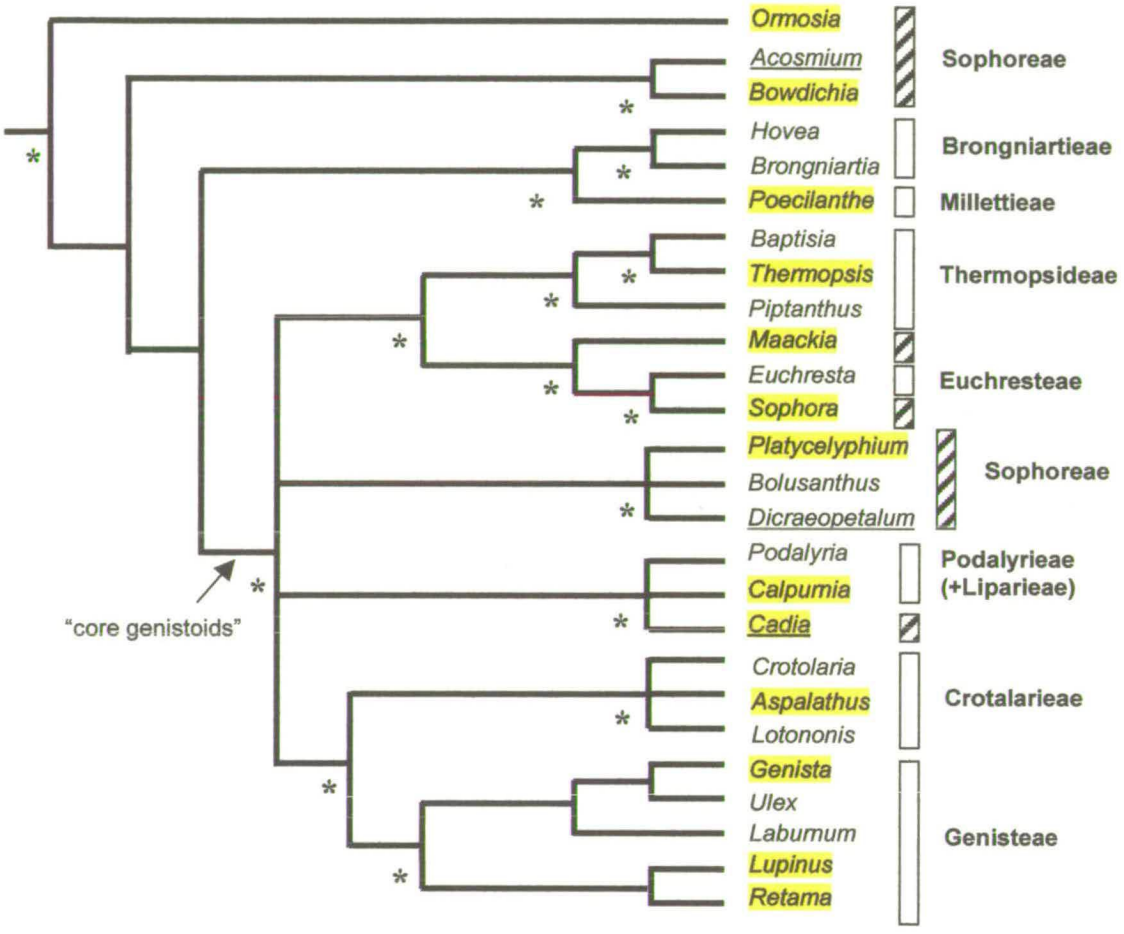


Figure 5-1. Summary of phylogenetic relationships within the genistoid clade (redrawn and modified from Wojciechowski, 2003), based on results from nrDNA ITS and *rbc*L (Crisp *et al.*, 2000; Kajita *et al.* 2001), and *trn*L intron (Pennington *et al.*, 2001) analyses. * denotes clades with bootstrap support greater than 50%, based mainly from Crisp *et al.*, (2000), and Pennington *et al.* (2001). Taxa highlighted in yellow were sampled for the LEGCYC sequence analyses. Taxa underlined have near-radially symmetrical flowers; their distribution suggests that radial symmetry evolved independently in the genistoid clade. Tribes are given on the right. The core genistoid clade is defined by Crisp *et al.* (2000) and Wojciechowski, 2003; a broader definition, with *Ormosia* as sister to all other genistoids, is given by Pennington *et al.* (2001).

5.2 MATERIALS AND METHOD

5.2.1 Taxon sampling

Taxa representing the range of the core genistoid clade and its sister group (*sensu* Wojciechowski, 2003; table 5-1 and figure 5-1) were sampled for isolation of orthologues of LEGCYC1A and LEGCYC1B. In particular, a putative sister taxon to *Cadia* based on recent phylogenetic evidence (Pennington *et al.*, 2001), *Calpurnia aurea*, which has typical zygomorphic papilionoid flowers, was included to detect changes at the sequence level in the actinomorphic branch. All taxa chosen here have typical papilionoid flowers, with the exception of *Acosmium subelegans*, which like *C. purpurea*, has near-radial flowers.

Genomic DNA for this study was extracted from fresh leaf material (*Crotalaria strigulosa, Maackia chinensis, Piptanthus nepalensis*, *Thermopsis villosa*) or floral material (*Retama monosperma*) following a small-scale 2X CTAB procedure modified from Doyle and Doyle (1987) (appendix 1A). Genomic DNA, from *Acosmium subelegans*, *Bowdichia vigilioides*, *Ormosia amazonica, Platycelyphium voense*, *Poecilanthes parviflora* and *Sophora velutina*, was provided by R.T. Pennington (RBGE). Genomic DNA from *Calpurnia aurea* (Aiton) Benth. was provided by M. Lavin (University of Montana), and from *Aspalathus carnosa* Bergius by D. Edwards (University of Reading).

Taxon	Source	Location
Acosmium subelegans (Mohl.) Yakovlev	S. Bridgewater 358	Mato Grosso do Sul, Brazil
Aspalathus carnosa Bergius	D. Edwards JAH 209	South Africa
Bowdichia vigilioides Kunth	R.T. Pennington 477	Goiás, Brazil
Calpurnia aurea (Aiton) Benth.	M. Lavin 6198	RBG Kew seed source
Crotalaria strigulosa Balf.f.	RBGE 1991 0080	Yemen
Lupinus angustifolius cv. Merrit	S. Barker	UWA, Perth
Maackia chinensis Takeda	RBGE 1966 0927	RBGE, cultivated material
Ormosia amazonica Ducke	R.T. Pennington 645	Napo, Ecuador
Piptanthus nepalensis (Hook.) D. Don	RBGE 1998 2708	RBGE, cultivated material
Platycelyphium voense (Eng.) Wild.	Kew 1953-10603	RBG Kew, cultivated material
Poecilanthe parviflora Bentham	Lima s.n.	Rio de Janeiro Botanic Garden, cultivated material
Retama monosperma (L.) Boiss	RBGE 1984 9032	Spain
Sophora velutina Lindl.	Kew 1983-3116	RBG Kew, cultivated material
Thermopsis villosa (Walter) Fernald &	RBGE 1955 0131	RBGE, cultivated material

Table 5-1. List of taxa from the core genistoid clade and sister group (*sensu* Wojciechowski, 2003) used to test the primers LEGCYC_F5-LEGCYC_R4/R3 and LEGCYC_iR4/iR3-LEGCYC_R8 specific to LEGCYC1A and LEGCYC1B (see appendix 2).

5.2.2. PCR and sequencing

Two fragments for each locus were amplified separately using specific primer pairs LEGCYC_F5 and LEGCYC_R4/R3, and LEGCYC_iR4/iR3 and LEGCYC_R8 (described in chapter 3 and appendix 2). These were designed from *C. purpurea* and *L. nanus* LEGCYC1A and LEGCYC1B sequences to amplify most of the ORF. PCR conditions were optimised to yield a single band, with an initial denaturation step at 94°C (3 minutes), followed by 30-35 cycles of: denaturation at 94°C (1 minute), annealing at 55°C (30 seconds) and extension at 72°C (30 seconds), then followed by a final extension step 72°C (5 minutes). PCR amplifications were carried out using Bioline *Taq* and reagents (Bioline, London NW2, UK), in a 50µl reaction mix containing sterile distilled water, polymerase buffer, $MgCl_2$ (2.5mM), dNTPs (20µM), primers (0.5µM each), *Taq* polymerase (1 unit), and 20 – 30 ng genomic DNA. PCR products were visualised on a 1% agarose gel. Some primer combinations failed to amplify the expected PCR product. In other cases multiple bands were amplified, so the appropriate fragment was either gel

extracted or cloned into pCR4-TOPO (Invitrogen Ltd, Paisley, UK). However, after purification using Qiaquick kits (Qiagen Ltd, Dorking, Surrey, UK), most PCR products were sequenced directly. Dye-terminator cycle sequencing was carried out using Thermosequenase II (Amersham Pharmacia, Buckinghamshire, UK). Samples were analysed on an ABI model 377 Prism Automatic DNA sequencer.

5.2.3 Sequence alignment and phylogenetic analyses

LEGCYC1A and LEGCYC1B sequences available prior to this study were included in the matrices: from *Cadia purpurea*, *Lupinus nanus* (chapter 3), *Lupinus densiflorus* (LEGCYC1A: AY338914, LEGCYC1B: AY338865), *Lupinus digitatus* (LEGCYC1A: AY338922, LEGCYC1B: AY338873), *Genista tenera* (LEGCYC1A: AY338924, LEGCYC1B: AY33875) (Ree *et al.*, 2004) and *Anarthrophyllum* sp. (LEGCYC1B, R.H. Ree pers. comm.). *Lupinus densiflorus* sequences were of particular interest as results from Ree *et al.* (2004) suggested instances of positive selection at some codon sites in the LEGCYC1B locus in this species, which has proportionally smaller standard petals and larger wing petals than other *Lupinus* species. Amino acid and nucleotide sequences were aligned manually. The intron region was excluded from all analyses.

Phylogenetic analyses of the separate LEGCYC1A and LEGCYC1B matrices were carried out with PAUP* 4.0b10 (Swofford, 2001) using both the maximum likelihood (ML) and parsimony (MP) methods. To find all shortest trees and identify tree islands, heuristic maximum parsimony searches with 1,000 random addition replicates and tree bisection reconnection (TBR) branch swapping were conducted. Bootstrap support for nodes was estimated using the parsimony criterion with 1,000 bootstrap replicates. Models of sequence evolution were selected using the Akaike Information Criterion with Modeltest v3.06 (Posada and Crandall, 1998). For LEGCYC1A the K81uf + G model was selected. In this model, base frequencies were estimated

empirically (Lset Base = 0.3558 0.2362 0.2106) and among-rate variation followed a gamma distribution (α = 1.6533). Substitution rates were assumed equal for transitions and for two types of transversions (A↔G = C↔T, A↔T = G↔C) (Rmat = 1.0000 1.8542 0.6719 0.6719 1.8542). For LEGCYC1B, the parameter-rich GTR + G model was selected. Base frequencies (Lset Base = 0.3544 0.2101 0.1852) and substitution rates (Rmat = 0.9273 1.6973 0.6048 0.9976 2.2438) were estimated empirically. Among-rate variation followed a gamma distribution (α = 0.5556). Neither model allowed for a site class to be invariable (Pinvar = 0). Heuristic searches under the ML optimality criterion were conducted using TBR branch swapping algorithm.

5.2.4 Analyses of LEGCYC coding sequence evolution

Estimation of substitution rates using a likelihood approach is the most powerful method of investigating adaptative molecular evolution (Yang, 1998; Yang & Bielawski, 2000). The likelihood method relies on explicit models of sequence evolution, such as taking into account transition/transversion rate bias and non-uniform codon usage. Furthermore, likelihood ratio tests allow for nested models to be tested statistically (Yang, 1998). Models of codon evolution and tests for selection on LEGCYC paralogues were evaluated on phylogenies generated by the MP analyses, using codeml from the PAML (Phylogenetic Analyses using Maximum Likelihood) package version 3.13 (Yang, 1997). Unrooted ML phylogenies (with a reduced sample for LEGCYC1B, see section 5.3.2) of each locus were used for the initial branch lengths estimates under the one-ratio model (M0). Regions with gaps were removed from the matrices as PAML does not have any methods for dealing with them (Yang, 1997).

Ten models of codon evolution (described in Nielsen & Yang, 1998; Yang *et al.*, 2000; Yang & Nielsen, 2002) were evaluated for each data set. Nested models were compared by the standard likelihood ratio test (LRT: twice the log-likelihood difference between two models

$2\Delta L$) against the χ^2 distribution with degrees of freedom equal to the difference in number of parameters. The one-ratio model (M0; Goldman & Yang, 1994) is the simplest model and assumes a single ω for all sites and branches of the phylogeny. Models M1- M3, M7, M8 (Nielsen & Yang, 1998; Yang *et al.*, 2000) are site models where ω varies among sites but is constant across the phylogeny. The "neutral" model M1 assumes two categories of sites in a gene: one category is neutral ($\omega_1 = 1$) whereas the other is conserved and non-synonymous substitutions are eliminated by selection ($\omega_0 = 0$). The "selection" model M2 is an extension of M1 with the addition of an ω_2 site class that can take any value. The "discrete" model M3 is an extension of M0, and allows for a set number K of site classes to be unconstrained. M7 and M8 (Yang *et al.*, 2000) describe ω variation according to a beta distribution (with parameters p and q). In M7, ω is constrained between 0 and 1. M8 is an extension of M7 by allowing a proportion of sites to have $\omega > 1$.

The branch models allow ω to vary among lineages. In the models evaluated here, the phylogeny is partitioned into "foreground" and "background" branches, which are allowed to have different ω values. Whereas the two-ratio model does not allow ω to vary along sites, the branch-site models (models A and B; Yang & Nielsen, 2002) assume two site classes, ω_0 and ω_1, on the background branch, with an additional site ω_2 on the foreground branch. Model A constrains $\omega_0 = 0$ and $\omega_1 = 1$ and is thus a branch-specific extension of M2, whereas model B places no constraint on the values of ω_0 and ω_1 and can therefore be compared to M3 ($K = 2$). In this study, each analysis was repeated with a different foreground branch, in order to obtain a separate ω_2 value for that branch. Although results from multiple tests using the same data may not be evaluated statistically (Yang, 1998), the foreground-specific ω values are considered here as descriptive of each branch.

5.2.5 Analysis of Lupinus nanus LEGCYC1A*

As discussed in chapter 3, a novel sequence LEGCYC1A* similar to LEGCYC1A (79.72% nucleotide sequence similarity) was isolated in *Lupinus nanus*. To place this sequence in a phylogenetic context, LEGCYC1A* was included in the LEGCYC1A matrix. A maximum parsimony analysis, with bootstrap support, was carried out as described above.

5.3 RESULTS

5.3.1 Range of LEGCYC primers

Primer pairs designed on *C. purpurea* and *L. nanus* sequences were found to amplify a single product in most genistoid taxa selected here (table 5-2). Primers specific for LEGCYC1A were found to work in fewer taxa than those for LEGCYC1B, which could reflect the faster rate of evolution of this locus (discussed in chapter 3).

Taxon	Primer combinations			
	LEGCYC1B		LEGCYC1A	
	F5-R3	iR3-R8	F5-R4	iR4-R8
Acosmium subelegans	Ø	Ø	√	√
Aspalathus sp.	√	Ø	√	√
Bowdichia vigilioides	Ø	Ø	√	√
Calpurnia aurea (Aiton) Benth.	√ mul	√	√	√
Crotalaria strigulosa	√	√	Ø	√
Dicraeopetalum stipulare	√	√ mul	Ø	Ø
Lupinus angustifolius cv. Merrit	√	√	√	√
Maackia chinensis	√ mul	√	√	√
Ormosia amazonica	√	√	Ø	Ø
Piptanthus nepalensis	√	√	Ø	Ø
Platycelyphium voense	√	√	Ø	Ø
Poecilanthe parviflora	√	√	Ø	√ mul
Retama monosperma	√	√	√	Ø
Sophora velutina	√	√	Ø	√
Thermopsis villosa	√	√	Ø	√

Table 5-2. Amplification results using primer combinations specific to LEGCYC1A (LEGCYC_R4/iR4) and LEGCYC1B (LEGCYC_R3/iR3) in a range of genistoid taxa. √ = amplification of a single band of the expected size, √ mul = amplification of multiple bands, Ø = no amplification.

5.3.2 Phylogenetic analyses of LEGCYC paralogues in the genistoid clade

5.3.2a Sequence data

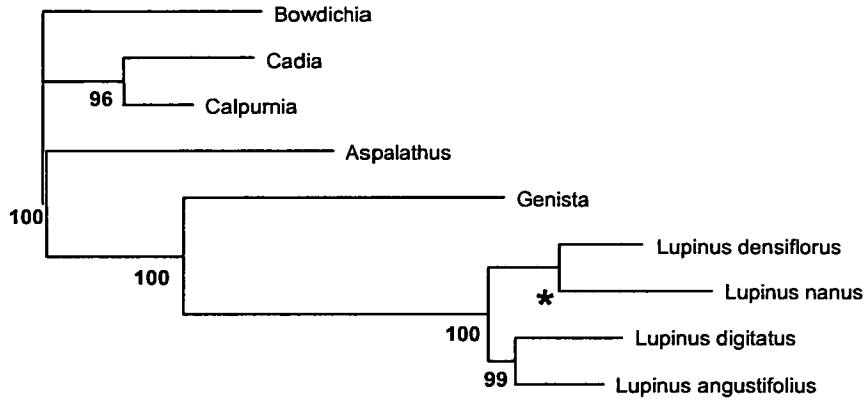
Two separate nucleotide matrices for LEGCYC1A and LEGCYC1B were compiled with 9 and 15 sequences respectively. At this taxonomic level, although sequences from the selected genistoid taxa were variable in length as well as in nucleotide sequence, putative LEGCYC1A and LEGCYC1B orthologues were easily alignable across the partial ORF. Alignments are given in appendix 7. Sequence lengths, excluding the intron, of putative LEGCYC1A and LEGCYC1B orthologues in selected genistoid taxa ranged from 937 bp (*C. purpurea*) to 988 bp (*G. tenera, L. digitatus*) for LEGCYC1A and 1,044 bp (*S. velutina*) to 1,143 bp (*R. monosperma*) for LEGCYC1B. Alignment of the eight sequences in the LEGCYC1A matrix was 1,028 characters in length, and required the insertion of 33 gaps between 3 and 33 bp. By comparison, alignment of the 15 sequences in the LEGCYC1B matrix was 1,308 characters in length, and required approximately four times the number of gaps (116 gaps between 3 and 66 bp) as the LEGCYC1A matrix. Although this may be accounted for by the greater number of sequences in the LEGCYC1B matrix, representing a wider range of taxa, this pattern is also in agreement with the pairwise comparison of LEGCYC1A and LEGCYC1B in *C. purpurea* and *L. nanus* (chapter 3), as well as between *Lupinus* species (Ree *et al.*, 2004). In contrast, pairwise similarity was higher between LEGCYC1B sequences (mean identity 90.14% at the nucleotide level, with a range of 96.61% - 84.57%) than between LEGCYC1A sequences (mean identity of 86.98%, with a range of 79.99% - 94.92%).

A number of indels were microsatellite-like repeats of codons, not only between sequences as described in chapter 3, but also within individuals. Allelic length variation was observed in LEGCYC1B *Retama monosperma* (CAA, glutamine) at nucleotide position 850. Allelic microsatellite regions were also observed in LEGCYC1B in *Lupinus* species (Ree *et al.*, 2004).

5.3.2b Phylogenetic analyses

Phylogenetic analyses of the LEGCYC1A and LEGCYC1B nucleotide matrices broadly recovered the species phylogeny based on current studies (Pennington *et al.*, 2001; Wojciechowski, 2003; figure 5-1). For both loci, sequences from members of the Genisteae (*Lupinus*, *Retama*, *Genista, Anarthrophyllum*) were recovered in a monophyletic clade, the sister relationship of *Cadia* and *Calpurnia* was also recovered, and sequences from the basal-most species (*Ormosia* and *Bowdichia*) were not found to be nested within more derived clades *e.g.* Genisteae. Parsimony analyses of nucleotide sequences resulted for LEGCYC1A in two most parsimonious trees of 447 steps (CI = 0.859, RI = 0.795), and for LEGCYC1B in two most parsimonious trees of 658 steps (CI = 0.781, RI = 0.711) (figure 5-2). Trees were rooted on the sequence from the basal-most species (*Bowdichia* for LEGCYC1A and *Ormosia* for LEGCYC1B) based on recent species phylogenies (Wojciechowski, 2003). The topology of the single ML trees for both data sets were identical to the MP trees shown here, with the exception of the position of the *Platycelyphium* branch which is nested between the *Cadia/Calpurnia* clade and the *Maackia* branch in the LEGCYC1B ML tree (figure 5-2). To simplify the PAML analysis, three LEGCYC1B sequences were removed from the data matrix (*Platycelyphium*, *Poecilanthe* and *Anarthrophyllum*), without any effect on topology based on parsimony analysis.

LEGCYC1A



LEGCYC1B

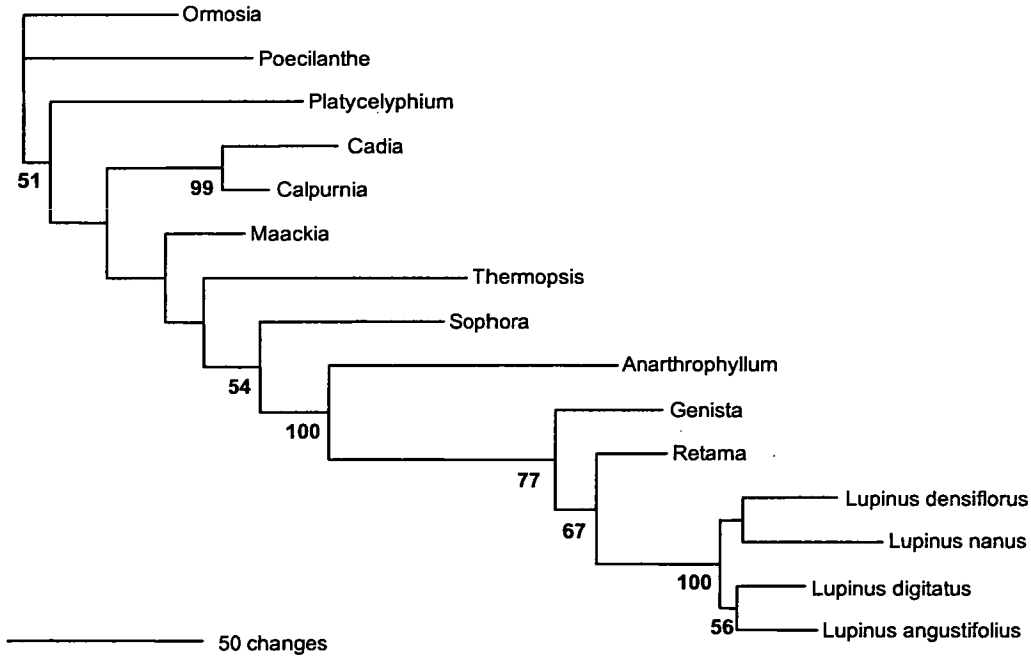


Figure 5-2. One of the two most parsimonious trees of LEGCYC1A nucleotide matrix (447 steps, CI = 0.859, RI = 0.795) rooted on *Bowdichia*, and of LEGCYC1B nucleotide matrix (658 steps, CI = 0.781, RI = 0.711) rooted on *Ormosia*, with bootstrap support shown in bold. * marks branches which collapse in the strict consensus tree.

5.3.3 Testing for positive selection

5.3.3a Site models

Parameter estimates for each of the site models investigated are summarised in table 5-3. None of the site models allowing for ω to be estimated across the entire phylogenies detected sites under positive selection, but some models were better than others at describing the data sets. For instance, allowing two site classes with unconstrained values (M3, $K = 2$) provided a significantly better fit to both LEGCYC1A and LEGCYC1B data sets than having a single unconstrained value for all sites (M0) (LEGCYC1A: $2\Delta L = 12.828$, df = 2, P = 0.0016, LEGCYC1B: $2\Delta L = 50.686$, df = 2, $P < 0.001$). This suggests that the selective constraint on sites in both copies is not homogeneous. Addition of a third site class (M3, $K = 3$) resulted in a similat likelihood to having only two estimated site classes (M3, $K = 2$) for either locus. In addition, the third estimated ω was in both cases less than 1 (LEGCYC1A: $\omega_2 = 0.60449$, LEGCYC1B: $\omega_2 = 0.19822$), suggesting that this additional site class, like the other two, was under intermediate purifying selection.

Comparison of the neutral model M1, which has two constrained site classes ($\omega_0 = 0$, $\omega_1 = 1$), and the selection model M2, which has an additional unconstrained site class ω_2, showed that the selection model fitted both data sets significantly better (LEGCYC1A: $2\Delta L = 44.183$, df = 2, $P < 0.001$, LEGCYC1B: $2\Delta L = 32.339$, df = 2, $P < 0.001$). This implies that across the entire tree, a large proportion of sites (LEGCYC1A: 69.6%, LEGCYC1B: 32%) are not evolving under strictly neutral or purifying selection, but somewhere in between.

Even with a continuous distribution of ω (M7 and M8), the additional unconstrained ω value estimated in M8 was less than 1 (LEGCYC1A: $\omega = 0.5689$, LEGCYC1B: $\omega = 0.68718$), and provided no significant improvement in either case over M7.

Model	p	LEGCYC1B			LEGCYC1A		
		lnL	Estimates of parameters	Positively selected sites	lnL	Estimates of parameters	Positively selected sites
M0: one ratio	1	-2641.747	ω= 0.2036	none	-2707.984	ω= 0.2536	none
Site-specific models							
M1: neutral	1	-2633.612	p_0 = 0.59859 (p_1 = 0.40141)	N/A	-2723.733	p_0 = 0.45649 (p_1 = 0.54351)	N/A
M2: selection	3	-2617.442	p_0 = 0.47843 p_1 = 0.0785 (p_2 = 0.44342); ω_2 = 0.31991	none	2701.642	p_0 = 0.25382 p_1 = 0.05041 (p_2 = 0.69576); ω_2 = 0.31068	none
M3: discrete (*K*=2)	3	-2617.404	p_0 = 0.67336 (p_1 = 0.32664); ω_0 = 0.04111 ω_1 = 0.57167	none	-2701.576	p_0 = 0.57674 (p_1 = 0.42326); ω_0 = 0.09355 ω_1 = 0.49595	none
M3: discrete (*K*=3)	5	-2617.184	p_0 = 0.43082 p_1 = 0.37096 (p_2 = 0.19822) ω_0 = 0.00000 ω_1 = 0.20751 ω_2 = 0.69694	none	-2701.542	p_0 = 0.19135 p_1 = 0.57705 (p_2 = 0.2316) ω_0 = 0.00001 ω_1 = 0.21554 ω_2 = 0.60449	none
M7: beta	2	-2618.428	p = 0.08589 q = 0.26704	N/A	-2701.606	p = 0.65452 q = 1.62236	N/A
M8: beta & ω	4	-2617.207	p_0 = 0.78757 p = 0.0857 q = 0.40712 (p_1 = 0.21243) ω = 0.68718	none	-2701.558	p_0 = 0.73698 p = 0.91882 q = 4.34242 (p_1 = 0.26302) ω = 0.5689	none

Table 5-3. Parameter estimates for LEGCYC1A and LEGCYC1B under site models. p is the number of free parameters for ω. lnL is the log likelihood of each model. p_n describes the proportion of sites having ω_n. For M7 and M8, p and q describe the beta distribution of ω values. None of these models detected sites under positive selection across the entire phylogeny in either locus.

5.3.3b Branch models

Results of the branch models are summarised in tables 5-4 and 5-5 for selected foreground branches that have ω greater than one for at least one of the LEGCYC copy. Values for all branches estimated with branch-site model B are shown in figure 5-3. The location of positively selected sites along these lineages is shown in figure 5-4.

The two-ratio model, where a single ω is estimated for the background and foreground branches, did not detect evidence of positive selection on any branch of the LEGCYC1B phylogeny. However, for LEGCYC1A, the ω value for the *Lupinus digitatus – L. angustifolius* foreground branch was found to be greater than 1 ($\omega = 3.5332$).

Unlike the two-ratio model, the branch-site models allow for sites to be partitioned into classes along the sequence, as well as allowing an additional ω parameter (ω_2) for the foreground branch. Branch-site model B is less constrained than model A, by estimating the two ω parameters for the background branches rather than fixing them at $\omega_0 = 0$ and $\omega_1 = 1$. It may therefore provide a better fit to the data (Yang & Nielsen, 2002). The foreground ω (ω_2) estimated under model B is shown for each branch in figure 5-3. Much variation in ω_2 was observed between lineages for both LEGCYC1A and LEGCYC1B, suggesting that selection pressures may not be acting uniformly across the trees. In the LEGCYC1B phylogeny, results suggest that positive selection may have acted on this gene along the *Cadia* branch (see figure 5-3). Both models A and B estimated a high foreground ω (model A: $\omega_2 = 17.91908$, model B: $\omega_2 = 19.65467$) for the *Cadia* branch. However, only three amino acids were identified under model A, of which two were identified under model B, with a posterior probability (*P*) greater than 0.5 of being positively selected (see table 5-4). None of these were in the conserved TCP domain (figure 5-4). By contrast, neither the branch of the sister taxon to *Cadia*, *Calpurnia*, or the branch of the common ancestor of *Cadia* and *Calpurnia*, have evidence of positive selection

with ω_2 close to zero for both branches model B (see figure 5-3). All other branches of the LEGCYC1B phylogeny have low ω_2 under model B (figure 5-3), with the exception of *Sophora*, where $\omega_2 = 1.0736$, with 17.3% of sites appear to be evolving under neutral selection (see table 5-5).

By contrast, indication of positive selection was found on a number of branches for LEGCYC1A. Both branch-site models, along with the two-ratio model, estimated high ω_2 for the *L. digitatus – L. angustifolius* lineage (model A: $\omega_2 = 19.4458$, model B: $\omega_2 = 10.9447$). A relatively high percentage of sites were estimated to be in the ω_2 site class (23.5%, under model B), and of particular interest one codon (tyrosine; $P = 0.89$ under model B) in the basic region of TCP domain was identified as having evolved under positive selection (figure 5-4). High ω_2 values were also obtained under model B, but not model A, for the *L. nanus* ($\omega_2 = 21.52457$), *Cadia* ($\omega_2 = 3.10706$), and *Bowdichia* ($\omega_2 = 3.43026$) branches (see figure 5-3, tables 5-4 and 5-5). The proportion of sites in this class along the foreground branch was low, particularly for *L. nanus* (0.48% under model B). No positively selected sites were identified with a posterior probability greater than 0.5 along the *Cadia* branch (table 5-4). However, along the *Bowdichia* branch, one of the positively selected sites (glycine; $P = 0.54$) was found in the loop region of the TCP domain (figure 5-4).

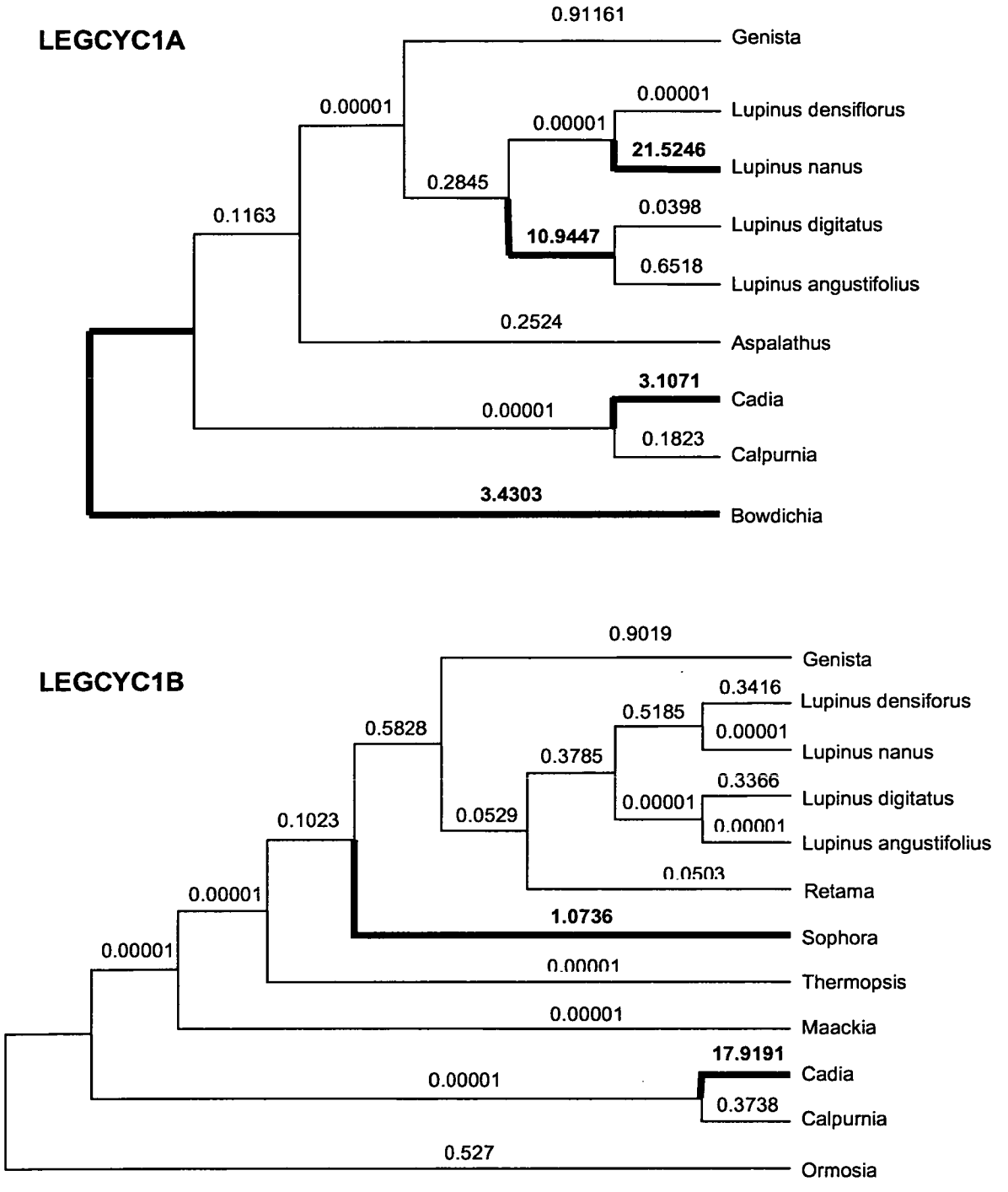


Figure 5-3. Cladograms of LEGCYC1A and LEGCYC1B showing the foreground ω_2 value obtained under model B for each branch. Branches with ω_2 values greater than one, indicative of positive selection on some sites on that particular lineage, are in bold. For LEGCYC1B, only *Cadia* has an ω_2 value much greater than 1, whereas for LEGCYC1A, these are scattered across the phylogeny.

Branch/ Model	p	LEGCYC1A lnL	LEGCYC1A Estimates of parameters	LEGCYC1A Positively selected sites	LEGCYC1B lnL	LEGCYC1B Estimates of parameters	LEGCYC1B Positively selected sites
foreground:							
Cadia							
2-ratio	2	-2707.957	ω_0 = 0.2515 ω_1 = 0.2819	N/A	-2642.618	ω_0 = 0.1625 ω_1 = 0.2071	N/A
Model A	3	-2719.576	p_0 = 0.05862 p_1 = 0.06284 (p_2 = 0.87855); ω_2 = 0.26315	none	-2631.276	p_0 = 0.60062 p_1 = 0.38089 (p_2 = 0.0185); ω_2 = **19.65467**	2L (*P*=0.95), 158C (*P*=0.64), 203N (*P*=0.52)
Model B	5	-2701.323	p_0 = 0.52923 p_1 = 0.43620 (p_2 = 0.03457); ω_0 = 0.08028 ω_1 = 0.47555 ω_2 = **3.10706**	none	-2614.886	p_0 = 0.65729 p_1 = 0.32863 (p_2 = 0.01407); ω_0 = 0.03631 ω_1 = 0.54605 ω_2 = **17.91908**	2L (*P*=0.96), 158C (*P*=0.73)
L.nanus							
2-ratio	2	-2707.499	ω_0 = 0.2631 ω_1 = 0.1758	N/A	-2640.258	ω_0 = 0.2200 ω_1 = 0.0801	N/A
Model A	3	-2719.423	p_0 = 0.15709 p_1 = 0.19488 (p_2 = 0.64802); ω_2 = 0.00001	none	-2626.851	p_0 = 0.10015 p_1 = 0.0672 (p_2 = 0.83265); ω_2 = 0.03377	none
Model B	5	-2701.1	p_0 = 0.59418 p_1 = 0.40095 (p_2 = 0.00487); ω_0 = 0.09863 ω_1 = 0.50318 ω_2 = **21.52457**	177Q (*P*=0.71)	-2615.001	p_0 = 0.26404 p_1 = 0.13056 (p_2 = 0.60539); ω_0 = 0.04295 ω_1 = 0.61210 ω_2 = 0.00001	none
L.digitatus/ L.angustifolius							
2-ratio	2	-2706.172	ω_0 = 0.2438 ω_1 = **3.5332**	N/A	-2642.345	ω_0 = 0.2048 ω_1 = 0.0001	N/A
Model A	3	-2722.367	p_0 = 0.43654 p_1 = 0.49410 (p_2 = 0.06935); ω_2 = **19.44588**	13I (*P*=0.84),28A (*P*=0.51), 49L (*P*=0.86), 63H (*P*=0.81), 163G (*P*=0.53)	-2632.9	p_0 = 0.00000 p_1 = 0.00000 (p_2 = 1.00000); ω_2 = 0.00001	none
Model B	5	-2699.918	p_0 = 0.46112 p_1 = 0.30403 (p_2 = 0.23486) ω_0 = 0.09519 ω_1 = 0.49631 ω_2 = **10.94474**	13I (*P*=0.93), 22S (*P*=0.69), 28A (*P*=0.89), 49L (*P*=0.94), 63H (*P*=0.93), 163G (*P*=0.88)	-2616.997	p_0 = 0.00000 p_1 = 0.00000 (p_2 = 1.00000); ω_0 = 0.04131 ω_1 = 0.57508 ω_2 = 0.00001	none

Table 5-4. Parameter estimates from the 2-ratio and branch-site models for selected LEGCYC1A and LEGCYC1B foreground branches where $\omega > 1$ under one of these models. p is the number of free parameters for ω. lnL is the log likelihood of each model. p_n describes the proportion of sites having ω_n. For the two-ratio model, ω_0 is the background estimate and ω_1 the foreground estimate. In the branch-site models, ω_2 is the additional parameter for a site class in the foreground branch and p_2 the proportion of sites in this class. For LEGCYC1B, only the *Cadia* branch was found to have a higher non-synonymous rate, whereas for LEGCYC1A more branches showed a signature of positive selection (also table 5-5). The location of positively selected sites (with a posterior probability $P > 0.5$) is shown in figure 5-4.

Branch/ Locus	Model	*p*	lnL	Estimates of parameters	Positively selected sites
Sophora	2-ratio	2	-2640.608	$\omega_0 = 0.1892$ $\omega_1 = 0.4903$	N/A
LEGCYC1B	Model A	3	-2630.227	$p_0 = 0.38885$ $p_1 = 0.23124$ ($p_2 = 0.3799$); $\omega_2 = 0.68908$	none
	Model B	5	-2615.899	$p_0 = 0.55151$ $p_1 = 0.27524$ ($p_2 = 0.07325$) $\omega_0 = 0.03075$ $\omega_1 = 0.55573$ $\omega_2 =$ **1.07360**	19E (*P*=0.71), 30P (*P*=0.64), 38H (*P*=0.71), 44L (*P*=0.66), 115E (*P*=0.7), 129V (*P*=0.69), 227G (*P*=0.68)
Bowdichia	2-ratio	2	-2705.788	$\omega_0 = 0.2312$ $\omega_1 = 0.4955$	N/A
LEGCYC1A	Model A	3	-2720.188	$p_0 = 0.23986$ $p_1 = 0.24181$ ($p_2 = 0.51833$); $\omega_2 = 0.54094$	none
	Model B	5	-2698.623	$p_0 = 0.54633$ $p_1 = 0.37120$ ($p_2 = 0.08248$) $\omega_0 = 0.08838$ $\omega_1 = 0.47220$ $\omega_2 =$ **3.43026**	4S (*P*=0.8), 95A (*P*=0.54), 224E (*P*=0.92), 231L (*P*=0.9), 252M (*P*=0.52)

Table 5-5. Parameter estimates for *Sophora* LEGCYC1B and *Bowdichia* LEGCYC1A from the two-ratio and branch-site models. Both branches have ω_2 greater than 1 under the model B, although the d_N/d_S is close to 1 for the *Sophora* branch suggesting a proportion of sites are evolving neutrally. p is the number of free parameters for ω. lnL is the log likelihood of each model. p_n describes the proportion of sites having ω_n. For the two-ratio model, ω_0 is the background estimate and ω_1 the foreground estimate. In the branch-site models, ω_2 is the additional parameter for a site class in the foreground branch and p_2 the proportion of sites in this class. Position and codon translation of sites identified in the ω_2 site class are given, along with their posterior probability (*P*). The location of positively selected sites (with a posterior probability $P > 0.5$) is shown for the *Bowdichia* branch in figure 5-4.

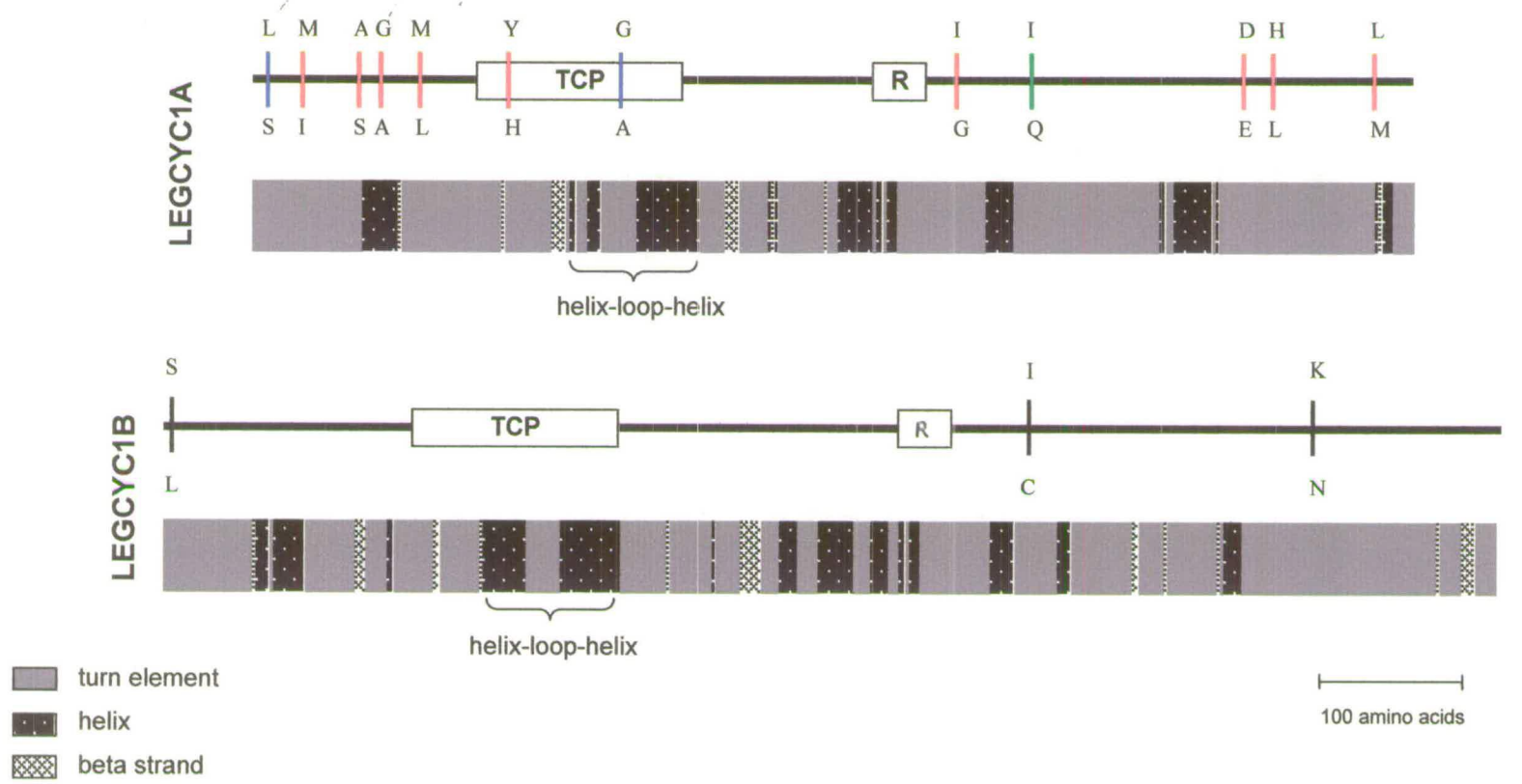


Figure 5-4. Location of the inferred non-synonymous mutations (with a posterior probability greater than 0.5 under model A or B) along the partial LEGCYC coding region, using *Genista tenera* sequences as reference. The predicted secondary structure (NNPREDICT ; Kneller *et al.*, 1990) is given for each locus, with helix and beta-strands regions shown., and the helix-loop-helix region of the TCP domain highlighted. Ancestral and derived amino acids are shown below and above the line respectively. For LEGCYC1B, derived amino acids are shown for the *Cadia purpurea* sequence. For LEGCYC1A, derived amino acids are shown for the *Lupinus digitatus*/*L. angustifolius* branch (red), *Bowdichia vigilioides* (blue) and *L. nanus* (green). One mutation was inferred in the TCP domain for *B. vigilioides* and one for the *L. digitatus*/*L. angustifolius* lineage.

5.3.4 Phylogenetic position of LEGCYC1A*

Parsimony analysis of the LEGCYC1A data set with the inclusion of *L. nanus* LEGCYC1A* (118 parsimony informative characters out of 724) resulted in two most parsimonious trees of 383 steps (CI = 0.830, RI = 0.733) (figure 5-5). Only one branch, related to LEGCYC1A*, collapsed in the strict consensus tree (figure 5-5). The position of the *L. nanus* LEGCYC1A* branch does not indicate that this copy is the product of a duplication specific to *L. nanus*, and suggests this copy may be found in other taxa. It also puts into question the initial orthology assessment of LEGCYC1A sequences.

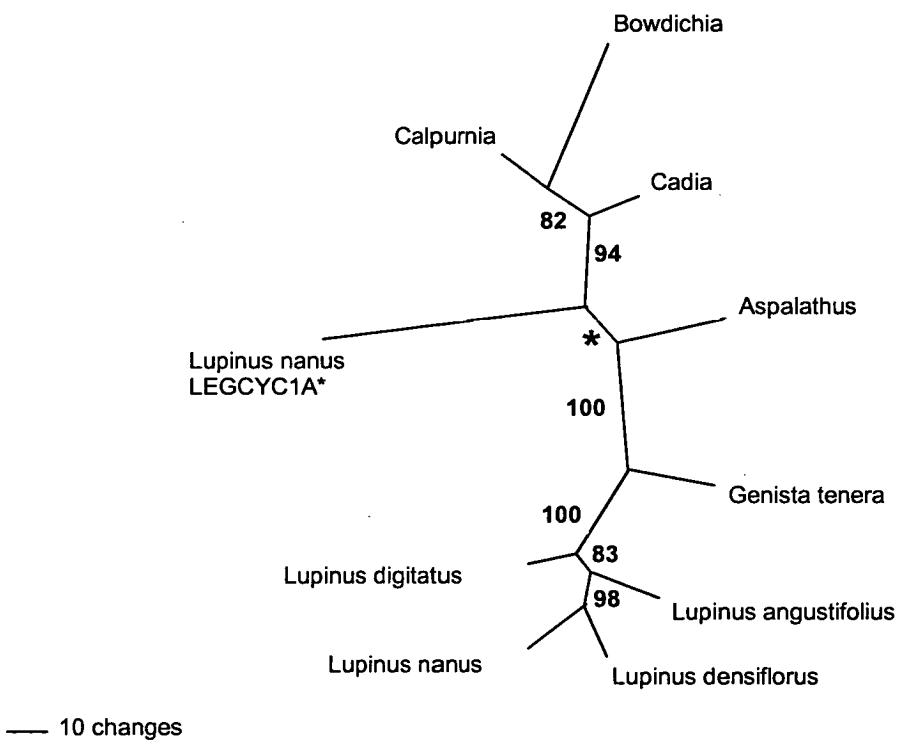


Figure 5-5. Unrooted phylogram of one most parsimonious tree out of two MP trees of 383 steps (CI = 830, RI = 733) of sequences amplified by LEGCYC1A specific-primers (LEGCYC_iR4/R4) and *L. nanus* LEGCYC1A*. The branch marked with * collapsed in the strict consensus tree.

5.4 DISCUSSION

5.4.1 Phylogenetic potential of LEGCYC genes in the genistoid clade

The locus-specific primers initially designed for *Cadia purpurea* and *Lupinus nanus* of the genistoid clade work in wide range of species within this group including taxa in the sister group of the core clade (*sensu* Wojciechowski, 2003, figure 5-1). However, the LEGCYC1B primers appear to work in a larger number of species than the LEGCYC1A primers, possibly reflecting the faster nucleotide substitution rate of LEGCYC1A.

These LEGCYC genes are potential sources of phylogenetic information. Although the internal transcribed spacers (ITS) of the nuclear 18S-26S ribosomal DNA gene family are the most commonly used nuclear regions for phylogenetic analysis at low taxonomic levels (Hershkovitz *et al.*, 1999), there is a need in systematic studies for other rapidly evolving low copy nuclear genes, particularly those that potentially underlie morphological variation (Doyle & Doyle, 1999). One reason is that multiple sources of informative molecular data are required for testing the congruence of topologies of different gene trees, in order to have more reliable estimates of taxic relationships, or to investigate hybridisation events (Doyle, 1992). Other reasons are related to the nature of the ITS region itself. The ITS region is part of a multigene family that is homogenised through concerted evolution. Reports of incomplete concerted evolution, or pseudogene evolution in this gene family suggest that sequencing of ITS may be subject to complicating factors (Doyle & Doyle, 1999). In addition, ITS divergence between closely related taxa may be too low to resolve relationships, in part due to the short length of the ITS region (*ca.* 450 bp), and to the homogenising effect of concerted evolution (Hershkovitz *et al.*, 1999). Molecular data from single copy nuclear genes providing more variable characters are therefore needed to resolve rapid radiations at the species level. The considerably higher rate of evolution of the two *CYC*-like loci described here compared to ITS at the species level (two to

four times greater than ITS, see chapter 3 and Ree *et al.* (2004)) suggests these would be a useful source of phylogenetic characters for species that have undergone rapid diversification.

Despite considerable work on members of the genistoid clade, relationships between certain genera or between species which have undergone rapid diversification are still unclear. As these primers appear to work in a relatively wide taxonomic range, they may be useful for molecular systematic studies in this clade, which contains many large genera such as *Crotalaria* (*ca.* 600 species), *Aspalathus* (*ca.* 250 species), *Genista* (*ca.* 90 species) and economically important ones such as *Lupinus* (*ca.* 250 species), *Sophora*, and gorse (*Ulex*). For instance, Ree *et al.* (2004) have found that these two LEGCYC copies provided greater phylogenetic information between recently diverged North American *Lupinus* species than ITS.

5.4.2 Selection pressures across LEGCYC paralogues

Codon-based models of sequence evolution suggest that both LEGCYC paralogues are under variable selection pressures across the sites and lineages examined. The average d_N/d_S over all sites was 0.25 for LEGCYC1A and 0.2 for LEGCYC1B, which are typical values for functional proteins where most amino acids are under strong constraints (Sharp, 1997). The majority of sites across both phylogenies appears to be under strong purifying selection, and around 30 to 40% of sites are under more relaxed purifying selection (*e.g.* $\omega > 0.5$, estimated from model 3 ($K = 2$), table 5-3). In some lineages, a small number of sites were found to be under positive selection, as detected by the branch-site models. This combination of selection pressures has been termed "selectional mosaic" by Ree *et al.* (2004), and reflects the heterogeneous and rapid evolution of LEGCYC genes.

Although specific differences in ω between lineages cannot be evaluated statistically, the variation in d_N/d_S between lineages estimated by branch-site models can be informative. For

LEGCYC1B, a higher rate of non-synonymous evolution was detected only on the *Cadia* branch and may be associated with a morphological shift from zygomorphy to radial symmetry. No evidence of positive selection was found on the *Lupinus densiflorus* branch. This result differs from that of Ree *et al.* (2004) where positive selection was detected for a small proportion of sites in this lineage characterised by unusual, small, wing-dominated flowers. Some of these sites were excluded here (in a region where a gap was required for alignment) although nearly half of the sites identified by Ree *et al.* (2004) were included in this analysis. It may be that because the sequences in this matrix are more divergent compared to a matrix of *Lupinus* sequences, the small number of sites putatively under positive selection may have been swamped out by the higher rate of non-synonymous substitutions across the whole tree.

Unlike the Ree *et al.* (2004) study, the branch site models also detected episodes of positive selection in LEGCYC1A along certain branches, including *Lupinus* lineages. These particular branches, however, were not tested for positive selection by Ree *et al.* (2004). For this locus, there is no obvious correlation between floral morphology and molecular evolution, although the *Cadia* branch does have a lineage-specific site class with d_N/d_S greater than 1.

The *Bowdichia* lineage is among those with sites that have a relatively higher non-synonymous rate. It would be worth sequencing LEGCYC1A from a sister taxon to *Bowdichia*, *Acosmium*, which has near radial flowers, to see if changes at these sites is shared by their common ancestor.

A high d_N/d_S was also detected for the *L. nanus* branch under model B, although only a small number of sites (0.48%) were estimated in that category. One possible explanation for instances of positive selection along this branch is the occurrence of a closely related gene, LEGCYC1A*, in *L. nanus*. It is unknown whether this copy is found in other genistoid taxa, but this duplication may have affected the molecular evolution of *L. nanus* LEGCYC1A.

It is also possible that LEGCYC1A has a role other than in floral development. LEGCYC1A transcripts, unlike LEGCYC1B, were found in vegetative (leaf) tissue (for both *C. purpurea* and *L. nanus*, chapter 4). It is worth noting that, unlike for LEGCYC1B, positively selected sites were identified in the TCP domain of LEGCYC1A for the *Bowdichia* and *L. digitatus/L. angustifolius* lineages (figure 5-4). This may be significant as the TCP domain is known to have DNA-binding properties (Kosugi and Ohashi, 2002), and therefore heterogeneity within this region may suggest novel binding interactions.

5.4.3 Limitations of this study and general conclusion

It is clear that the molecular evolution of the two LEGCYC paralogues is complex and not uniform across the genistoid clade. Positive selection on LEGCYC1B is correlated with a change in expression pattern of this gene in the *Cadia* lineage, which represents a homeotic shift in expression from adaxial to all domains of the corolla (see chapter 4), which may have contributed to the evolution of radial symmetry in this genus. By contrast, the molecular evolution of LEGCYC1A, with its occasional episodes of diversifying (positive) selection, does not seem to correlate with any single identifiable feature.

A number of factors may affect the estimate of d_N/d_S. One of these is taxon sampling. For genes like LEGCYC, which are evolving rapidly not only by nucleotide substitutions but also by insertion and deletion events, multiple sequence alignment requires the insertion a large number of gaps between divergent sequences. However, likelihood analysis using PAML requires that regions where gaps have been inserted are removed from the data matrix (Yang, 1997). Therefore, estimates of d_N/d_S may be based on a fraction of the codons that make up the gene in a matrix that contains sequences from a wide taxonomic range. It is clear that more sites were excluded from the matrix containing sequences spanning the range of the genistoid clade than that of Ree *et al.* (2004) containing *Lupinus* sequences. Some of these sites may have been

under positive selection, as discussed in section 5.4.2, and therefore analysing divergent sequences may result in an inaccurate estimate of past selection pressures. In addition, the detection of positively selected sites along a particular lineage depends on estimates in the rest of the tree (Ree *et al.*, 2004). This may be particularly sensitive when the proportion of sites with higher non-synonymous rates is low, or the proportion of sites under relaxed purifying selection across the tree is high, as is the case here. It may be that increasing sampling, or reducing the taxonomic range may improve estimates of substitution rates.

Two other genistoid genera, *Acosmium* and *Dicraeopetalum*, have evolved near-radially symmetrical flowers independently of each other and of *Cadia* (Pennington *et al.*, 2000; see figure 5-1). It would be interesting to test if evidence of positive selection is found in LEGCYC genes in these lineages, particularly LEGCYC1B which is strongly implicated in the evolution of the floral phenotype in *Cadia*. Independent evidence of high non-synonymous substitution rates in these actinomorphic lineages would provide greater confidence in relating the signature of positive selection to an indication of functional change.

CHAPTER 6: GENE SILENCING IN *LUPINUS ANGUSTIFOLIUS*

6.1 INTRODUCTION

6.1.1 Investigating gene function

Establishing the effect of a gene on phenotype is crucial for understanding its function. The mRNA expression studies described in chapter 5 provide only an indication of the possible function of the genes of interest. As factors other than mRNA levels alone determine gene activity, such as modifications at the RNA processing and translational-level, these studies cannot prove causal relationship between gene and phenotype. The reverse genetic approach, by studying phenotypes resulting from loss of gene expression, directly implicates a gene in pathways controlling the affected traits. The most widely used reverse genetic approach is insertional mutagenesis, which relies on the insertion of a DNA fragment, used as an identifiable tag, into the genome, and has been extremely successful in characterising genes in diverse model plant species. This approach relies on either transferred DNA (T-DNA) insertions when transformation efficiency is high (examplified in the numerous *Arabidopsis* T-DNA lines), or transposon tagging, for example in *Antirrhinum* (Carpenter & Coen, 1990).

Insertional mutagenesis, however, has several limitations (Thorneycroft *et al.*, 2001; Waterhouse & Helliwell, 2003). First of all, it is untargeted, labour intensive and only suitable for a limited number of model plant species. In addition, this method is not suitable for investigating duplicated genes which are functionally redundant, and may also cause disruption to genes other than those into which the DNA tag is inserted. Previously used targeted methods used for interferring with gene expression, such as cosuppression (*i.e.* the suppression of endogenous gene by insertion of a homologous transgene) or insertion of antisense RNA, are often unpredictable in their outcome (Waterhouse & Helliwell, 2003).

A novel method of gene discovery that circumvents some of these problems is double stranded RNA (dsRNA)-induced gene silencing (Waterhouse & Helliwell, 2003). The introduction of dsRNA in an organism has been found to induce sequence-specific RNA degradation that effectively silences the target gene (reviewed in Bosher & Labouesse, 2000; Guru, 2000, Hammond *et al.*, 2001). This naturally occurring phenomenon, referred to as RNA interference (RNAi) or post-trancriptional gene silencing (PTGS), has evolved as a defense against viruses and transposable DNA elements (Waterhouse *et al.*, 2001). This mechanism appears to be evolutionarily conserved and has been described in wide range of organisms, including invertebrates (*e.g. Caenorhabditis elegans* (Montgomery *et al.*, 1998), *Drosophila* (Hammond *et al.*, 2000)), vertebrates *e.g.* mouse (Yang *et al.*, 2001), as well as plants (Vaucheret *et al.*, 2001) and fungi (*Neurospora*; Pickford *et al.*, 2002).

6.1.2 Mechanism of RNA interference (RNAi)

A simplified model of RNAi is shown in figure 6-1. The process can be divided into two steps: cleavage of introduced dsRNA and subsequent cleavage of endogenous mRNA that is homologous to the short dsRNA fragments (reviewed in Matzke *et al.*, 2001; Waterhouse *et al.*, 2001). The introduction of dsRNA into a host cell triggers a degradation system mediated by a Dicer nuclease. The Dicer-containing complex recognises the ends of dsRNA, and cleaves it in succession to produce short 21-25 nucleotide dsRNA fragments known as short interfering RNAs (siRNAs). These siRNAs assemble and serve as guides for a RNA-induced silencing complex (RISC) that has nuclease activity. The antisense strand of the siRNA then pairs with homologous endogenous mRNA, which is cleaved at approximately the middle of the recognised sequence.

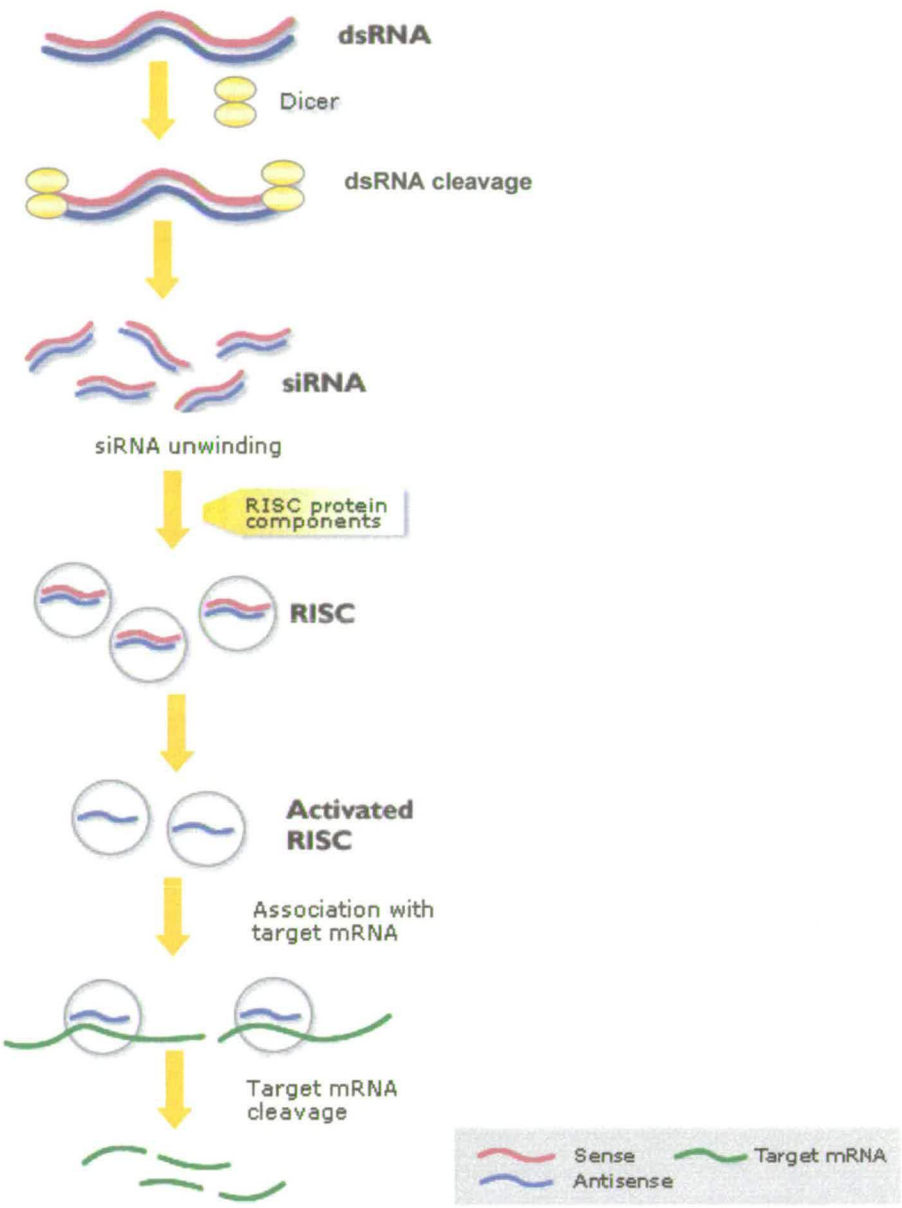


Figure 6-1. Current model of RNA interference (redrawn from the Ambion RNAi resource: http://www.ambion.com/techlib/append/RNAi_mechanism.html). Similar models have been described in plants (Waterhouse *et al.*, 2001), animals (*e.g.* nematodes, Montgomery *et al.*, 1998) and fungi (Pickford *et al.*, 2002). Upon introduction into an organism, long double stranded RNAs (dsRNAs) are processed by a dicer-containing complex into 21-25 bp small interfering RNAs (siRNAs). These siRNAs assemble with an endonuclease-containing complex, known as RNA-induced silencing complexes (RISCs). The anti-sense strand of the siRNA guides the RISC to complementary mRNA, where cleavage is induced.

6.1.3 Use of RNAi in the discovery of gene function

RNA interference has successfully been exploited as a gene silencing technology in several organisms. In *C. elegans*, a large scale genome-wide project was carried out, inhibiting ~86% of the 19,427 predicted genes by expression of dsRNA (Kamath *et al.*, 2003). In plants, insertion of dsRNA-expressing constructs have resulted in effective silencing of the target genes (Chuang & Meyerowitz, 2000; Smith *et al.*, 2001; Stoutjesdijk *et al.*, 2002; Wesley *et al.*, 2001). For example, in *Arabidopsis thaliana*, the insertion of dsRNA fragments from previously characterised floral developmental genes were found to produce phenotypes similar to those of loss-of-function mutants (Chuang & Meyerowitz, 2000). RNAi-inducing transgenes were also found to repress the expression of multiple orthologues in the polyploid *Arabidopsis suecica*, highlighting the potential of this technology for gene discovery in species less amenable to genomic research (Lawrence & Pikaard, 2003).

6.1.4 Experimental background

Gene silencing technology mediated by RNAi was used in this project to investigate the function of *CYC*-like genes in papilionoid legumes. Two *CYC*-like genes, LEGCYC1A and LEGCYC1B, were found to be expressed in the developing flower of *Lupinus nanus*, a genistoid legume with typical papilionoid zygomorphic flowers, in a way that is very similar to *Antirrhinum CYC* (chapter 5). Locus-specific dsRNA constructs were designed to silence each gene individually in another *Lupinus* species, *L. angustifolius*, which can be routinely transformed by *Agrobacterium tumefaciens*-mediated gene transfer (Pigeaire *et al.*, 1997).

6.2 MATERIALS AND METHODS

6.2.1 Silencing construct design

Genomic DNA from *Lupinus angustifolius* cv. Merrit was provided by Susan Barker (University of Western Australia, Perth). Isolation of the partial open reading frame (ORF) of LEGCYC1A and LEGCYC1B from *L. angustifolius* was achieved by PCR amplification and sequencing using primers LEGCYC_F9 (5'- CTT CTA CTT ACA YWT CYT CAG GC -3') close to the start of the ORF, and LEGCYC_R4/R3 respectively (see appendix 2). As silencing specificity is critical to investigate gene function, fragments for the double stranded RNA (dsRNA) constructs were selected based on sequence divergence (*i.e.* no strings of identical 20 bases) between the two *CYC*-like paralogues. In addition, the location of the fragment was specifically chosen upstream of the conserved TCP domain, to prevent any extension 5' of the target region, as observed in *Caenorhabditis elegans* (Sijen *et al.*, 2001), which may compromise silencing specificity if the 5' sequence is conserved between paralogous genes. Oligonucleotide primers for specific amplification of LEGCYC1A and LEGCYC1B fragments (205 bp and 236 bp respectively) were flanked with attB1 or attB2 recognition sites for directional insertion by homologous recombination into the GATEWAY donor vector pDONR207 (Invitrogen, Life Technologies, Inc.). Primers were synthesized by Life Technologies, Inc.: LEGCYC1A (forward) 5'-attB1- TCA AGC AAC AAC AAC AAC AAC CAC -3'; and (reverse), 5'-attB2- TTG GCT GGT TTC TTT GTG -3'; LEGCYC1B (forward) 5'-attB1- TCT TCA AAC AAC ACA TTT TCT C -3' and (reverse), 5'-attB2- TGT CTT TCT TTG GAG CAG -3'. The pDONR207 plasmids containing the locus-specific PCR products were then used to transfer via homologous recombination the gene sequences into pFGC5149 (ChromDB, Arizona, USA), a vector designed for the synthesis of dsRNA, and modified to have GATEWAY recombination sites. This vector contains a spliceable intron from the petunia

Chalcone synthase A gene between the target gene sequences (figure 6-2). Intron-containing constructs have been shown to significantly increase silencing efficiency compared to, for instance, hairpin-RNA constructs which have a spacer region between recombination sites (Smith *et al.*, 2000; Wesley *et al.*, 2001). Correct insertion of the target sequence was verified by sequencing using pFGC5149 specific primers, designed in the regions spanning the two insertion sites (pFGC-F1: 5'- GTA AGG GAT GAC GCA CAA TC -3', pFGC-R1: 5'- TTT CTA CCT TCC CAC AAT TCG -3'; pFGC-F2: 5'- GAA TCT TAC TAA CTT TGT GGA AC -3', pFGC-R2: 5'- GTA AGG ATC TGA GCT ACA C -3'; figure 6-3).

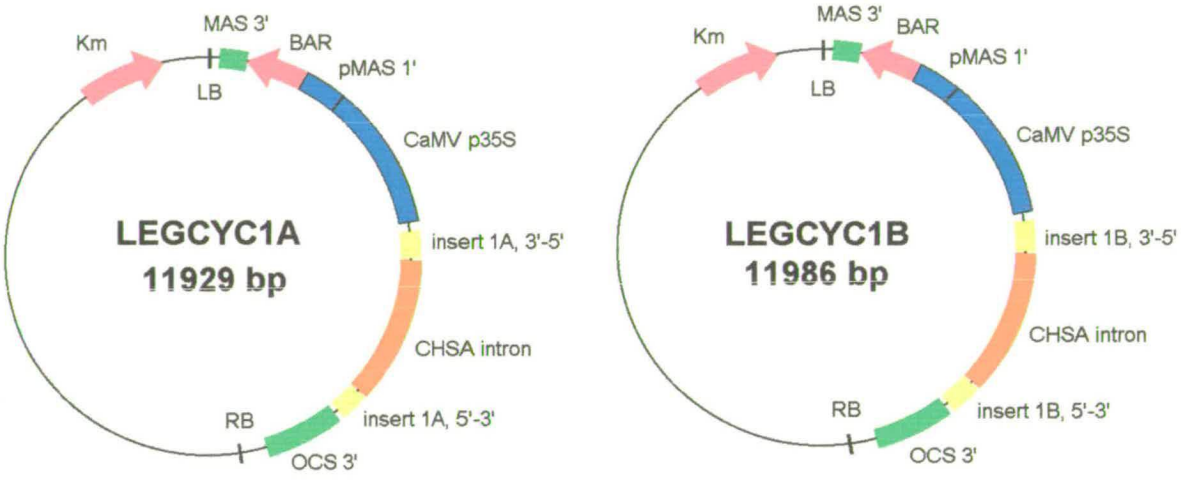


Figure 6-2. Plasmid maps showing the transformed pFGC514 RNAi vector (ChromDB, Arizona, USA) with inserted *CYC* fragments (in yellow), generated with BioEdit v5.0.9 (Hall, 2001). Details of the portion transferred to *L. angustifolius* generating *CYC*-specific dsRNA fragments are given in figure 6.3. The plasmids have a kanamycin resitant gene (Km) for selection of *Agrobacterium tumefaciens*. CaMV p35S: cauliflower mosaic virus promoter, CHSA intron: 1,353 bp fragment from the petunia Chalcone synthase A gene, OCS -3': poly adenylation signal sequence from *A. tumefaciens*, for trancription termination. The selectable marker BAR gene conveys resistance to the herbicide Basta. pMAS 1': plant promoter from *A. tumefaciens*, MAS 3': poly adenylation signal sequence from *A. tumefaciens*. LB: left border repeat from T-DNA; RB: right border repeat from T-DNA.

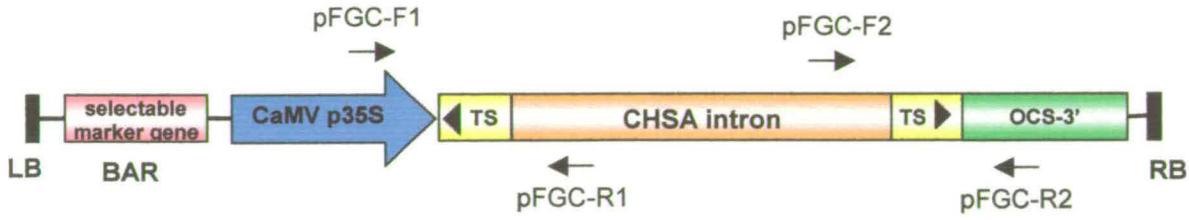


Figure 6-3. Schematic outline of the intron-spliced hairpin RNA construct tranferred to lupins for RNA-mediated gene silencing, from the pFGC5149 vector (ChromDB, Arizona, USA), modified with GATEWAY adaptors for directional insertion of DNA target sequence (TS). The target sequence (TS) fragments are inserted in opposite orientation to form a dsRNA structure. Primers pFGCF1/R1/F2/R2 specifically bind to regions flanking the two cloning sites of pFGC5941, and are therefore transgene specific. Abbreviations are given in figure 6-2.

6.2.2 Gene transfer in *Lupinus angustifolius*

6.2.2a *Agrobacterium tumefaciens* transformation

The Agl0 strain of *Agrobacterium tumefaciens* was transformed with the plasmids described above. Bacterial cells were grown (5.10^8 cells/ml) for inoculation of the explants in a selective tetracycline (50µg/ml) medium as described in Pigeaire *et al.* (1997).

6.2.2b Explant preparation

Approximately 2000 seeds (946 seeds infected with LEGCYC1A construct, and 885 seeds with LEGCYC1B construct) of *L. angustifolius* cv. Merrit were prepared for co-cultivation with Agl0. Details of the protocol and media recipes are given in Pigeaire *et al.* (1997), and illustrated in figure 6-4. Germination of sterile seeds was induced overnight (figure 6-4A), followed by excision of the whole shoot axis including the first two pairs of leaves in the plumule (figure 6-4B). After wounding the apical dome and primordia, the embryonic axis was

placed in co-cultivation medium (figure 6-4C). The wounded shoot apex was inoculated with a drop of Agl0 suspension.

After two days of co-cultivation, explants were transferred to a regeneration medium (figure 6-4D). Selection was initiated two days later by placing a drop of a phosphinothricin (PPT) solution (2mg/ml), the active ingredient of the herbicide Basta, on the apical dome of each were then subcultured every two weeks on the same selective medium (figure 6-4G). After a minimum of 6 months subculture on selective medium, explants are transferred to a root regenerating medium containing indole-3-butyric acid (IBA) (3mg/L). Plants were eventually transferred to a sterile soil mix under glasshouse conditions.

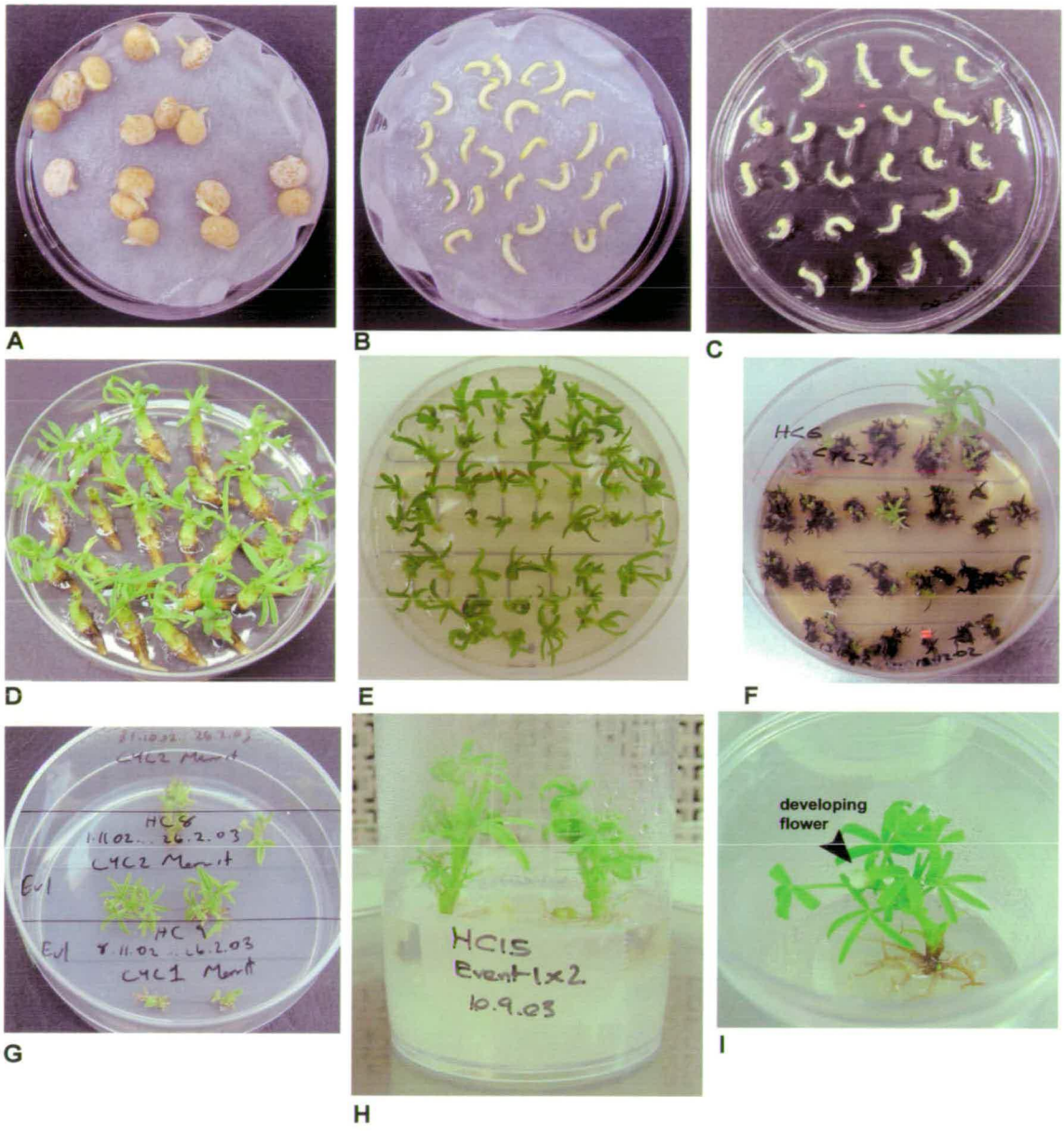


Figure 6-4. Stages in *Lupinus angustifolius* transformation and explant regeneration (following the protocol of Pigeaire *et al.*, 1997). *L. angustifolius* seeds were germinated overnight (A), dissected to expose the apical meristem (B), and co-cultivated with *Agrobacterium* containing the dsRNA construct (C). Explant were regenerated over two days (D). Shoots were then dissected and placed on selective medium containing PPT (20mg/l), the active ingredient of the herbicide Basta (E). Surviving shoots (F) were then subcultured on selective medium (G). When explants reached a certain size (~ 5cm in height), roots were induced (H). At this stage, sterile flowers were observed (I).

6.2.3 Transformant screening

The presence of the transgene in surviving explants was confirmed by PCR using transgene-specific primers pFGC-F2 and pFGC-R2 (see section 6.2.1). DNA was extracted from leaf material from cultured explants 10 months after the initial transformation.

6.3 RESULTS

6.3.1 Frequency of transformation

The frequency of transformation based on PCR screen results (figure 6-5) was 0.85% for LEGCYC1A and 0.23% for LEGCYC1B. This is similar to the average of 0.4% transformation frequency obtained for cv. Merrit by Pigeaire *et al.* (1997).

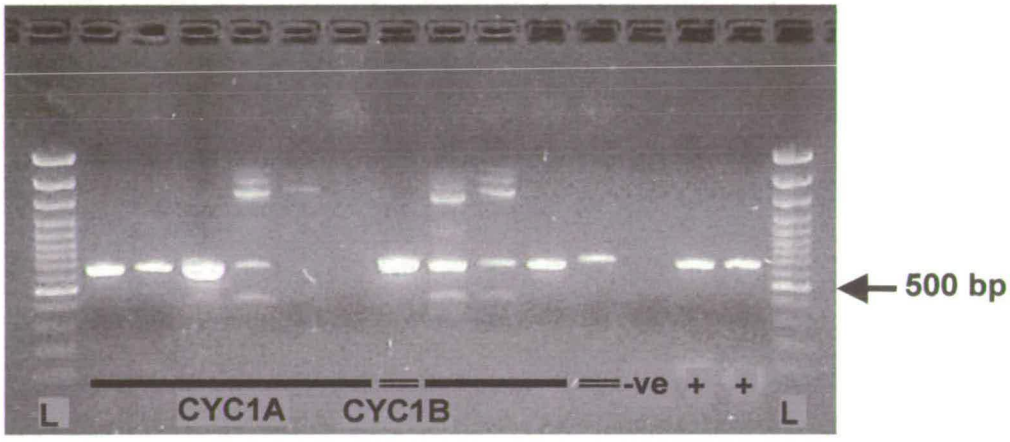


Figure 6-5. Amplification of transgene in surviving explants (L. Hogdson, UWA) using the pFGC5149 specific primers pFGC-F2 and pFGC-R2. Lanes with products from plants transformed with the LEGCYC1A construct are marked by ▬, lanes with products from plants transformed with the LEGCYC1B construct are marked by ═. –ve: negative control, +: positive control (plasmid DNA), L: 100 bp ladder.

6.3.2 Phenotypes of putative transformants

No obvious differences were observed between flowers from T0 putative transformants, shown here with a LEGCYC1A inverted fragment insert, and wild type *L. angutifolius* (figure 6-6). In addition, no differences in vegetative parts were apparent, even though LEGCYC1A was found to be expressed in developing leaves (chapter 4). However, T0 plants are frequently chimeric, containing both transformed and untransformed sectors (Pigeaire *et al.*, 1997), and are not usually informative for examining transgenic phenotype. Recovery of wholly transformed plants is expected in seeds (T1 generation) of T0 plants.

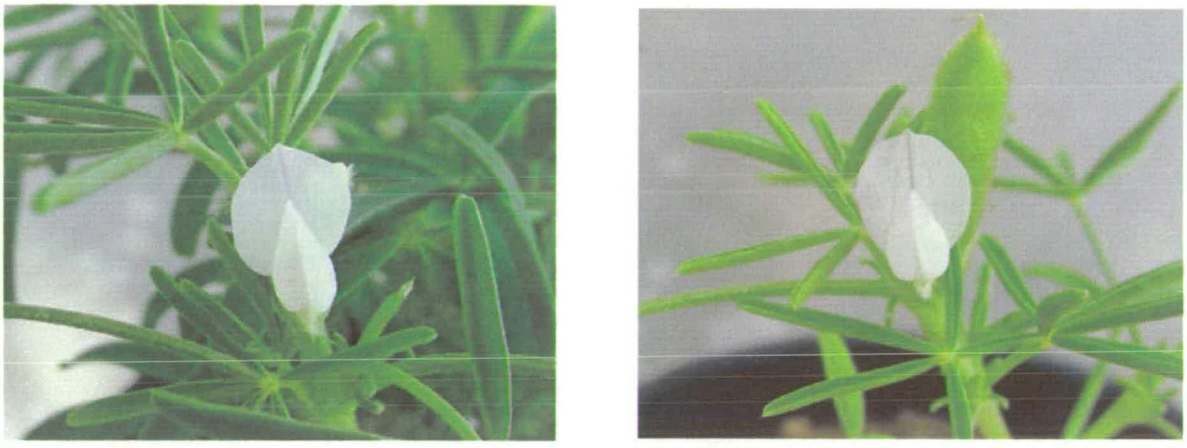


Figure 6-6. Mature flower of T0 plant with LEGCYC1A inverted repeat insert (A) and wild type (B) *L. angustifolius* cv. Merrit. Although no differences were visible, T0 plants are often chimeric and therefore seldom informative in transformation experiments.

6.4 DISCUSSION

6.4.1 Transformation efficiency

Transformation frequencies of plants infected with constructs containing fragments of LEGCYC1A and LEGCYC1B, measured at this stage by the presence of the transgene in T0 explants on selective medium, were within the range expected for cv. Merrit (Pigeaire *et al.*, 1997). However, a greater proportion of plants transformed with the LEGCYC1A construct survived than with LEGCYC1B. Although this could be due to chance, this may also suggest that expression of LEGCYC1B dsRNA may be harmful in some way to the plant. Unlike LEGCYC1A, however, expression of LEGCYC1B was not detected by RT-PCR in developing leaf tissue of *Lupinus* (chapter 4).

6.4.2 Predicted results and limitations of this study

It is not possible at this stage to evaluate the extent and effect of gene silencing mediated by RNAi in transformed cv. Merrit. Seeds from T0 plants, which show no deviation in floral phenotype from the wild type, were collected a year and four months after initiating the experiments. T1 plants will be screened for the presence of the transgene and examined for modification in phenotype.

It is likely that silencing of the LEGCYC copies independently will not cause profound changes in floral morphology, as these genes have overlapping expression patterns and are believed to be partially redundant (chapter 4). However, silencing each copy separately will help define their relative importance in establishing floral symmetry which cannot be inferred by expression pattern alone.

A number of limitations associated with *Lupinus* transformation and gene silencing via RNAi are likely to affect the outcome of this study. First of all, although the effectiveness of

gene silencing by the introduction of intron-spliced inverted repeats was found to be high, the degree of silencing was variable and unpredictable (Wesley *et al.*, 2001). Silencing of *Arabidopsis* developmental genes showed that a majority of transformed plants with dsRNA constructs had reduced but detectable endogenous gene expression, with a low percentage exhibiting near-complete knock-out of the target gene (Chuang & Meyerowitz, 2000; Wesley *et al.*, 2001). Nevertheless, although little is known about the dosage-dependent effect of *CYC*, it is likely that a reduction of expression of *CYC*-like genes would result in significant phenotypic changes. For instance, in teosinte, the lesser accumulation of *TB1* mRNA compared with cultivated maize corresponded to greater branch elongation (Doebley *et al.*, 1997). In this study, however, with the low transformation rate found for cv. Merrit, a range of phenotypes corresponding to different levels of endogenous gene expression may not be recovered.

It is not expected that the small size of the LEGCYC fragments, approximately 200 nucleotides, should affect silencing efficiency. Although Wang & Waterhouse (2001) suggest that silencing is more efficient with constructs of 300 nucleotides or more, effective silencing was obtained with constructs only 98 nucleotides long (Wesley *et al.*, 2001). Synthetic short interfering RNAs (siRNAs), that resemble the 21-23 nucleotide duplexes produced by Dicer from dsRNA, have also been found to mediate specific silencing in cultured cell lines from mammals (Semizarov *et al.*, 2003) and plants (Vanitharani *et al.*, 2003). In mammalian cells, it was found that unlike long dsRNA fragments, siRNA fragments do not trigger an unspecific immune response which generally inhibits gene expression (Stark *et al.*, 1998). The development of siRNA-mediated gene silencing has become one of the fastest growing tools in genetic research.

The silencing efficiency of the contructs may be improved by incorporating promoter as well as exon sequence (Wang & Waterhouse, 2001). In addition to RNA degradation, dsRNAs containing promoter sequence have been found to direct specific methylation of target promoters

resulting in transcriptional gene silencing (Mette *et al.*, 2000). Combining the effects of transcriptional (promoter methylation) and post-transcriptional (RNA degradation) gene silencing mediated by dsRNA may result in a greater reduction in gene expression.

6.4.3 Future work

Phenotypic examination and genetic analysis of T1 plants, through transgene detection (as in section 6.3.1) and quantification of mRNA accumulation of endogenous target gene by real time RT-PCR, will evaluate the success of this study. The extent of functional redundancy between LEGCYC1A and LEGCYC1B will be estimated, and it may be that an observable phenotype may only be observed in double mutants. These may be obtained by crossing stable T1 transformants.

Further transformation experiments may be informative although it is clear that *Lupinus* transformation is a long, labour intensive process with a low success rate (Pigeaire *et al.*, 1997). Efficient transformation systems are being developed for many other legumes species, although the vast majority of these are within the Phaseoleae and Hologalegina clades, and are closely related to the model legumes *Lotus* and *Medicago* (Somers *et al.*, 2003). Nevertheless, *Lupinus* transformation may be valuable to test the role of LEGCYC genes in changes in floral symmetry in taxa from the genistoid clade. Expression data in *Cadia* suggest that radial symmetry may result from an expansion of the expression of one LEGCYC copy. To test this hypothesis, it would be valuable to homeotically express this LEGCYC copy in the lateral and ventral regions of the corolla and androecium of *Lupinus*. This could be achieved by inserting a full length ORF construct under the control of a B class promoter (controlling petal and stamen identity) such as that of *APETALA3*.

CHAPTER 7: CONCLUSIONS AND FUTURE WORK

7.1 Summary of findings

7.1.1 Phylogenetic framework

This study proposed to examine the evolution and function of putative floral symmetry genes across legume lineages, with particular emphasis on taxa with unusual floral morphology. Much of the research on the genetic control of development has focused on a few model organisms to gain insights on the general mechanisms involved in the evolution of morphological traits. In the Leguminosae, these model organisms, such as *Lotus japonicus*, *Medicago truncatula* and *Pisum sativum* are all members of derived clades (*i.e.* Phaseoleae and Hologalegina) of the Papilionoideae, where there is little natural variation in floral morphology especially in floral symmetry. A study such as this one, with a wide taxonomic scope and encompassing clades containing species with diverse floral forms, has revealed aspects of the processes influencing morphological evolution that cannot be obtained by examining model legumes alone.

The phylogenetic component of this project has shown that homologues of *CYCLOIDEA* in the Leguminosae belong to a complex gene family. Unravelling the relationships between of the members of this family was complicated by the rapid and variable rate of evolution of LEGCYC copies, and may have also been affected by unequal taxon sampling. It is difficult to study developmental gene evolution in such a large family as the Leguminosae, and even by narrowing the focus to the Papilionoideae, accounting for two-third of species within this family, the determination of orthology relationship of LEGCYC copies was still problematic. The rapid rate of sequence evolution of LEGCYC copies, two to four times faster than ITS, and the abundance of insertion/deletion events means that unambiguous alignment and robust

phylogenetic analyses of these genes can only be carried out over a reduced taxonomic range, such as the genistoid clade.

Despite the problematic nature of the data, general conclusions can be drawn from the phylogenetic study. Results suggest that *CYC*-like gene duplication has occurred during the evolution of the Leguminosae, probably early in, or prior to, the evolution of the Papilionoideae. The independent duplication of *CYC*-like genes, and maintenance of these duplicate copies, has been found in a variety of plant groups including Antirrhineae (Gübitz *et al.*, 2003; Hileman & Baum, 2003), Gesneriaceae (Citerne *et al.*, 2000) and Solanaceae (K. Coenen, unpublished). The maintenance of duplicate copies in the Papilionoideae does not seem to be affected by deviations from the typical zygomorphic papilionoid flower, either in taxa which have lost lateral and ventral petals (*e.g. Swartzia*) or with radially symmetrical flowers (*e.g. Cadia*, *Acosmium*). *CYC* homologues are also found in the Mimosoideae, characterised by radially symmetrical flowers, suggesting that actinomorphy has not evolved by complete loss of *CYC* genes in this subfamily.

This phylogenetic framework enabled the identification in a number of legume taxa of orthologues of two LEGCYC copies, found to be expressed in the adaxial region of *Lotus japonicus* floral meristems (D. Luo, unpublished) and which are thus candidates for studying the control of floral symmetry in this family. In particular, homologues were found in *Cadia purpurea*, a papilionoid species with unusual radially symmetrical flowers, and in *Lupinus* (*L. nanus*), a taxon closely related to *Cadia* but with typical zygomorphic papilionoid flowers.

7.1.2 Functional inferences from expression data

The expression pattern of the two LEGCYC candidate genes in *Lupinus* was highly similar to that of *Antirrhinum CYC*, and strongly suggests these are involved in the control of floral symmetry in papilionoid legumes. This result is important because it implies that *CYC* genes have been recruited independently in the evolution of zygomorphy in distant angiosperm

lineages, such as Leguminosae and Antirrhineae. It is also suspected that *CYC* genes are involved in the control of zygomorphy in the Asteraceae, another lineage that has evolved bilateral symmetry independently from *Antirrhinum* (Gillies *et al.*, 2002). These separate lines of evidence support the theory that the transient dorsal expression of *CYC* genes in the early stages of axillary meristems as found in *Arabidopsis*, may be a pleisiomorphic "pre-pattern" that has been modified repeatedly in various angiosperm lineages.

Furthermore, these genes are implicated in the evolution of one of the "reversals" from zygomorphy to actinomorphy within the the Papilionoideae. In the unusual papilionoid legume *Cadia* with radially symmetric flowers, one LEGCYC copy (LEGCYC1B) was found to be expressed in all five petals, suggesting that the lateral and ventral petals have acquired dorsal identity through extension of the LEGCYC1B expression domain. This differs from the radial *Antirrhinum* and *Linaria* mutants (Veronicaceae, Lamiales), which develop as a result of loss of expression through transposon insertion (*Antirrhinum*, Luo *et al.*, 1996) or methylation (*Linaria*, Cubas *et al.*, 1999b) of *CYC* genes. It appears that in *Cadia*, radial symmetry is not an evolutionary reversal resulting from a loss-of-function mutation or a loss of *CYC* expression during the later stages of floral development, but a morphological novelty correlated with the expansion of LEGCYC expression. Circumstancial evidence for changes in protein function of LEGCYC1B was provided by the study of sequence evolution, where positive selection may have acted in the *Cadia* lineage.

7.2 Future work

7.2.1 Detailed characterisation of LEGCYC function

Typical papilionoid flowers are similar in their zygomorphic form, with well differentiated standard, wings and keel; the main differences lie in the size of the standard relative to the keel and wings, and in staminal fusion, which is absent in certain lineages, for

instance within the genistoid clade (Crisp *et al.*, 2000). It seems likely that the genetic control of floral symmetry in papilionoid legumes should be similar across members of the subfamily. A gene silencing approach is required to demonstrate this. The results from the gene silencing experiments in *Lupinus* are still pending because *Lupinus* transformation is a long process with a low success rate, particularly in the year-long regeneration phase. Characterising the role of all LEGCYC copies by gene silencing, including LEGCYC2 which is also florally expressed, may be better achieved in model legumes where transformation is more efficient. Legume transformation is the focus of considerable research, and a variety of transformation systems have been developed and improved for many taxa, usually from derived clades within the Papilionoideae (Somers *et al.*, 2003). For example, improvements in strain virulence such as Agl1 in *Medicago truncatula* (Chabaud *et al.*, 2003), or the development of new starting material such as dedifferentiated root cells highly susceptible to *Agrobacterium* infection in *Lotus japonicus* (Lombari *et al.*, 2003), have increased the success rate and decreased regeneration time to four to five months in these taxa. In addition to dsRNA-mediated gene silencing, TILLING (Targeted Induced Local Lesions in Genomes) reverse genetic methodology has been developed for *Lotus japonicus* (M. Parniske, Sainsbury Laboratory, Norwich). This technique allows the identification of induced point mutations in specific genes by PCR. Using a high-throughput method developed by Colbert *et al.* (2001), identification of specific mutant individuals can be achieved by pooling PCR products from different lines and digesting them with an endonuclease that recognises mistmatches in heteroduplexes. There are, therefore, different avenues for investigating gene function by reverse genetics in a variety of legume taxa. Transformation systems can also be used to specifically over-express LEGCYC genes in floral organs, in order to reproduce the expression pattern of LEGCYC1B in *Cadia purpurea*.

7.2.2 Examination of other unusual papilionoid legumes

The *Cadia* case study provides an example of how changes in expression of transcription factors can result in the evolution of novel morphological traits. In the Papilionoideae, many other species have unusual flowers, particularly in the basal lineages in the subfamily, but also within more derived clades (described in Pennington *et al.*, 2000; see chapter 1, figure 1-6). For instance, in the genistoid clade, *Acosmium* and *Dicraeopetalum* also have radially symmetrical flowers, which have evolved independently from each other and from *Cadia* (Pennington *et al.*, 2000). The genetic basis underlying convergent evolution is poorly understood. In *Drosophila*, recent work has shown certain cases of morphological convergence relied on the same genetic mechanisms (reviewed in Richardson & Brakefield, 2003). For instance, the independent loss of trichomes in different *Drosophila* species was correlated with a reduction in levels of expression of the gene *SHAVENBABY* (Sucena *et al.*, 2003). A framework has been established here to study whether the expression of *CYC*-like genes has been modified in a similar way to *Cadia* in *Acosmium* and *Dicraeopetalum*. Such a study would test in flowering plants whether morphological convergence is coupled with parallel genetic changes.

The only known actinomorphic mutant in Papilionoideae is found in cultivars of the butterfly pea *Clitoria ternatea* L. (Phaseoleae) (figure 7-1). Wild type *C. ternatea* flowers are strongly zygomorphic, inverted at maturity with an enlarged standard acting as a platform, and a diadelphous androecium (stamen filaments fused with the exception of the dorsal stamen). By contrast, mutants have five equal large petals similar to the wild type standard, and free stamens.

A: WILD TYPE

B: MUTANT

Figure 7-1. Wild type (A) and mutant (B) *Clitoria ternatea* flowers. In the mutant, all petals are equal and resemble the wild type standard.

Crosses between wild type and mutant plants suggested that two genes may be responsible for the mutant phenotype (Fazlullah *et al.*, 1996). Three partial LEGCYC sequences have already been isolated in *Clitoria* in this study. As the mutant phenotype of *Clitoria* is clearly dorsalised and reminiscent of *Cadia* flowers, it would be very interesting to investigate if LEGCYC genes have expanded their expression domain in a way similar to that found in *Cadia* flowers.

7.2.3 Evolution of floral symmetry in other lineages

Perception of evolutionary trends in the legume family have suggested that less specialised, near-radial flowers as found in certain caesalpinioids are primitive compared to typical entomophilous papilionoid flowers. However, it is still unclear when bilateral symmetry evolved in this family. Within the basal-most lineage of the Leguminosae, the caesalpinioid tribe Cercideae (Wojciechowski, 2003), the genus *Cercis* has flowers which superficially resemble those of papilionoid legumes. Shared features include an enlarged reflexed standard petal,

differentiated asymmetric lateral and ventral petals, arched stamen filaments and a style lying within the keel petals. However, a number of differences led Tucker (2002a) to suggest that the resemblance between the specialised flowers of *Cercis* and papilionoids probably resulted from convergent evolution. These differences include the floral aestivation, organ fusion (absent in *Cercis*), and the onset of asymmetric development, which is apparent only after organ enlargement in *Cercis* flower buds, whereas it is evident from organ inception in typical papilionoids (Tucker, 2002a). *Duparquetia* Baill. is another genus with a basal and isolated position in the family based on recent molecular data, that has heteromorphic petals, with imbricate aestivation characteristic of papilionoids (Klitgaard *et al.*, 2002). Flowers of *Duparquetia* are unique within the Leguminosae in that they resemble those of orchids with stamens united in a hood-like synandrium (Klitgaard *et al.*, 2002).

Among the closest allies of the legume family is the family Polygalaceae, some of whose members have highly zygomorphic flowers described as "pseudo-papilionaceaous" (reviewed in Doyle & Luckow, 2003). The small tropical family Surianaceae, and the genus *Quillaja* (Chilean soap tree), characterised by actinomorphic flowers are also sister groups of the legume family (see figure 7-2).

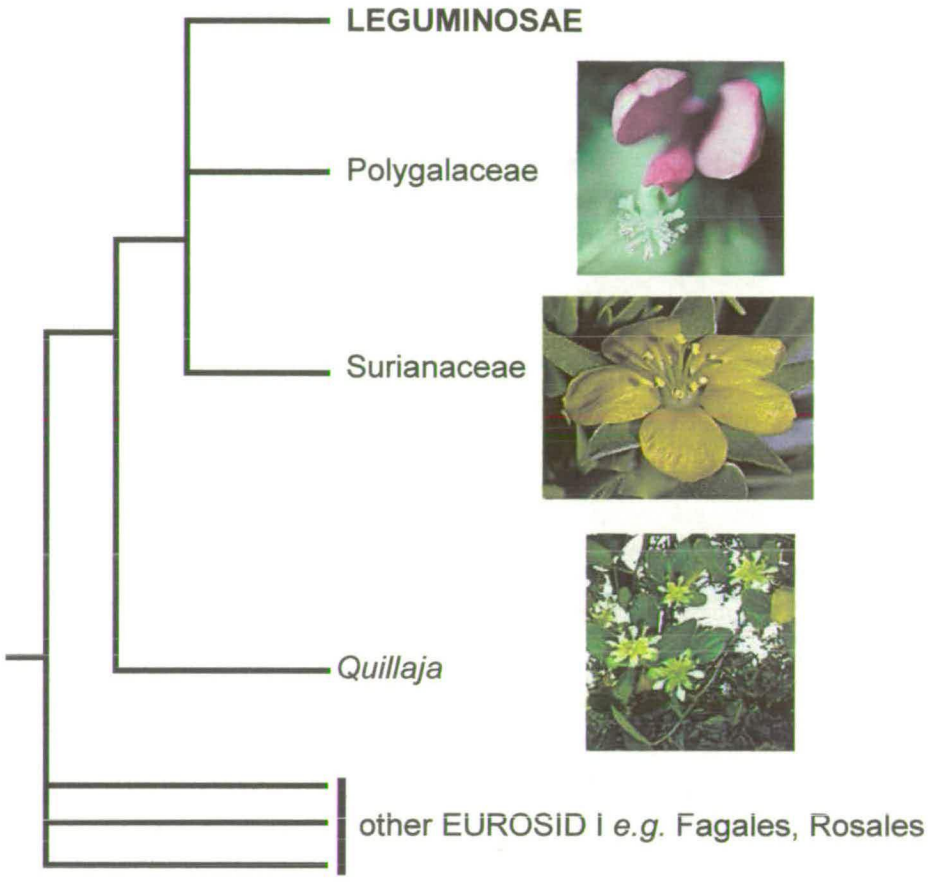


Figure 7-2. Schematic representation of the the Leguminosae and sister clades, based on molecular data (from Doyle & Luckow, 2003). The Polygalaceae (*Polygala paucifolia*; Ken Systma, UW Madison, dept Botany Plant Systematics Collection) have strongly zygomorphic flowers, whereas Surianaceae (*Suriana maritima*; Tim Motley University of Hawaii Botany dept.) and *Quillaja* (*Quillaja saponaria*; San Marcos growers) have radially symmetric flowers.

Examination of *CYC*-like gene expression in these lineages may help understand the evolution of bilateral symmetry in legumes, and the genetic changes that contributed to the development of the highly specialised papilionoid flowers. In particular, comparison of LEGCYC expression in *Cercis* and papilionoid legumes may highlight some similarities in the

genetic control of their floral development, and may change the perception that their flowers are fundamentally different as suggested by Tucker (2002a). The Mimosoideae is another clade of particular interest because it forms a large actinomorphic-flowered group derived from within the Caesalpinioideae. Studying the function and evolution of LEGCYC genes in this subfamily would provide insights into the developmental control of the mimosoid flower.

Continuing advances in legume research, including complete sequencing of *Medicago* and *Lotus* genomes, improvements in transformation systems, and a good phylogenetic framework, are highly favourable for evolution and development research. In *Antirrhinum*, other genes interact with *CYC*, such as the MYB genes *RAD* and *DIV*, conferring lateral and ventral identity respectively (Galego & Almeida, 2002). Although this system may be specific to the Antirrhineae, a better understanding of the control of floral symmetry in papilionoid legumes may be achieved by identifying the genes which affect the development of the strongly differentiated lateral and ventral floral domains, and understanding their interactions with LEGCYC genes.

REFERENCES

Alfaro M.E., Zoller S., Lutzoni F. 2003. Bayes or bootstrap? A simulation study comparing the performance of Bayesian Markov chain Monte Carlo sampling and bootstrapping in assessing phylogenetic confidence. *Mol. Biol. Evol.* 20: 255-266.

Almeida J., Rocheta M., Galego L. 1997. Genetic control of flower in *Antirrhinum majus*. *Development* 124: 1387-1392.

Angiosperm Phylogeny Group. 2003. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG II. *Bot. J. Linn. Soc.* 141: 399-436.

Barrier M., Robichaux R.H., Purugganan M.D. 2001. Accelerated regulatory gene evolution in an adaptive radiation. *Proc. Nat. Acadd. Sci. USA* 98: 10208-10213.

Baum D.A. 1998. The evolution of plant development. *Curr. Opin. Plant Biol.* 1: 79-86.

Baum D.A., Doebley J., Irish V.F., Kramer E.M. 2002. Response: missing links: the genetic architecture of flower and floral diversification. *Trends Plant Sci.* 7: 1360-1385.

Belting H.-G., Shashikant C.S., Ruddle F.H. 1998. Modification of expression and *cis*-regulation of *Hoxc8* in the evolution of diverged axial morphology. *Proc. Nat. Acad. Sci. USA* 95: 2355-2360.

Bosher J.M., Labouesse M. 2000. RNA interference : genetic wand and genetic watchdog. *Nat. Cell. Biol.* 2: E31-E36.

Bowman J.L., Smyth D.R., and Meyerowitz E.M. 1991. Genetic interactions among floral homeotic genes of *Arabidopsis*. *Development* 112: 1-20.

Bradley D., Carpenter R., Sommer H., Hartley N., Coen E. 1993. Complementary floral homeotic phenotypes result from opposite orientations of a transposon at the *plena* locus of *Antirrhinum*. *Cell* 72: 85-95.

Bradley D., Carpenter R., Copsey L., Vincent C., Rothstein S., Coen E. 1996. Control of inflorescence architecture in *Antirrhinum*. *Nature* 379: 791-797.

Caroll S.B. 2000. Endless forms: the evolution of gene regulation and morphological diversity. *Cell* 101: 577-580.

Carpenter R., Coen E.S. 1990. Floral homeotic mutations prodiced by transposon-mutagenesis in *Antirrhinum majus*. *Gene Dev.* 4: 1483-1493.

Chabaud M., de Carvalho-Niebel F., Barker D.G. 2003. Efficient transformation of *Medicago truncatula* cv. Jemalong using the hypervirulent *Agrobacterium tumefaciens* strain AGL1. *Plant Cell Rep.* 22: 46-51.

Chuang C.-F., Meyerowitz E.M. 2000. Specific and heritable genetic interference by double-stranded RNA in Arabidopsis thaliana. *Proc. Nat. Acad. Sci. USA* 97: 4985-4990.

Citerne H.L., Moeller M., Cronk Q.C.B. 2000. Diversity of *cycloidea*-like genes in Gesneriaceae in relation to floral symmetry. *Ann. Bot.* 86: 167-176

Citerne, H.L., Luo, D., Pennington, R.T., Coen, E., and Cronk, Q.C.B. 2003. A phylogenomic investigation of *CYCLOIDEA*-like TCP genes in the Leguminosae. *Plant Physiol.* 131: 1042-1053.

Clark J., Coen E.S. 2002. The *cycloidea* gene can respond to a common dorsoventral prepattern in *Antirrhinum*. *Plant J.* 30: 639-648.

Coen E.S., Meyerowitz E.M. 1991. The war of the whorls: genetic interactions controlling flower development. *Nature* 353: 31-37.

Coen E.S., Nugent J.M. 1994. Evolution of flowers and inflorescences. *Development* Suppl.: 107-116.

Colbert T., Till B.T., Tompa R., Ryenolds S., Steine M.N., Yeung A.T., McCallum C.M., Comai L., Henikoff S. 2001. High throughput screening for induced point mutations. *Plant Physiol.* 126: 480-484.

Crane P.R., Friis E.M., Pedersen K.R. 1995. The origin and early diversification of angiosperms. *Nature* 374: 27-33.

Crisp M.D., Gilmore S., Van Wyk B.-E. 2000. Molecular phylogeny of the genistoid tribes of papilionoid legumes. In *Advances in Legume Systematics, part 9* (eds Heredeen P.S., Bruneau A.), pp 249-276. Royal Botanic Gardens, Kew.

Cronk Q.C.B. 2001. Plant evolution and development in a post-genomic context. *Nat. Rev. Genet.* 2: 607-619.

Cubas P. 2002. Role of TCP genes in the evolution of morphological characters in angiosperms. In *Developmental Genetics and Plant Evolution* (eds Cronk Q.C.B., Bateman R.M., Hawkins J.A.), pp 247-266, Taylor & Francis, London.

Cubas P., Lauter N., Doebley J., Coen E. 1999a. The TCP domain: a motif found in proteins regulating plant growth and development. *Plant J.* 18: 215-222.

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

Cubas P., Coen E., Zapater J.M.M. 2001. Ancient asymmetries in the evolution of flowers. *Curr. Biol.* 11: 1050-1052.

Darwin C. 1868. *The Variations of Animals and Plants under Domestication.* Volume 2. Murray, London.

Dilcher D. 2000. Toward a new synthesis: major evolutionary trends in the angiosperm fossil record. *Proc. Nat. Acad. Sci. USA* 97: 7030-7036.

Doebley J., Stec A., Hubbard L. 1997. The evolution of apical dominance in maize. *Nature* 386: 485-386.

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

Donoghue M.J., Ree R.H., Baum D.A 1998. Phylogeny and the evolution of flower symmetry in the Asterideae. *Trends Plant Sci.* 3: 311-317.

Doyle J.J. 1992. Gene trees and species trees : molecular systematics as one-character taxonomy. *Syst. Bot.* 17: 144-163.

Doyle J.J. 1994. Evolution of a plant homeotic multigene family: toward connecting molecular systematics and molecular developmental genetics. *Syst. Biol.* 43: 307-328.

Doyle J.J., Doyle J.L. 1987. A rapid DNA isolation procedure for small amounts of fresh leaf tissue. *Phytochemical Bulletin* 19: 11-15.

Doyle J.J., Doyle J.L., Ballenger J.A., Dickson E.E., Kajita T., Ohashi H. 1997. A phylogeny of the chloroplast gene *rbc*L in the Leguminosae: Taxonomic correlations and insights into the evolution of nodulation. *Amer. J. Bot.* 84: 541-554.

Doyle J.J., Doyle J.L. 1999. Nuclear protein-coding genes in phylogeny reconstruction and homology assessment: some examples from Leguminosae. In *Molecular Systematics and Plant Evolution* (eds Hollingsworth P.M., Bateman R., Gornall R.J.), pp 229-254. Taylor & Francis, London.

Doyle J.J., Luckow M.A. 2003. The rest of the iceberg. Legume diversity and evolution in a phylogenetic context. *Plant Physiol.* 131: 900-910.

Eisen J.A. 1998. Phylogenomics: improving functional predictions for uncharacterized genes by evolutionary analysis. *Genome Res.* 8: 163-167.

Eisen J.A., Wu M. 2002. Phylogenetic analysis and gene functional predictions: phylogenomics in action. *Theor. Popul. Biol.* 61: 481-487.

Endress P.K. 1997. *Antirrhinum* and Asterideae – evolutionary changes in floral symmetry. *Symposium Series for the Society of Experimental Biology* 53: 133-140.

Endress P.K. 1999. Symmetry in flowers: diversity and evolution. *Int. J. Plant Sci.* 160 (Suppl.): S3-S23.

Endress P.K. 2001. Evolution of floral symmetry. *Curr. Opin. Plant Biol.* 4: 86-91.

Erixon P., Svennblad B., Britton T., Oxelman B. 2003. Reliability of bayesian posterior probabilities and bootstrap frequencies in phylogenetics. *Syst. Biol.* 52: 665-673.

Fazlullah K.A.K., Amirthadevarathinam A., Sudhakar D., Vaidyanathan P. 1996. Inheritance of flower colour and petal shape in blue pea. *Madras Agricultural Journal* 83: 642-643.

Felsenstein J. 1993. PHYLIP (Phylogeny Inference Package) version 3.5c. Department of Genetics, University of Seattle, Washington.

Fukuda T., Yokoyama J., Maki M. 2003. Molecular evolution of *cycloidea*-like genes in Fabaceae. *J. Mol. Evol.* 57: 588-597.

Galego L., Almeida J. 2002. Role of *DIVARICATA* in the control of dorsoventral asymmetry in *Antirrhinum* flowers. *Gene Dev.* 16: 880-891.

Gaudin V., Lunness P.A., Fobert P.R., Towers M., Riou-Khanlichi C., Murray J.A.H., Coen E., Doonan J.H. 2000. The expression of D-cyclin genes defines distinct developmental zones in snapdragon apical meristems and is locally regulated by the cycloidea gene. *Plant Physiol.* 122: 1137-1148.

Gillies A.C.M., Cubas P., Coen E.S., Abbott R.J. 2002. Making rays in the Asteraceae: genetics and evolution of radiate versus discoid flower heads. In *Developmental Genetics and Plant Evolution* (eds Cronk Q.C.B., Bateman R.M., Hawkins J.A.), pp 233-246, Taylor & Francis, London.

Giurfa N.M., Dafni A., Neal P.R. 1999. Floral symmetry and its role in plant-pollinator systems. *Int. J. Plant Sci.* 160 (supplement): S41-S50.

Goldman N., Yang Z. 1994. A codon based model nucleotide substitution model for protein-coding DNA sequences. *Mol. Biol. Evol.* 11: 725-736.

Gübitz T., Caldwell A., Hudson A. 2003. Rapid molecular evolution of CYCLOIDEA-like genes in *Antirrhinum* and its relatives. *Mol. Biol. Evol.* 20: 1537-1544.

Guru T. 2000. A silence that speaks volumes. *Nature* 404: 804-808.

Hall T. 2001. BioEdit v5.0.9. Department of Microbiology, North Carolina State University.

Hammond S.M., Bernstein E., Beach D., Hannon G.J. 2000. An RNA-directed nuclease mediates post-transcriptional silencing in *Drosophila* cells. *Nature* 404: 293-296.

Hebsgaard S.M., Korning P.G., Tolstrup N., Engelbrecht J., Rouze P., Brunak S. 1996. Splice site prediction in *Arabidopsis thaliana* DNA by combining local and global sequence information. *Nucleic Acids Res.* 24: 3439-3452.

Henikoff S., Henikoff J.G. 1992. Amino acid substitution matrices from protein blocks. *Proc. Nat. Acad. Sci. USA* 89 : 10925-10919

Hershkovitz M.A., Zimmer E.A., Hahn W.J. 1999. Ribosomal DNA sequences and angiosperm systematics. In *Molecular Systematics and Plant Evolution* (eds Hollingsworth P.M., Bateman R., Gornall R.J.), pp 268-326. Taylor & Francis, London.

Hileman L.C., Baum D.A. 2003. Why do paralogs persist? Molecular evolution of CYCLOIDEA and related floral symmetry genes in Antirrhineae (Veronicaceae). *Mol. Biol. Evol.* 20: 591-600

Hileman L.C., Kramer E.M., Baum D.A. 2003. Differential regulation of symmetry genes and the evolution of floral morphologies. *Proc. Nat. Acad. Sci. USA* 100: 12814-9

Hillis D.M. 1996. Inferring complex phylogenies. *Nature* 383: 130-131.

Holder M., Lewis P.O. 2003. Phylogeny estimation: traditional and Bayesian approaches. *Nat Rev. Genet.* 4: 275-284.

Hu J.-M., Lavin M., Wojciechowski M., Snaderson M.J. 2000. Phylogenetic systematics of the tribe Millitteae (Leguminosae) based on chloroplast *trn*K/*mat*K sequences, and its implications for evolutionary patterns in Papilionoideae. *Amer. J. Bot.* 87: 418-430

Hubbard L., McSteen P., Doebley J., Hake S. 2002. Expression patterns and mutant phenotype of *teosinte branched1* correlated with growth suppresion in maize and teosinte. *Genetics* 162: 1927-1935.

Huelsenbeck J.P., Ronquist F. 2001. MRBAYES: Bayesian inference of phylogeny. *Bioinformatics* 17: 754-755

Huelsenbeck J.P., Ronquist F., Nielsen R., Bollback J.P. 2001. Bayesian inference of phylogeny and its impact on evolutionary biology. *Science* 294: 2310-2314.

Jesson J.K., Barrett S.C.H. 2002. The genetics of mirror-image flowers. *Proc. R. Soc. Lond. B* 269: 1835-1839.

Kajita T., Ohashi H., Tateishi Y., Bailey D., Doyle J.J. 2001. *rbc*L legume phylogeny, with particular reference to Phaseoleae, Millitteae, and allies. *Syst. Bot.* 26: 515-536.

Kamath R.S., Fraser A.G., Dong Y., Poulin G., Durbin R., Gotta M., Kanapin A., Le Bot N., Moreno S., Sohrmann M., Welchman D.P., Zipperlen P., Ahringer J. 2003 Systematic functional analysis of the *Caenorhabditis elegans* genome using RNAi. *Nature* 421: 231-237.

Klitgaard B.B., Forest F., Bruneau A., Banks H. 2002. *Duparquetia* (Leguminosae: Caesalpinioideae): an egnimatic basal legume with orchid-like flowers. Flowers: Diversity, Development and Evolution, Institute of Systematic Botany and Botanic Garden, Zürich 2002.

Kneller D.G., Cohen F.E., Langridge R. 1990. Improvements in protein secondary structure prediction by an enhanced neural network. *J. Mol. Biol.* 214: 171-182.

Kosugi S., Ohashi Y. 1997. PCF1 and PCF2 specifically bind to *cis*-elements in the rice proliferating cell nuclear antigne gene. *Plant Cell* 9: 1607-1619.

Kosugi S., Ohashi Y. 2002. DNA binding and dimerization specificity and potential targets for the TCP protein family. *Plant J.* 30: 337-348.

Lamb R.S., Irish V.F. 2003. Functional divergence within the *APETALA3/PISTILLATA* floral homeotic gene lineages. *Proc. Nat. Acad. Sci. USA* 100 : 6558-6563.

Lavin M., Pennington R.T., Klitgaard B.B., Sprent J.I., De Lima H.C., Gasson P.E. 2001. The dalbergioid legumes (Fabaceae): delimitation of a pantropical monophyletic clade. *Amer. J. Bot.* 88: 503-533.

Lawrence R.J., Pikaard C.S. 2003. Transgene-induced RNA interference: a strategy for overcoming gene redundancy in polyploids to generate loss-of-function mutations. *Plant J.* 36:114-21.

Lawton-Rauh A.L., Alvarez-Buylla E.R., Puruggana M.D. 2000. Molecular evolution of flower development. *TREE* 15: 144-149.

Lewis, E.B. 1978. A gene complex controlling segmentation in *Drosophila. Nature* 276: 565-570.

Linnaeus C. 1749. *De Peloria*. Diss. Ac. Amoenitates Acaddemicae III, Uppsala.

Lombari P., Ercolano E., El Alaoui H., Chiurazzi M. 2003. A new transformation-regeneration procedure in the model legume *Lotus japonicus*: root explants as a source of large numbers of cells susceptible to *Agrobacterium*-mediated transformation. *Plant Cell Rep.* 21: 771-777.

Lopez P., Casana D., Philippe H. 2002. Heterotachy, an important process in protein evolution. *Mol. Biol. Evol.* 19: 1-7

Lukens L., Doebley J. 2001. Molecular evolution of the teosinte branched gene among maize and related grasses. *Mol. Biol. Evol.* 18:627-38.

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

Luo D., Carpenter R., Copsey L., Vincent C., Clark J., Coen E. 1999. Control of organ asymmetry in flowers of *Antirrhinum. Cell* 99: 367-376.

Lynch M., Force A. 2000. The probability of duplicate gene preservation by subfunctionalization. *Genetics* 154: 459-73.

Matzke M., Matzke A.J., Kooter J.M. 2001. RNA: guiding gene silencing. *Science* 293: 1080-1083.

McSteen P., Hake S. 1998. Genetic control of plant development. *Curr. Opin. Biotech.* 9: 189-195.

Mette M.F., Aufsatz W., van der Winden J., Matzke M.A., Matzke A.J.M. 2000. Trancriptional silencing and promoter methylation triggered by double-stranded RNA. *EMBO J.* 19: 5194-5201.

Moniz de Sa M., Drouin G. 1996 Phylogeny and substitution rates of angiosperm actin genes. *Mol. Biol. Evol.* 13:1198-1212.

Montgomery M.K., Xu S., Fire A. 1998. RNA as a target of double stranded RNA-mediated genetic interference in *Caenorhabditis elegans*. *Proc. Nat. Acad. Sci. USA* 95: 15502-15507.

Muse S.V., Gaut B.S. 1994. A likelihood approach for comparing synonymous and non-synonymous substitution rates, with application to the chloroplast genome. *Mol. Biol. Evol.* 11: 715-724.

Nath U., Crawford B.C.W., Carpenter R., Coen E. 2003. Genetic control of surface curvature. *Science* 299: 1404-1407.

Neal P.R., Dafni A., Guirfa M. 1998. Floral symmetry and its role in plant-pollinator systems: terminology, distribution and hypotheses. *Annu. Rev. Ecol. Syst.* 29: 345-373.

Nielsen R., Yang Z. 1998. Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics* 148: 929-936.

Ochman H., Gerber A.S., Hartl D.L. 1988. Genetic applications of inverse polymerase chain reactions. *Genetics* 120: 621-623.

Ohta T. 2002. Near-neutrality in evolution of genes and gene regulation. *Proc. Nat. Acad. Sci. USA* 99: 16134-16137.

Palatnik J.F., Allen E., Wu X., Schommer C., Schwab R., Carrington J.C., Wiegel D. 2003. Control of leaf morphogenesis by microRNAs. *Nature* 425: 257-263.

Pennington R.T., Klitgaard B.B., Ireland H., Lavin M. 2000. New insights into floral evolution of basal Papilinoideae from molecular phylogenies. In *Advances in Legume Systematics, part 9* (eds Heredeen P.S., Bruneau A.), pp 233-248. Royal Botanic Gardens, Kew.

Pennington R.T., Lavin M., Ireland H., Klitgaard B., Preston J., Hu J.-M. 2001. Phylogenetic relationships of basal papilinoid legumes based upon sequences of the cholorplast *trnL* intron. *Syst. Bot.* 26: 537-556.

Pickford A.S., Catalanotto C., Cogoni C., Macino G. 2002. Quelling in *Neurospora crassa*. *Adv. Genet.* 46: 277-303.

Pigeaire A., Abernethy B., Smith P.M., Simpson K., Fletcher N., Lu C.-Y., Atkins C.A., Coenish E. 1997. Transformation of a grain legume (*Lupinus angustifolius* L.) via *Agrobacterium tumefaciens*-mediated gene transfer to shoot apices. *Mol. Breeding* 3: 341-349.

Polhill R.M. 1981. Papilionoideae. In *Advances in Legume Systematics, part 1* (eds Polhill R.M. Raven P.H.), pp 191-208. Royal Botanic Gardens, Kew.

Posada D., Crandall K.A. 1998. Modeltest: testing the model of DNA substitution. *Bioinformatics* 14: 817-818.

Purugganan M.D., Suddith J.I. 1998. Molecular population genetics of the *Arabidopsis* CAULIFLOWER regulatory gene: noneutral evolution and naturally occurring variation in floral homeotic function. *Proc. Nat. Acad. Sci. USA* 95: 8130-8134.

Ree R.H., Donoghue M.J. 1999. Inferring rates of change in flower symmetry in asterid angiosperms. *Syst. Biol.* 48: 633-641.

Ree R.H., Citerne H.L., Lavin M., Cronk Q.C.B. 2004. Heterogeneous selection on LEGCYC paralogs in relation to flower morphology and the phylogeny of *Lupinus* (Leguminosae). *Mol. Biol Evol.* 21: 321-331.

Remington D.L., Purugganan M.P. 2002. *GAI* homologues in the Hawaiian silversword alliance (Asteraceae-Madiinae): molecular evolution of growth regulators in a rapidly diversifying plant lineage. *Mol. Biol. Evol.* 19: 1563-1574

Richardson M.K., Brakefield P.M. 2003. Hotspots for evolution. *Nature* 424: 894-895.

Rudall P.J., Bateman R.M. 2003. Evolutionary changes in flowers and inflorescences: evidence from naturally occurring terata. *Trends Plant Sci.* 8: 1360-1385.

Schmidt H.A., Strimmer K., Vingron M., von Haeseler A. 2002. TREEPUZZLE: maximum likelihood phylogenetic analysis using quatets and parallel computing. *Bioinformatics* 18: 502-504.

Semizarov D., Frost L., Sarthy A., Kroeger P., Halbert D.N., Fesik S.W. 2003. Specificity of short interfering RNA determined through gene expression signatures. *Proc. Nat. Acad. Sci. USA* 100: 6347-6352.

Sharp P.M. 1997. In search of molecular darwinism. *Nature* 385: 111-112.

Shepard K.A., Purugganan M.D. 2002. The genetics of plant morphological evolution. *Curr. Opin. Plant Biol.* 5: 49-55.

Siebert P.D., Chenchick A., Kellogg D.E., Lukyanov K.A., Lukyanov S.A. 1995. An improved PCR method for walking in uncloned genomic DNA. *Nucleic Acids Res.* 23: 1087-1088.

Sijen T., Fleenor J., Simmer F., Thijssen K.L., Parrish S., Timmons L., Plasterk R.H., Fire A. 2001. On the role of RNA amplification in dsRNA-triggered gene silencing. *Cell* 107:465-76.

Smith N.A., Singh S.P., Wang M.-B., Stoutjesdijk P.A., Green A.G., Waterhouse P.M. 2000. Total silencing by intron-spliced hairpin RNAs. *Nature* 407: 319-320.

Soltis P.S., Soltis D.E., Chase M.W. 1999. Angiosperm phylogeny inferred from multiple genes as a tool for comparative biology. *Nature* 402: 402-404.

Somers D.A., Samac D.A., Olhoft P.M. 2003. Recent advances in legume transformation. *Plant Physiol.* 131: 892-899.

Stark G.R., Kerr I.M., Williams B.R., Silverman R.H., Schreiber R.D. 1998. How cells respond to interferons. *Annu. Rev. Biochem.* 67: 227-64.

Stebbins G.L. 1974. *Flowering Plants: Evolution Above the Species Level.* Harvard University Press.

Stoutjesdijk P.A., Singh S.P., Liu Q., Hurlstone C.J., Waterhouse P.M., Green A.G. 2002. hpRNA-mediated targeting of the Arabidopsis FAD2 gene gives highly efficient and stable silencing. *Plant Physiol.* 129: 1723-1731.

Stubbe H. 1966. *Genetik und Zytologie von* Antirrhinum *L. sect.* Antirrhinum. Veb Gustav Frischer Verlag, Jena.

Sucena E., Delon I., Jones I., Payre F., Stern D.L. 2003. Regulatory evolution of *shavenbaby/ovo* underly multiple cases of morphological parallelism. *Nature* 424: 935-938.

Suzuki Y., Glazko G.V., Nei M. 2002. Overcredibility of molecular phylogenies obtained by Bayesian phylogenetics. *Proc. Nat. Acad. Sci. USA* 99: 16138-16143.

Swofford D.L. 2001. PAUP*: phylogenetic analysis using parsimony (* and other methods). Version 4. Sinauer Associates, Sunderland, Mass.

Theiβen G., Becker A., Winter K.-U., Münster T., Kirchner C., Saedler H. 2002. How the land plants learned their floral ABCs: the role of MADS-box genes in the evolutionary origin of flowers. In *Developmental Genetics and Plant Evolution* (eds Cronk Q.C.B., Bateman R.M., Hawkins J.A.), pp 173-205, Taylor & Francis, London.

Thompson J.F., Gibson F., Plewmiak F., Jenamougin F., Higgins D.G. 1997. The ClustalX window interface: flexible strategies for multiple sequence alignmnet aided by quality analysis tool. *Nucleic Acids Res.* 25: 4876-4882.

Thorley J.L., Page R.D.M. 2000. RadCon: Phylogenetic tree comparison and consensus. *Bioinformatics* 16: 486-487.

Thorneycroft D., Sherson S.M. Smith S.M. 2001. Using gene knockouts to investigate plant metabolism. *J. Exp. Bot.* 52: 1593-1601.

Triglia T., Peterson M.G., Kemp D.J. 1988. A procedure for in vitro amplification of DNA segments that lie outside the boundaries of known sequences. *Nucleic Acids Res.* 16: 81-86.

Tucker S.C. 1984. Origin of symmetry in flowers. In Contemporary Problems in Plant Anatomy. (eds White R.A., Dickison W.C.), pp: 351-395. Acaddemic Press, Inc., London.

Tucker S.C. 1999. Evolutionary lability of symmetry in early floral development. *Int. J. Plant Sci.* 160 (supplement): S25-S39.

Tucker S.C. 2002a. Floral ontogeny of *Cercis* (Leguminosae: Caesalpinioideae: Cercideae): does it show convergence with Papilionoids? *Int. J. Plant Sci.* 163: 75-87.

Tucker S.C. 2002b. Floral ontogeny in Sophoreae (Leguminosae: Papilinoideae) III: *Cadia purpurea*, with radia symmetry and random petal aestivation. *Amer. J. Bot.* 89: 748-757.

Tucker S.C. 2003. Floral development in legumes. *Plant Physiol.* 131: 911-926.

Van der Maesen L.J.G. 1970. Primitiae Abricanae VIII. A revision of the genus *Cadia* Forskae (Caes.) and some remarks regarding *Dicraeopetalum* Harms (Pap.) and *Platycelyphium* Harms (Pap.). *Acta Bot. Neerl.* 19: 227-248.

Vanitharani R., Chellappan P., Fauquet C.M. 2003. Short interefering RNA-mediated interference of gene expression and viral DNA accummulation in cultured plant cells. *Proc. Nat. Acad. Sci. USA* 100: 9632-9636.

Vaucheret H., Béclin C., Fagard M. 2001. Post-trancriptional gene silencing in plants. *J. Cell Sci.* 114: 3083-3091.

Vollbrecht E.B., Veit N., Sinha N., Hake S. 1991. The developmental gene *Knotted-1* is a member of a maize homeobox gene family. *Nature* 250: 241-243.

Wang M.-B., Waterhouse P.M. 2001. Application of gene silencing in plants. *Curr. Opin. Plant Biol.* 5: 146-150.

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

Waterhouse P.M., Wang M.-B., Lough T. 2001. Gene silencing as an adaptive defence against viruses. *Nature* 411: 834-842.

Waterhouse P.M., Helliwell C.A. 2003. Exploring plant genomes by RNA-induced gene silencing. *Nat. Rev. Genet* 4: 29-38.

Watson L., Dallwitz M. J. 1992 (onwards). The Families of Flowering Plants: Descriptions, Illustrations, Identification, and Information Retrieval. Version: 14th December 2000. http://biodiversity.uno.edu/delta/

Weberling F. 1989a. *Morphology of flowers and inflorescences*. Cambridge University Press.

Weberling F. 1989b. Structure and evolutionary tendencies of inflorescences in the Leguminosae. In *Advances in Legume Biology* (eds Stirton C.H., Zarucchi J.L.) *Mongr. Syst. Bot. Missouri Gard. 29: 35-58.*

Wesley S.V., Helliwell C.A., Smith N.A., Wang M.B., Rouse D.T., Liu Q., Gooding P.S., Singh S.P., Abbott D., Stoutjesdijk P.A., Robinson S.P., Gleave A.P., Green A.G., Waterhouse P.M. 2001. Construct design for efficient, effective and high-througput gene silecing in plants. *Plant J.* 27: 1-12.

Wojciechowski M.F. 2003. Reconstructing the evolution of legumes (Leguminosae): an early 21st century perspective. In *Advances in Legume Systematics, part 10, Higher Level Systematics.* (eds Klitgaard B.B., Bruneau A.), pp 5-35. Royal Botanic Gardens, Kew.

Yang S., Tutton S., Pierce E., Yoon K. 2001. Specific double-stranded RNA interference in undifferentiated mouse embryonic stem cells. *Mol. Cell. Biol.* 21: 7076-7016.

Yang Z. 1997. PAML: a computer package for phylogenetic analysis by maximum likelihood. *Comput. Appl. Biosci.* 13:555-556.

Yang Z. 1998. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. *Mol. Biol. Evol.* 15:568-573.

Yang Z., Bielawski J.P. 2000. Statistical methods for detecting molecular adaptation. *TREE* 15: 497-503.

Yang Z., Nielsen R., Goldman N., Pedersen A.M. 2000. Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* 155:431-449.

Yang Z., Nielsen R. 2002. Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. *Mol. Biol. Evol.* 19: 908-917.

Zhang Z., Gurr S.J. 2000. Walking into the unknown: a 'step down' PCR-based technique leading to the direct sequence analysis of flanking genomic DNA. *Gene* 253: 145-50.

APPENDIX 1: MOLECULAR PROTOCOLS

Appendix 1A. Small scale total DNA extraction using a 2X CTAB method modified from Doyle and Doyle (1987).

One to two discs of silica dried or fresh leaf material were harvested for each extraction. These were flash frozen in liquid nitrogen then finely ground with the addition of fine grade acid purified dry sand, and PVPP (polyvinylpolypyrrolidone) to help remove secondary plant compounds such as polyphenolics, tannins and quinones. 1ml of 2X CTAB extraction buffer (2% CTAB, 20 mM EDTA, 100 mM Tris-HCl pH 8.0, 1.4 M NaCl) with added 0.2% mercapto-ethanol added to the ground leaf material was incubated at 65°C for 30 to 45 minutes. The samples were extracted two to three times with 500µl 24:1 chloroform: isoamylalcohol to precipitate proteins and carbohydrates. The samples were inverted to obtain a momentary single phase, left on a shaker for 15 minutes, then centrifuged for 10 minutes at 13,000 rpm. The aqueous supernatant was transferred to a clean tube after each extraction. Nucleic acids were precipitated in 600µl of –20°C isopropan-2-ol overnight at –20C, then centrifuged for 10 minutes at 13,000 rpm. After discarding the supernatant, the pellet was washed with 1ml wash buffer (76% ethanol, 10mM NH_4Ac) and left on a shaker at least 2 hours to dissolve the CTAB from the CTAB-nucleic acid complex, then centrifuged for 5 minutes at 13,000 rpm. The wash buffer was discarded and the pellet vacuum-dried for 5 minutes. The dried pellet was resuspended in 50 to 75µl TE (10mM Tris-HCl, 1mM EDTA pH 8.0). DNA concentration was estimated by electrophoresis on a 1% agarose gel run for 1 hour at 80V in 1X TBE buffer with a concentration marker.

Appendix 1B. Protocols for tissue fixation in FAA (and paraformaldehyde (PFA) (similar to the Barton laboratory protocol, http://www-ciwdpb.stanford.edu/research/barton/in_situ_protocol.html).

Fixative	FAA (2% formaldehyde, 5% HOAc, 60% ethanol)	4% PFA (in 1X PBS)
	FAA, vacuum 10 min, at least x3	PFA, vacuum 10 min, at least x3
	FAA O/N 4°C	PFA O/N 4°C
Tissue dehydration	70% ethanol 5 min on ice	1 X PBS 30 min x2
	70% ethanol 1hr	30% ethanol 1hr
	80% ethanol 1hr	40% ethanol 1hr
	95% ethanol O/N	50% ethanol 1hr
	100% ethanol 1hr x2	60% ethanol 1hr
	ethanol:histoclear 2:1 1hr	70% ethanol 1hr
	1:1 1hr	80% ethanol 1hr
	1:2 1hr	95% ethanol O/N
	1:3 1hr	100 % ethanol 30 min x2
	100% histoclear 1hr x2	100% ethanol 1hr x2
		ethanol:histoclear 2:1 1hr
		1:1 1hr
		1:2 1hr
		1:3 1hr
		100% histoclear 1hr x2
Wax embedding	Paraplast changed twice a day for at least 3 days	

Appendix 1C. RNA probe synthesis protocols from E. Coen's laboratory (described in Bradley *et al.*, 1993) at the John Innes Centre, Norwich (JIC) and Justin Goodrich's labotatory (similar to the Barton laboratory protocol, http://www-ciwdpb.stanford.edu/research/barton/in_situ_protocol.html) at the Institute of Cell and Molecular Biology (ICMB), University of Edinburgh.

	JIC protocol	ICMB protocol
Reaction mix (25µl) incubated 1hr at 37°C	template 4µg 10X transcription buffer 5mM ATP,GTP,CTP 2.5 µl 1mM DIG-UTP 2.5 µl RNAse inhibitor 1µl RNAse polymerase 1µl	template 1µg 10X transcription buffer 5mM ATP,GTP,CTP 2.5 µl 1mM DIG-UTP 2.5 µl RNAse inhibitor 1µl RNAse polymerase 1µl
Reaction end	1X mineral salts 75 µl tRNA (100mg/ml) 2 µl DNase (RNAse free) 1 µl in reaction mix incubated at 37°C for 20 min	dH_20 75 µl tRNA (100mg/ml) 1 µl DNase (RNAse free) 1 µl in reaction mix incubated at 37°C for 10 min
Precipitation	NH_4Ac 3.8M 100 µl 100% ethanol 600 µl 10 min on dry ice centrifuge 15 min, wash in 200 µl 70% ethanol, centrifuge again and dry resuspended in 50 µl dH_20	NH_4Ac 4M 100 µl 100% ethanol 600 µl 20 min on ice resuspended in 100 µl dH_20
Carbonate hydrolysis ~ 30 min at 60°C	equal amount of X2 carbonate buffer (80mM $NaHCO_3$, $120Na_2CO_3$)	
Precipitation	10% Hac 10 µl 3M NaAC 12 µl 100% ethanol 312 µl	10% Hac 10 µl 3M NaAC 21 µl 100% ethanol 420 µl

Appendix 1D. Protocols for RNA *in situ* hybridisation from E. Coen's laboratory (described in Bradley *et al.*, 1993) at the John Innes Centre, Norwich (JIC) and Justin Goodrich's labotatory (similar to the Barton laboratory protocol, http://www-ciwdpb.stanford.edu/research/barton/in_situ_protocol.html) at the Institute of Cell and Molecular Biology (ICMB), University of Edinburgh.

	JIC protocol	ICMB protocol
Section pretreatment		
1. tissue rehydration	100 % histoclear 10 min x2 100% ethanol 1 min x2 95% ethanol 45s 85% ethanol, 0.85% saline 45s 50% ethanol, 0.85% saline 45s 30% ethanol, 0.85% saline 45s 0.85% saline 2 min 1X PBS 2 min	100 % histoclear 10 min x2 100% ethanol 2 min x2 95% ethanol 2 min 90% ethanol 2 min 80% ethanol 2 min 60% ethanol 2 min 30% ethanol 2 min water 2 min 2X SSC 15 min
2. protease treatment	pronase (0.125mg/ml in 100mM Tris-HCl and 50mM EDTA) 12 min	proteinase K (1µg/ml in 100mM Tris-HCl and 50mM EDTA) 30 min, 37°C
3. tissue fixation	glycine (0.2% in 1X PBS) 3 min 1X PBS 2 min 4% PFA 10 min 1X PBS 2 min x2	glycine (2mg/ml in 1X PBS) 2 min 1X PBS 2 min x2 4% PFA 10 min 1X PBS 5 min x2
4. acetic anhydride treatment	acetic anhydride and 0.1 M triethanolamine for 10 min, stirring	acetic anhydride and 0.1 M triethanolamine for 10 min, stirring
5. wash and dehydration	1X PBS 2 min 0.85% saline 2 min 30% ethanol, 0.85% saline 30s 50% ethanol, 0.85% saline 30s 85% ethanol, 0.85% saline 30s 95% ethanol 30s 100% ethanol 30s	1X PBS 5 min x2 30% ethanol 30s 60% ethanol 30s 80% ethanol 30s 90% ethanol 30s 95% ethanol 30s 100% ethanol 30s
Hybridisation	<u>hybridsation buffer</u> (800ml) 10X in situ salts DEPC dH_20 70µl 100X Denhardts salts tRNA (100mg/ml) 10µl 50% dextran sulfate 200µl <u>probe</u> probe 4 µl formamide 4µl soak towels at bottom of container with 2X SSC, 50% formamide slides with probe and hybridisation buffer O/N at 50°C	<u>hybridsation buffer</u> (800ml) 10X in situ salts DEPC dH_20 64µl 100X Denhardts salts tRNA (100mg/ml) 8µl formamide 320µl 50% dextran sulfate 160µl <u>probe</u> probe 1µl DEPC dH_20 19µl formamide 20µl slides with probe and hybridisation buffer O/N at 55°C

Staining		
1. washing	wash buffer (2X SSC, 50% formamide) 30 min, 50°C wash buffer 1h30 x2, 50°C NTE 5 min x2, 37°C RNAse (20µg/ml in NTE) 30 min, 37°C NTE 5 min x2 wash buffer 1hr, 50°C 1X SSC 2 min 1X PBS 5 min x2	0.2 SSC 1hr x2, 55°C NTE 5 min x2, 37°C RNAse (20µg/ml in NTE) 30 min, 37°C NTE 5 min x2, 37°C 0.2 SSC 1hr, 55°C 1X PBS 5 min
2. antibody staining	100mM Tris, 150mM NaCl, 5 min 0.5% blocking reagent in 100mM Tris, 150mM NaCl, 1 hr 1% BSA in 100mM Tris, 150mM NaCl, 0.3% Triton X-100, 30 min anti-DIG antibody (1:3000) in 1% BSA in 100mM Tris, 150mM NaCl, 0.3% Triton X-100, 1h 30	1% blocking reagent in 100mM Tris, 150mM NaCl, 45 min 1% BSA in 100mM Tris, 150mM NaCl, 0.3% Triton X-100, 45 min anti-DIG antibody (1:1250) in 1% BSA in 100mM Tris, 150mM NaCl, 0.3% Triton X-100, 2hr
3. washing	1% BSA in 100mM Tris, 150mM NaCl, 0.3% Triton X-100, 20 min x4 100mM Tris, 150mM NaCl, 5 min 100mM Tris, 100mM NaCl, 50mM $MgCl_2$, 5 min	1% BSA in 100mM Tris, 150mM NaCl, 0.3% Triton X-100, 15 min x4 100mM Tris, 100mM NaCl, 50mM $MgCl_2$, 10 min
4. substrate application	NBT /BCIP leave in dark 1-3 days	NBT /BCIP leave in dark 1-3 days
5. stop enzyme reation	dH_20 < 5s 70% ethanol < 5s 95% ethanol < 5s 100% ethanol < 5s 95% ethanol < 5s 70% ethanol < 5s dH_20 < 5s	30% ethanol < 5s 50% ethanol < 5s 70% ethanol < 5s 85% ethanol < 5s 95% ethanol < 5s 100% ethanol < 5s 100% histoclear < 5s

Abbreviations and reagents
PBS: Phosphate buffered saline
10X PBS: 1.3M NaCl, 0.03M NaH_2PO_4

SSC: sodium chloride-sodium citrate buffer
20X SSC: 3M NaCl, 0.3 Na_3citrate

10X in situ salts: 3M NaCl, 0.1M Tris-HCl, 0.1M $NaPO_4$, 50mM EDTA

APPENDIX 2: PRIMERS

Table summarising LEGCYC primer sequences and melting temperature (Tm), with a brief description of primer specificity.

Figures show the binding site of each primer (location indicates 5' position on *L. nanus* LEGCYC1A and LEGCYC1B sequences and *C. purpurea* LEGCYC2 sequence). The hatched region in each sequence identifies the position of the intron. ▲ primer amplifying multiple loci, ▲ locus-specific primer, ▲ genome-walking primer. Forward primers are shown above the sequence, and reverse primers below the sequence.

Primer	Sequence (5'-3')	Length	Tm	Note
LEGCYC_F1	TCA GGG SYT GAG GGA CCG	18	61.7	general forward primer in TCP domain, will amplify cyc from legumes from all 3 subfamilies
LEGCYC_R1	TCC CTT GCT CTT GCT CTT GC	20	59.4	general reverse primer in R domain, will amplify cyc from legumes from all 3 subfamilies
LEGCYC_iF1	TCA CCC TCC GGT CCC TCA	18	60.5	inverse primer in TCP domain, used as nested primer in inverse PCR
LEGCYC_iR1	AAA GCA AGA GCA AGA GCA AGG	21	57.9	inverse primer in R domain, used as nested primer in inverse PCR
LEGCYC_F2	GCI MGI AAG TTC TTY GAY CTI CAR GATG	28	63.7	highly degenerate forward primer in TCP domain
LEGCYC_R2	GTY CKY TCC CTS GCY CKY GCT CTY GC	26	71.9	highly degenerate reverse primer in R domain, appears to bind to non cyc genes like atpB and actin
LEGCYC_F4	CTT YGA TCT HCA RGA CAT GYT RGG RTT YGA YAA	33	66.8	highly degenerate forward primer in TCP domain
LEGCYC_F3	CAA GAC ATG YTA GGG TTT GAC	21	56.9	forward primer in TCP domain, designed to amplify both loci in *Cadia* and *Lupinus*
LEGCYC_R3	CAA GCS GGT TCC TTY TGT G	19	57.7	specific reverse primer amplifying Cadia 1 and Lupinus 1, in hypervariable region between TCP and R (close to R)
LEGCYC_R4	CTA CYA CTA CCC CTT CTG G	19	57.7	specific reverse primer amplifying Cadia 2 and Lupinus 2, in hypervariable region between TCP and R (close to R)
LEGCYC_iF3	GTC AAA CCC TAR CAT GTC TTG	21	56.9	inverse primer specific for Cadia 1 and Lupinus 1
LEGCYC_iR3	CAC ARA AGG AAC CWG CTT G	19	55.6	inverse primer specific for Cadia 2 and Lupinus 2
LEGCYC_iR4	CCA GAA GGG GTA GTR GTA G	19	57.7	inverse primer amplifying both loci in *Cadia* and *Lupinus*
LEGCYC_F5	CTT TCY TTA ACC CTG AAA ATG CTT C	25	58.9	forward primer close to start of ORF, amplifying both loci in *Cadia* and *Lupinus*
LEGCYC_R5	YAT TSG CAT CCC AAT TTG GAG	21	56.9	reverse primer at 3' end of ORF, before intron, amplifying both loci in *Cadia* and *Lupinus*
LEGCYC_R6	AGC ARA CAA GAA AGS CCA TAG TG	23	59.8	reverse primer close to beginning of TCP domain, specific for Cadia 1 and Lupin 1
LEGCYC_R7	GGT TTC TTW GYA AGA AAA TTG GAG	24	56.7	reverse primer close to beginning of TCP domain, specific for Cadia 1 and Lupin 1
LEGCYC_R8	CAC TCY TCC CAR GAY TTT CC	20	58.3	reverse primer at 3'end of ORF, spanning putative intron, amplifying both loci for *Cadia* and *Lupinus*
LEGCYC_R9	TTC CAA AGA TTT CAA GCT C	19	50.2	reverse primer at 3' end of LEGCYC2 ORF
LEGCYC_F9	CTT CTA CTT ACA YWT CYT CAG GC	23	58.9	forward primer at start of ORF, amplifying both loci in *Lupinus*
LEGCYC_F10	SAW CRA CAC RTC AAA TGA G	19	52.4	forward primer between the TCP and R domains, specific to Cadia 3
LEGCYC_F12	GAG AAA GTA GCA TCA TTG	18	49.1	forward primer between the TCP and R domains, specific to Lupinus 3
LEGCYCI_GW1	CCT ARC ATG TGT TGW AGA TCR AAG AAC	27	64.0	genome walking primer amplifying 5'-end of *Cadia* and *Lupinus* LEGCYC1
LEGCYC1A_GW2	CMG GTT TGT TWG YAA GAA AAT TGG AG	26	60.6	nested genome walking primer (5'), specific for Cadia 2 and Lupinus 2
LEGCYC1B_GW2	GTC TTG TTT SGG CAT TGW AGC AG	23	60.1	nested genome walking primer (5'), specific for Cadia 1 and Lupinus 1
LEGCYCI_RGW1	GGA ATG CAT TGT GAT MAR GAG AAA RTT GAA GC	32	65.0	genome walking primer amplifying 3'-end of *Cadia* and *Lupinus* LEGCYC1
LEGCYCI_RGW2	CAG CAT GAA TCT MTC WAC AGG TAT	25	60.5	nested genome walking primer (3'), for *Cadia* and *Lupinus* LEGCYC1

LUPINUS NANUS 1
LEGCYC1B
F9
8
F5
77
F4
423
F2
414
F1
369
F3
433
iR3
686
iR1
736
RGW1
980
RGW2
1,188
TCP
R
ATG
0
239
GW2-
1B
298
R6
377
iF1
420
GW1
433
iF3
686
R3
739
R1
743
R2
1,137
R5
1,314
R8
LUPINUS NANUS 2
LEGCYC1A
iR4
583
TCP
R
ATG
0
267
R7
267
GW2-1A
583
R4
CADIA PURPUREA 3
LEGCYC2
F10
195
partial
TCP
R
0
iF1
831
R9

APPENDIX 3: TCP amino acid matrix, with GenBank accession numbers for each sequence when available. The basic helix-loop-helix structure is shown (from Cubas *et al.*, 1999a).

	BASIC	HELIX I	LOOP	HELIX II	GenBank accession no.
Arabidopsis TCP1	KDRHSKIQTAQGIRDRRVRLSIGIARQFFDLQDMLGFDKASKTLDWLLKKSRKAIKEV				AC002130
Arabidopsis TCP2	KDRHSKVLTSKGPRDRRVRLSVSTALQFYDLQDRLGYDQPSKAVEWLIKAAEDSISEL				AL161548
Arabidopsis TCP3	KDRHSKVCTAKGPRDRRVRLSAPTAIQFYDVQDRLGFDRPSKAVDWLITKAKSAIDDL				AF072134
Arabidopsis TCP4	KDRHSKVCTAKGPRDRRVRLSAHTAIQFYDVQDRLGFDRPSKAVDWLIKKAKTSIDEL				AP000370
Arabidopsis TCP5	KDRHSKVCTVRGLRDRRIRLSVPTAIQLYDLQDRLGLSQPSKVIDWLLEAAKDDVDKL				AB008269
Arabidopsis TCP6	KDRHLKV---EG-RGRRVRLPPLCAARIYQLTKELGHKSDGETLEWLLQHAEPSILSA				AB010072
Arabidopsis TCP9	KDRHTKV---EG-RGRRIRMPATCAARIFQLTRELGHKSDGETIRWLLENAEPAIIAA				AF370606
Arabidopsis TCP10	KDRHSKVFTSKGPRDRRVRLSAHTAIQFYDVQDRLGYDRPSKAVDWLIKKAKTAIDKL				AC005311
Arabidopsis TCP11	KDRHTKV---NG-RSRRVTMPALAAARIFQLTRELGHKTEGETIEWLLSQAEPSIIAA				AC006922
Arabidopsis TCP12	RDRHSKICTAQGPRDRRMRLSLQIARKFFDLQDMLGFDKASKTIEWLFSKSKTSIKQL				AC011914
Arabidopsis TCP13	KDRHSKVCTLRGLRDRRVRLSVPTAIQLYDLQERLGVDQPSKAVDWLLDAAKEEIDEL				AB014465
Arabidopsis TCP16	KDRHLKI---GG-RDRRIRIPPSVAPQLFRLTKELGFKTDGETVSWLLQNAEPAIFAA				AL138649
Arabidopsis TCP17	KDRHSKVCTVRGLRDRRIRLSVMTAIQVYDLQERLGLSQPSKVIDWLLEVAKNDVDLL				AL357612
Arabidopsis TCP18	TDRHSKIKTAKGTRDRRMRLSLDVAKELFGLQDMLGFDKASKTVEWLLTQAKPEIIKI				AP001303
Arabidopsis TCP19	KDRHTKV---EG-RGRRIRMPAGCAARVFQLTRELGHKSDGETIRWLLERAEPAIIEA				AB025623
Arabidopsis TCP23	KDRHIKV---DG-RGRRIRMPAICAARVFQLTRELQHKSDGETIEWLLQQAEPAIIAA				AC007887
Arabidopsis TCP24	KDRHSKVLTSKGLRDRRIRLSVATAIQFYDLQDRLGFDQPSKAVEWLINAASDSITDL				AC073506
Rice PCF1	SDRHSKV---AG-RGRRVRIPAMVAARVFQLTRELGHRTDGETIEWLLRQAEPSIIAA				D87260
Rice PCF2	RDRHTKV---EG-RGRRIRMPAACAARIFQLTRELGHKSDGETIRWLLQQSEPAIIAA				D87261
Antirrhinum CYC	KDRHSKIYTSQGPRDRRVRLSIGIARKFFDLQEMLGFDKPSKTLDWLLTKSKTAIKEL				Y16313
Antirrhinum DICH	KDRHSKINRPQGPRDRRVRLSIGIARKFFDLQEMLGFDKPSKTLDWLLTKSKEAIKEL				AF1994665
Linaria LCYC	KDRHSKIYTAQGPRDRRVRLSIGIARKFFDLQEMLGFDKPSKTLDWLLTKSKTAIKEL				AF161252
Maize TB1	KDRHSKICTAGGMRDRRMRLSLDVARKFFALQDMLGFDKASKTVQWLLNTSKSAIQEM				AF340199
Gossypium AUX	KDRHTKV---DG-RGRRIRMPALCAARVFQLTRELGHKYNGETIEWLLQQAEPAVIAA				AF165924
Lupinus albus TCP1	KDRHSKVCTAKGPRDRRVRLSAHTAIQFYDVQDRLGYDRPSKAVDWLIKKAKTAIDQL				AJ426419
Lotus japonicus 1	KDRHSKIYTSQGLRDRRVRLSIEIARKFFDLQDMLGFDKARNTLEWLFNKSKRAIKDF				-
Lotus japonicus 2	KDRHSKIHTSQGLRDRRVRLSIEIARKFFDLQDMLGFDKASNTLEWLFSKSNKAIEEL				-
Cadia 1	KDRHSKIYTSQGLRDRRVRLSIEIARKFFDLQDMLGFDKASNTLEWLFNKSKKAIKDL				AY225825
Cadia 2	KDRHSKIHTSQGLRDRRVRLSIEIARKFFDLQDMLGFDKASNTLEWLFNKSKKAMKEL				AY225826
Cadia 3	????????????????RVRLSSEIARKFFDLQDMLEFDKPSNTLEWLFTKSENAIKEL				AY225827
Cadia 4	????????????????RMRLSLEVAKRFFGLQDILGFDKASKTVEWLLNQAKVEIKQL				AY225828

APPENDIX 4: Aligned nucleotide sequences, including regions between the TCP and R domains, of LEGCYC genes (GenBank accession numbers in table 2-3). Excluded regions are not shown here.

```
Dussia3        CAGGGTGAGATTGTCCAGTGAAATCGCTCGAAAGTTCTTTGATCTTCAGGACATGCTAGAGTATGACAAACCCAGCAATACTCTTGAGTGG
Pisum CYC2     ??????????????????????????????????????????????????????????????????????????????AGCAATACACTTGAGTGG
L.nanus3       GAGGGTGAGGCTTTCAAGTGAAATAGCAAGGAAGTTCTTTGACCTTCAGGACATGCTTGAGTTTGACAAACCTAGCAATACCCTTGAGTGG
Lupinus sp.3   GAGGGTGAGGCTTTCAAGTGAAATAGCAAGGAAGTTCTTTGACCTTCAGGACATGCTTGAGTTTGACAAACCTAGCAATACCCTCGAGTGG
Cadia3         CAGGGTGAGACTGTCAAGTGAAATAGCCCGCAAGTTCTTTGATCTTCAGGACATGCTAGAGTTTGACAAACCTAGCAATACCCTTGAGTGG
Acosmium3      CAGGGTGAGGTTGTCAAGTGAAGTAGCCCGCAAGTTCTTTGATCTTCAGGACATGCTAGAGTTTGACAAACCTAGCAATACCCTTGAGTGG
Clitoria3      CAGGGTGAGGTTATCAAGCGAAATAGCCCGCAAGTTCTTTGATCTTCAGGACATGTTAGAGTTTGACAAACCAAGTAACACCCTTGAGTGG
Lupinussp.4    GAGGGTGAGACTTTCAAGTGACATTGCAAGAAAGTTCTTTGATCTTCAGGAGATGTTGGACTTTGACAAACCTAGCAATACCCTTGAGTGG
Anthyllis3     CCGCGTGAGGCTATCGAGCGAGATAGCGCGCAAGTTCTTTGATCTTCAGGACATGTTGGAGTTTGACAAGCCAAGCAACACACTTGAGTGG
Indigofera3    CAGGGTGAGGTTATCAAGTGAAATAGCTCGCAAGTTCTTTGATCTTCAGGACATGCTTGAGTTTGACAAACCTAGTAACACTCTTGAGTGG
Swartzia3      AAGGGTGAGATTGTCAAACCAAATCGCTAGAAAGTTCTTTGATCTTCAGGACATGCTTGAATTTGACAAACCCAGCAATACCCTTGAGTGG
Acosmium2      GAGGGTAAGATTGTCCATCGACATTGCGCGCAAGTTCTTCGATCTTCAAGACATGTTAGGGTTTGACAAAGCCAGCAACACCCTTGAGTGG
Cadia2         CAGGGTGAGATTGTCCATTGAGATTGCACGCAAGTTCTTTGATCTTCAAGACATGTTAGGGTTTGACAAAGCCAGCAACACCCTTGAGTGG
Lupinus sp.2   ??????????????????????????????????????????????????????????TGTTAGGGTTTGACAAGGCTAGTAACACACTTGAGTGG
L.nanus2       GAGGGTGAGATTATCAATCGAGATCGCGCGAAAGTTCTTCGATCTTCAAGATATGTTAGGGTTTGACAAGGCTAGTAACACACTTGAGTGG
L.ang2         GAGGGTGAGATTGTCGATCGAGATCGCGCGAAAGTTCTTTGATCTTCAAGATATGTTAGGGTTTGACAAGGCCAGTAACACACTTGAGTGG
Machaerium2    AAGGGTGAGGCTCTCTATTGAGATTGCACGCAAGTTCTTTGACCTTCAAGAGATGCTAGGGTTTGACAAGGCCAGCAACACGCTTGAGTGG
Dussia1        CAGAGTAAGGTTGTCCATCGAGATCGCGCGCAAGTTCTTTGATCTTCAAGACATGCTAGGGTTTGGCAAAGCCAGCAACACCCTTGAGTGG
Dussia2        CAGGGTGAGATTGTCCATCGAGATCGCACGCAAGTTCTTTGATCTTCAAGACATGCTAGGGTTTGACAAGGCCAGCAACACCCTTGAGTGG
Acosmium1      TAGGGTGAGGTTGTCGATCGAGATCGCCCGCAAGTTCTTTGATCTACAAGATATGCTAGGGTTTGACAAAGCTAGCAACACCCTCGAGTGG
L.berth2       AAGGGTGAGGCTCTCGATCGAGATCGCGAGAAAGTTCTTTGATCTTCAAGACATGCTAGGGTTTGATAAAGCCAGCAACACCCTCGAGTGG
L.jap2         GAGGGTGAGGCTCTCAATCGAGATCGCAAGAAAGTTCTTTGATCTTCAAGACATGCTGGGGTTTGATAAGGCCCGCAACACCCTCGAGTGG
Anthyllis2     GAGGGTGAGGCTCTCGATCGAGATCGCGCGCAAGTTCTTCGATCTTCAAGACATGCTAGGATTCGACAAGGCCAGCAACACCCTTGAGTGG
Clitoria1      CAGGGTGAGGTTGTCCATTGAGATTGCTCGAAAGTTCTTTGATCTTCAAGACATGTTAGGGTTTGACAAAGCCAGCAACACCCTTGAGTGG
Soya1          AAGGTGGAGGTTGTCCATTGCGATTGCTCGCAAGTTCTTTGATCTTCAAGACATGCTAGGGTTTGACAAAGCCAGTAACACCCTTGAGTGG
Cadia1         CAGGGTGAGGTTGTCCATTGAGATCGCCCGCAAGTTCTTTGATCTACAAGACATGCTAGGGTTTGACAAAGCCAGTAACACTCTTGAGTGG
Lupinus sp.1   GAGGGTGAGGCTTTCGATTGAGATCGCGCGAAAGTTCTTCGATCTACAAGACATGCTAGGGTTTGACAAAGCAAGCAACACCCTTGAGTGG
L.nanus1       GAGGGTGAGGCTTTCGATTGAGATCGCGCGAAAGTTCTTCGATCTACAAGACATGCTAGGGTTTGACAAAGCAAGCAACACCCTTGAGTGG
L.ang1         GAGGGTGAGGCTTTCGATCGAGATCGCACGAAAGTTCTTCGATCTACAAGATATGCTAGGGTTTGACAAAGCAAGCAACACCCTTGAGTGG
Machaerium1    AAGGGTGAGGCTATCCATCGAGATTGCTCGCAGGTTCTTCGATCTCCAGGACATGCTAGGGTTCGACAAGGCCAGCAACACCCTCGACCGG
Medicago1      AAGAGTGAGGCTTTCGATTGAGATCGCTCGAAAGTTCTTCGATCTTCAAGACATGTTAGGGTTTGACAAAGCTAGCAACACACTTGATTGG
Swartzia2      AAGGGTGAGATTGTCAATTGACATAGCGCGCAAGTTCTTTGATCTTCAGGACATGTTAGGGTTCGACAAAGCCAGCAACACCCTCGAGTGG
```

```
Clitoria2     AAGGGTGAGGCTTTCCATAGATATTGCACGCAAGTTCTTTGATCTTCAAGACATGTTAGGGTTTGACAAAGCCAGCAACACCCTTGAATGG
L.berth1      GAGGGTGAGGCTTTCGATCGAGATCGCGCGGAAGTTCTTCGATCTTCAAGACATGTTAGGGTTTGACAAAGCTAGCAACACCCTTGAGTGG
L.jap1        GAGGGTGAGGCTTTCAATCGAGATCGCGCGAAAGTTCTTCGATCTTCAAGACATGTTAGGGTTTGACAAAGCCAGCAATACCCTCGAGTGG
Anthyllis1    GAGGGTGCGGCTCTCGATCGAGATCGCGCGCAAGTTCTTCGATCTCCAGGACATGTTAGGGTTTGACAAGGCCAGCAACACCTTAGAGTGG
Pisum CYC1    ??????????????????????????????????????????????????????ATGTTAGGGTTTGACAAAGCTAGCAACACACTTGAGTGG
Pisum1        GAGGGTGAGACTCTCGATCGAGATAGCGCGGAAGTTCTTCGATCTTCAAGACATGTTAGGGTTTGACAAAGCTAGCAACACACTTGAGTGG

Dussia3       CTCTTCAACAAGTCTGAGAATGCAATTAAAGAACTAGCTCGAAGTAAGCACAACTCGTTGGGT---GCTTCC------------------
Pisum CYC2    CTTTTCAATAAATCTGATACCGCAATCAAAGAACTCGCCAGAACTAAAAAC---TCGTTCGGT---TGTTCG------------------
L.nanus3      CTTTTCGCAAAGTCAGAGAACGCAATCAAAGAACTTGCTAGAAGTAAGAATAGTTCATTGGGTGATGCTTCT------------------
Lupinus sp.3  CTTTTCACAAAGTCAGAGAACGCAATCAAAGAACTTGCTAGAAGTAAGAATAGTTCATTGGGTGATGCTTCT------------------
Cadia3        CTCTTCACCAAGTCTGAGAATGCAATCAAAGAACTGGCTAGAAGTAAGCATAGCTCATTGGGTGATGGTTCT------------------
Acosmium3     CTATTCACAAAGTCGGAGAATGCAATCAAAGAACTAGCCAGAAGCAAGCATAGCTCATTCGGTGATGCTTCT------------------
Clitoria3     CTCTTCACAAAGTCCGAGAATGCAATCAAGGAGCTTGCTCGAAGTAAGCATAGCTCTTTTGGTGATAGTTCC------------------
Lupinussp.4   CTATTCACAAAGTCAGAGAATGCAATCACAGAACTTGCAAGAAGTAAGCATAATCCGTTGGGTGATAGTTCT------------------
Anthyllis3    CTCTTCACAAAGTCTGAGAGTGCAATCAAAGAGCTTGCAAGGAGTAAGAAC---TCATTGGCTGAT---TCA------------------
Indigofera3   CTCTTCACAAAGTCAGAGAATGCAATTAAGGAACTTGCTAGGAGTAAGAACAGTTCATTGGGTGAAGCTTCC------------------
Swartzia3     CTCTTAACAAAATCTGAGAATGCAATTAAAGAACTAGCACGTGCCAAG------TCAATAGCTAGTGCTTCT------------------
Acosmium2     CTCTTCAACAAGTCAAAGAAAGCAATTGAAGAGCTTGCTAGAAGCAAGAACAGT------GGT---GCTGCCAATAGCTTCTCCTCCTCTG
Cadia2        CTCTTCAACAAATCAAAGAAAGCAATGAAAGAGCTAGCTCGAAGCAAGCAAAGT------GGT---GCTGCCAATAGCTTTTCCTCCTCTG
Lupinussp.2   CTATTCAACAAGTCAAAGAAAGCAATGAAAGAATTAGCTAGAAGCAAAATCAGT------GGTGTTGTTGCAAATAGCTTCTCCTCTTCGG
L.nanus2      CTATTCAACAAGTCAAAGAAAGCAATGAAAGAATTAGCTAGAAGCAAAATCAGT------GGTGTTGTTGCAAATAGCTTCTCCTCTTCGG
L.ang2        CTATTCAACAAGTCAAAGAAAGCAATGAAGGAATTAGCTAGAAGCAAAAACAGT------GGTGTTGTTGCAAATAGCTTCTCCTCTTCGG
Machaerium2   CTCCTAACAAAGTCAAAGAGAGCAATTAAGGAGCTTGCAAGGAGCAAGAACAGT------GCT------GCTAATAGCTTCTCTTCCTCTG
Dussia1       CTCTTCACAAAATCTAATAAAGCAATTGAAGAGCTAGCTCGAAGCAAGCACAGC------GGGGTTGCC---AACAGCTCCACCTCCTCTG
Dussia2       CTCTTCACCAAGTCCAAGAAAGCAATCAAAGAGCTAGCTCGAAGCAAGAACAGC------GGCGGTGGC---AAGAGCTTCTCCTCCTCTG
Acosmium1     CTCTTCAACAAGTCCAAGAAAGCAATTAAAGAACTAGCTAGAAGCAAGAACAGC---------GAAGGCGCTAAGAGTTTCTCCTCATCTG
L.berth2      CTCTTCAACAAGTCCAAGAGAGCCATGAAGGATCTCGCTCGGAGCAAAAACAGC------GGTGGTGGTGACAAGAGCTTCTCTTCC---G
L.jap2        CTCTTCAACAAGTCCAAGAGAGCCATCAAGGATTTCGCTCGGAGCAAGAACAGC------GGTGGTGGTGACAAGAGCTTCTCTTCC---G
Anthyllis2    CTCTTCAACAAGTCCAAGAGAGCCATCAAGGATCTCGCCCAGAGCAGCAACAAC------GGAGATGGTGCC---AGCTTCTTCTCA----
Clitoria1     CTCTTCAACAAGTCCAAGAGAGCAATTAAGGAGCTAGCAAGGAGCAAGAACAGC------GAATTAGGAGGCAAGAGCTTCTCTTCTTCAG
Soya1         CTCTTCAACAAGTCCAAGAGAGCAATTAAGGAGCTTGCAAGGAGCAAGCACAGC------GATGAAGGAGCCAAGAGCTTCTCTTCTTCAG
Cadia1        CTCTTCAACAAGTCCAAGAAAGCAATTAAAGATCTAGCCAGAAGCAAGCACAGC------GAAGGTGCC---AAGAGCTTCGCCTCATCTG
Lupinussp.1   CTCTTCAACAAGTCCAAGTGAGCAATTAAGGACCTAGCTAGAAGCAAGAAA---------GAAGGTGATGCTAATAGTTTATCCTCATCTG
L.nanus1      CTCTTCAACAAGTCCAAGAGAGCAATTAAGGACCTAGCTAGAAGCAAGAAA---------GAAGGTGATGCTAATAGTTTATCCTCATCTG
```

```
L.ang1       CTCTTCAACAAATCCAAGAGAGCAATTAAGGAGCTAGCTAGAAGCAAGAAA---------GAAGGTGATGCTAATAGCTTCTCCTCATCTG
Machaerium1 CTCTTCACAAAGTCCAAGAAGGCAATTAAGGAGCTTGCAAGGACCAAGCACAGT------GCCAGCGAAGGTAAGAGCTTCTCCACATCCG
Medicago1   CTTTTCACAAAATCTAAGAAAGCAATTAAGGATCTAACTAAGAGTAAGCAAAGA------GGTGGTGATGCTAAAAGCTTCACATCTTCCA
Swartzia2   CTCTTCAACAAGTCCAAGAAAGCAATCAAAGATCTAACCGCCGCTAGA----------GGTGATGGC---AGGAGCCTCTCTTCTTCTG
Clitoria2   CTCTTCACAAAGTCCAAGAAAGCAATTAAGGAGCTAACTAGAAGCAATAAG---------GTTGTTGAT------AGCTTCTCTTCTTCTG
L.berth1    CTCTTCAGCAAATCAAACAAAGCAATTGAAGAGCTTTTCAGAAGCAAGCACAGT------GCAGGTGCTTGTTATAGCTTCTCCTCTTCCG
L.jap1      CTCTTCAGCAAATCAAACAAAGCAATTGAAGAGCTTTTCAGAAGCAAGCATAGT------GGTGCTTGTGCT---AGCTTCTCCTCTTCCG
Anthyllis1  CTCTTCAGCAAATCAGACAAAGCAATTGAAGAGCTCTTCCAAAGCGAAAACAGT------GGCGGCGGCGGCCATAGCTTCTCCTCTTCCG
Pisum CYC1  CTTTTCAACAAATCAGAAGAAGCAATTGAGGAGTTAACTAGAAGCAAGAAC---------TCGGGTGACGACCATAGCTTCTCCACTTCGA
Pisum1      CTTTTCAACAAATCAAAAGAAGCAATTGAAGAGTTAACTAGAAGCAAGAAC---------TCGGGTGACGACCATAGCTTCTCCACTTCGA


Dussia3     ----------------------------------AAAGAGAGGATGTTGAAATGTGCAGAGAAGGAAAATGTTTGTGTTCAGGCAAAG--
Pisum CYC2  ----------------------------------AAGGGGAGAAAACTGAAATGGACACAGAAAGAA---------------ACAAAG--
L.nanus3    ----------------------------------AAAGGGAGGAAGTTGAAATGTGGACAGAGGGATGATGTTTCTGTTCAGACTAAA--
Lupinussp.3 ----------------------------------AAAGGGAGGAAGTTGAAATGTGGACAGAGGGATGATGTTTCTGTTCAGACTAAA--
Cadia3      ----------------------------------AAAGGGAGGAAGTTGAAATGGGCACAGGGAGAAGATGTTTGTGTTCAGACAAAA--
Acosmium3   ----------------------------------AAAGGGAGGAAGTTGAAATGGGCAGAGAGGGAAGATGTTTGTGTTCAGACAAAA--
Clitoria3   ----------------------------------AAAGGGAGGAAGTTGAAATGGGCACAGAGAGATGATGCTTGTGTTCTAACCAAA--
Lupinussp.4 ----------------------------------AAAGGGAAGAAGTCCAAATGGGCACAGAGGGATGGTATTTGTATTCAGACTAAA--
Anthyllis3  ----------------------------------GGGAGGAGCAAGTTGAAGTGGACACAGAGGGATGATGTTTGTCTGCAGAACAAG--
Indigofera3 ----------------------------------AAAGGAAGGAAGTTGAAGTGGGGACAGAGGGAAGATGTTTGTGTTCAGATCAAG--
Swartzia3   ----------------------------------AAAGGGAGGAAATTGAAATGGGTGCAGAGGGAAGATGTGGGTGTTCAGACCAAA--
Acosmium2   TGGTTTCAGTGCCAGAAGGGGTAGTAGTAGATTCAAAAGAGAGGAAGCTG---------------------------------------
Cadia2      TGGTTTCAGTGCCAGAAGGGGTAGTAGTAGAATCAAAAGAGAGGAAGCTGAAAAGAGCA----------------------------AAGAT
Lupinussp.2 TGGTTTCGATGCCAGAAGGGGTAGTGGTAGATTCAAAAGATAGGAAGCTGAAAAGGGCA----------------------------AAGAT
L.nanus2    TGGTTTCGATGCCAGAAGGGGTAGTGGTAGATTCAAAAGATAGGAAGCTGAAAAGGGCA----------------------------AAGAT
L.ang2      TCGTTTCAATGCCAGAAGGG????????????????????????????????????????????????????????????????????
Machaerium2 TTGATTCAGTACAACAAGGGGTTGTG---GACTCAGAAGAGAGGAGGCTAAGTAGGGCACAGAAGGAA---------TCAAGGGCAAAGAT
Dussia1     TGGTTTCAGGGCCAGACGGGGTT------GATTCAAAAGAGAGGAAGTTGAAAAGGGCACAGAAGGAACCTGCTTGTGTTCGAGCAAAGAT
Dussia2     TGGTTTCAGGGCCAAACGGGTTA------GATTCAAAAGAGAGGAAGTTGAAAAGGGCACAGAAGGAACCTGCTAGTATTCGGGCAAAGAT
Acosmium1   TGGTTTCATGGCCAAACGGGTTA------GATTTAAAAGAGAGGAAGTTGAAGAGGGCAGAGAAGGAACCTCCTGGTGTTCGTGCAAAGAT
L.berth2    TTGTCTCA---TCAAACAGGTTA------GATTCAAAAGAGATGAAGTTGAAACGGGCACAGAAGGAACCTTCTTGTGCTCGTGCAAAGAT
L.jap2      TTGTCTCA---TCAAACAGGTTA------GATTCAAAAGAGCTGAAGTTGAAAAGGGCACAGAAGGAACCTTCTTGTGCTCGTGCAAAGAT
Anthyllis2  --GTTTGTGAATCAAACGGGTTA------GATTCAAAAGAGATTAAGCTGAAAAGGGCACAAAAGGAACCTTCTTGTGCTCGTGCAAAAAT
Clitoria1   TGGTTTCTGAG---AACGGGTTA------GATTCAAGAGAGAGGAAGATGAAAAGGGCACAAAAGGAACCT------------GCAAAGAT
```

```
Soya1        TGGTTTCTGAGCACAACGGGTTG------GATTCAAGG?AGAGGAAGTTGAAGAGGAACAAGAAGGAACCT------------GCAAAGAT
Cadia1       TGGTTTCAGGGCTAAACGGGTTA------AATTCAAAAGAAAGGAAGTTGAAAAGGACACAGAAGGAACCTGCTTGTGTTCGTGCAAAGAT
Lupinussp.1  TTGTTTCCGGG-----------------GATTCAAAAGATATGAAGTTGAAAAGGGCACAGAAGGAACCAGCTTGTGTAAGAGCAAAGAT
L.nanus1     TTGTTTCCGGG-----------------GATTCAAAAGATATGAAGTTGAAAAGGGCACAAAAGGAACCAGCTTGTGTAAGAGCAAAGAT
L.ang1       TTGTTTCAGGG-----------------GATTCAAAAGATATGAAGTTGAAAAGGGCACAAAAGGAATCCG???????????????????
Machaerium1  TTGTTTCAGGGCCACAAGGGTTGTTG---GATTCAAAAGAAAAGAAGCTGAAGAGGGCACAGAAGGAAGCTAGTACTGCAAGGGCGAAGAT
Medicago1    TTGCT------TCAAAC-----------GGTGCAGAA?AGAA?AAGTTGAAAAGA??????????????????????????????????
Swartzia2    TGGTT------TCAAACGGGTTA------AATTCAAAGGAGAGA---TTGAAAAGGGCACAAAAGGAACCTGATTCTGATAGGGCAAAGAT
Clitoria2    TGGTT------CAACAA------ATGGTGGATTTGGAAGAG---AAGTTAAAAGAA-----------CCAGCTTTTGGTAAGGCAAAGAT
L.berth1     TTGTTTCTGTG-----------------------------------------AAAAGGGCACAGAAAGAACCTTCCTGTGTTCAGGCAAAGAT
L.jap1       TTGTTTCTGTG-----------------------------------------AAAAGGGCACAGAAAGAACCTTCTGGTGTTCAAGCAAAGAT
Anthyllis1   TTGTATCTGTG-----------------------------------------AAAAGGGCACAGAAAGAGCCTTCTAACGTTCAGGTAAAGAT
Pisum CYC1   TGTTTTGAGCA-----------------------------------------------CAGAAGGAATCCTCA-----------AAGAT
Pisum1       TGAAAAGAGCA-----------------------------------------------CAGAAGGAATCCTCA-----------AAGAT


Dussia3      -AAGGAGTCAAGGGAAAAA
Pisum CYC2   -AAAGAGTCAAGAGAAAGA
L.nanus3     -AAAGAGTCAAGGGAAAAG
Lupinussp.3  -AAAGAGTCAAGGGAAAAG
Cadia3       -AAGGAGTCACGGGAAAAG
Acosmium3    -AAGGAGTCAAGGGAAAAG
Clitoria3    -AAGGAGTCAAGGGAAAGG
Lupinussp.4  -AAGGAGTCAAGGGAAAAA
Anthyllis3   -AAGGAGTCAAGGGAAAGG
Indigofera3  -AAGGAGTCAAGAGAAAGG
Swartzia3    -AAGGAGTCAAGGGAAAAG
Acosmium2    -AAAAGGGAAAGGGAAAAG
Cadia2       GAAGGAATCAAGGGAAAAA
Lupinussp.2  TAAGGAATCAAGGGAAAAA
L.nanus2     TAAGGAATCAAGGGAAAAA
L.ang2       ????????????????????
Machaerium2  GAAGGAATCAAGGGAAAAA
Dussia1      GAAGGAATCAAGGGAAAAA
Dussia2      GAAGGAGTCAAGGGAAAAA
Acosmium1    GAAGGAGTCAAGAGAAAAA
```

```
L.berth2     AAAGGAGTCAAGGGAGAAA
L.jap2       AAAGGAATCAAGGGAGAAA
Anthyllis2   GAAGGAGTCAAGGGAGAAA
Clitoria1    GAAGGAGTCAAGGGAAAAA
Soya1        AAAGGAGTCAAGGGGAAAA
Cadia1       GAAGGAGTCCAGAGAAAAA
Lupinussp.1  GAAAGAGTCAAGGGAAAAA
L.nanus1     GAAAGAGTCAAGGGAAAAA
L.ang1       ????????????????????
Machaerium1  GAAGGAGTCAAGGGAGAAA
Medicago1    ????????????????????
Swartzia2    GAAGGAGTCAAGGGAGAAA
Clitoria2    AAAGGAAACAAGGGAAAAA
L.berth1     GAAGGAATCAAGGGAGAAA
L.jap1       GAAGGAATCAAGGGAAAAA
Anthyllis1   GAAGGAATCTAGGGAGAAA
Pisum CYC1   GAAAGACTCAAGAGAAAAA
Pisum1       GAAAGACTCAAGAGAAAAA
```

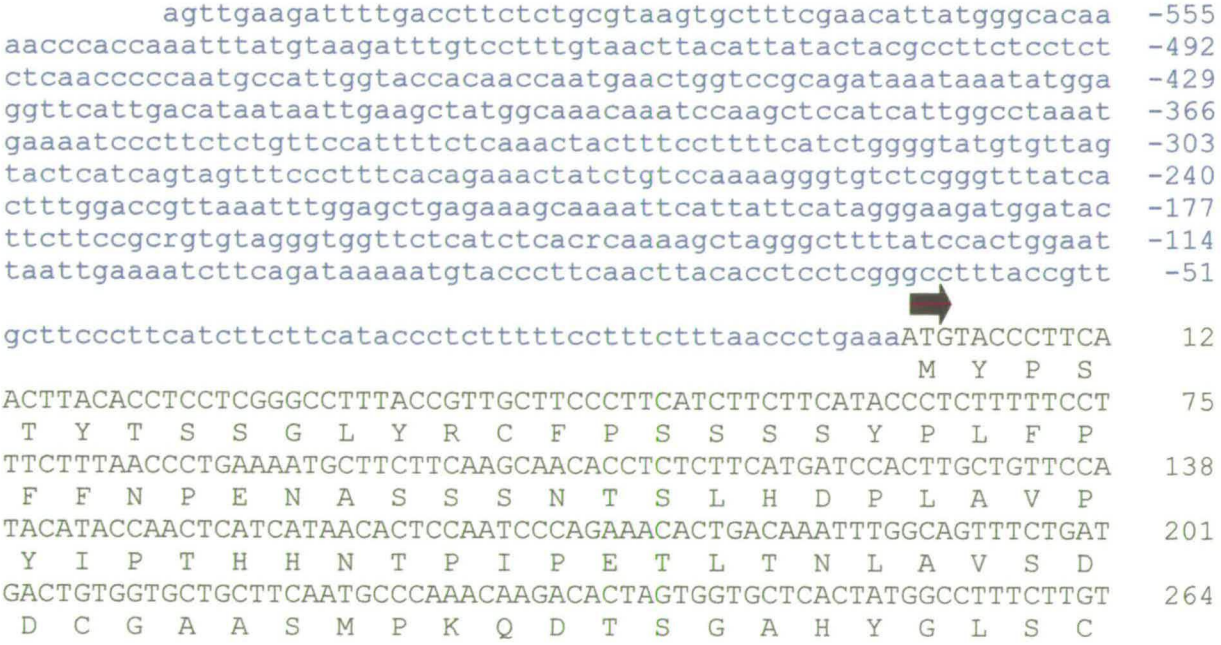
APPENDIX 5

Genomic sequence of *Cadia purpurea* and *Lupinus nanus* LEGCYC1A and LEGCYC1B. Coding sequence is shown in black (start codon: ➡, stop condon: •) , with predicted amino acid translation below; TCP and R domains are underlined; upstream and downstream regions (untranslated) of the open reading frame are shown in blue, the intron (predicted splice sites are marked by ▲) is shown in red.

Lupinus nanus 1 (LEGCYC1B)

```
gttactggcactattacttctacacctttctctctttaaaccccactccatttaacaattgaa  -387
cctggtcctcagataaataaatatggaggttcatagttcattcattttcacataataattgaa  -324
actatgcaaagttccatcattgttgctaaaatgaaatcccttcctcttatcatttttttcccaa  -261
acacacactttccttttcttatgggatagtgttattattattagtagtactaatcagtaacat  -198
agtttcactttcacagaaactatttgtataaaagggtgtcttgggtttatcactatggaccgt  -135
gtaagttgaagttgaagaaaggaagagttctttattcaaagggaagatctgatttgaagggtg   -72
ttccaattcatatttcacataaacaaaagctagggtttttatccactagaatcaattgaaaat    -9

cttcatatcATGTACCCTTCTACTTACACTTCTTCAGGCCCTTATTCTTGTTACTCTTCAGCT    54
          M  Y  P  S  T  Y  T  S  S  G  P  Y  S  C  Y  S  S  A
TCGAATTCATACCCTTTTTTCCCTTTTCTTAACCCTGAAAATGCTTCTTCAAGCAACAACAAC   117
 S  N  S  Y  P  F  F  P  F  L  N  P  E  N  A  S  S  S  N  N  N
AACAACCATAACCTTCTTCATGATCCACTTGTTCATGTTCCTTACAACTTACCAAGTCATCAT   180
 N  N  H  N  L  L  H  D  P  L  V  H  V  P  Y  N  L  P  S  H  H
CATATTCATAACACACCTATAATCCAAGAAACACTGACCAATTTGGCTGTTTCTGATGCTGCT   243
 H  I  H  N  T  P  I  I  Q  E  T  L  T  N  L  A  V  S  D  A  A
ACAATGCCGAAACAAGACCCTATTATGAGTGGTGGTGGTGGTGGTGTTCATCATCACTATGGG   306
 T  M  P  K  Q  D  P  I  M  S  G  G  G  G  G  V  H  H  H  Y  G
CTTTCTTCTCTGCTCACAAAGAAACCAGCCAAAAAGGATAGGCACAGCAAGATTTACACCTCT   369
 L  S  S  L  L  T  K  K  P  A  K  K  D  R  H  S  K  I  Y  T  S
CAGGGCTTGAGGGATCGGAGGGTGAGGCTTTCGATTGAGATCGCGCGAAAGTTCTTCGATCTA   432
 Q  G  L  R  D  R  R  V  R  L  S  I  E  I  A  R  K  F  F  D  L
CAAGACATGCTAGGGTTTGACAAAGCAAGCAACACCCTTGAGTGGCTCTTCAACAAGTCCAAG   495
 Q  D  M  L  G  F  D  K  A  S  N  T  L  E  W  L  F  N  K  S  K
AGAGCAATTAAGGACCTAGCTAGAAGCAAGAAAAACAATGGTAGTGAAGGTGATGCTAATAGT   558
 R  A  I  K  D  L  A  R  S  K  K  N  N  G  S  E  G  D  A  N  S
TTATCCTCCTCTTCGGATCGCGAGGAATGTAATGAAGTTGTTTCCGGGATCAATAATGAACAA   621
 L  S  X  S  S  D  R  E  E  C  N  E  V  V  S  G  I  N  N  E  Q
CAAGGTATCACCATTGCTGATCATGATTCAAATGGTGTGAAAGATATGAAGAAGTTGAAAAGG   684
 Q  G  I  T  I  A  D  H  D  S  N  G  V  K  D  M  K  K  L  K  R
GCACAAAAGGAACCAGCTTGTGTAAGAGCAAAGATGAAAGAGTCCAGGGAAAAAGCAAGAGCA   747
 A  Q  K  E  P  A  C  V  R  A  K  M  K  E  S  R  E  K  A  R  A
AGAGCAAGAGAAAGAACTAGTAACAAGATGTGTAACAATAACAATGGAAGGGTAGTTCAAGTG   810
 R  A  R  E  R  T  S  N  K  M  C  N  N  N  N  G  R  V  V  Q  V
```

Cadia 1 (LEGCYC1B)




TTGCTTACAAAGAAACCAGCCAAGAAAGATAGGCACAGCAAGATTTACACCTCCCAGGGCTTG 327
L L T K K P A K K D R H S K I Y T S Q G L
AGGGACCGCAGGGTGAGGTTGTCCATTGAGATCGCCCGCAAGTTCTTTGATCTACAAGACATG 390
R D R R V R L S I E I A R K F F D L Q D M
CTAGGGTTTGACAAAGCCAGTAACACTCTTGAGTGGCTCTTCAACAAGTCCAAGAAAGCAATT 453
L G F D K A S N T L E W L F N K S K K A I
AAAGATCTAGCCAGAAGCAAGCACAGCAACAGTGAAGGTGCCAAGAGCTTCGCCTCATCTTCT 516
K D L A R S K H S N S E G A K S F A S S S
GACTGTGAGGACTGGGAAGTGGTTTCAGGGATCAATGAAACTGATACTCTAAACCTAAAACAA 579
D C E D W E V V S G I N E T D T L N L K Q
GGGTTAAATTCAAATGACAATAAGTTATTGATGGGTAATGGTGGTGGTGGTGGTTCAGATGCT 642
G L N S N D N K L L M G N G G G G G S D A
GTGAAAGAAAGGAAGTTGAAAAGGACACAGAAGGAACCTGCTTGTGTTCGTGCAAAGATGAAG 705
V K E R K L K R T Q K E P A C V R A K M K
GAGTCCAGAGAAAAAGCAAGAGCAAGAGCAAGGGAAAGGACTAGTAACAAGATGTGCAACAGT 768
E S R E K A R A R A R E R T S N K M C N S
AACACCACAAGTAATGGGAGGGTGCAAGTGCAAGACTTGAAGAAAAAGATCCTTGCAACTGAA 831
N T T S N G R V Q V Q D L K K K I L A T E
AACCCTCAAACTCTGCACCAATTTAGGTCACCCCTTCAGCCTGAGGACTGTGCAAGATCACCT 894
N P Q T L H Q F R S P L Q P E D C A R S P
AATAAGCTGTTTCACCCTATACCTCATCACCTTGTGGGTAGTGAAGCACCTAGAGATGACTTC 957
N K L F H P I P H H L V G S E A P R D D F
AACGTGATTGAGGAATCCATTTTGATAAGGAGAAAGTTGAAGCCAACGTTGATGTCTTCTCAT 1020
N V I E E S I L I R R K L K P T L M S S H
CATCATCACCAAAAACTTGTGATCCCAAAGGAAGCTAGTTTCAACAGCAATGACTACCACTCC 1083
H H H Q K L V I P K E A S F N S N D Y H S
TTCCCCAATTTGTCTCCAAATTGGGATGCTAATAATGGTACCAATGCCACTGGCCGCGCCAAC 1146
F P N L S P N W D A N N G T N A T G R A N
TTTTGTACAATAGCCAGCATGAATCTATCTACAGgtatgtttcatgtttgtgtttcatgacaa 1209
F C T I A S M N L S T
gctgatagtgctaagtgctcttcttgaaacctaccattttgatgatatttacagttctaagtc 1272
ttaattttcagGGCTTCAAATCTTTGGAAAGTCTTGGGAGTAGtgcaccaatccaagtctaca 1335
G L Q I F G K S W E •
ctagtatgttagctttcagtattatctgatccgaatganctctctagtgctttgccaaggaat 1398
catmaaggcatctttctgtgttttccaccagtaacttttctgtcctatattccctytcgacaa 1461
tgtttgtacctgatgttttgctcatgatcagccaatggcgtgtgatagttggcacaaaggttg 1524
ctgcgtgtattatttctgagttctgaacaagatttgaagtgtggttggcattatataatgcca 1587
attagttatcaagaactgttcctttctagcagcctttaatatttatatattyggttaagtaat 1650
gttcaacagtaactaatatatgccatattcgaaaacatttcaagcagttaaataccttggctg 1713
gtaagagagggtggtacggaagaaattaagtcttcagatttgtttgc 1759

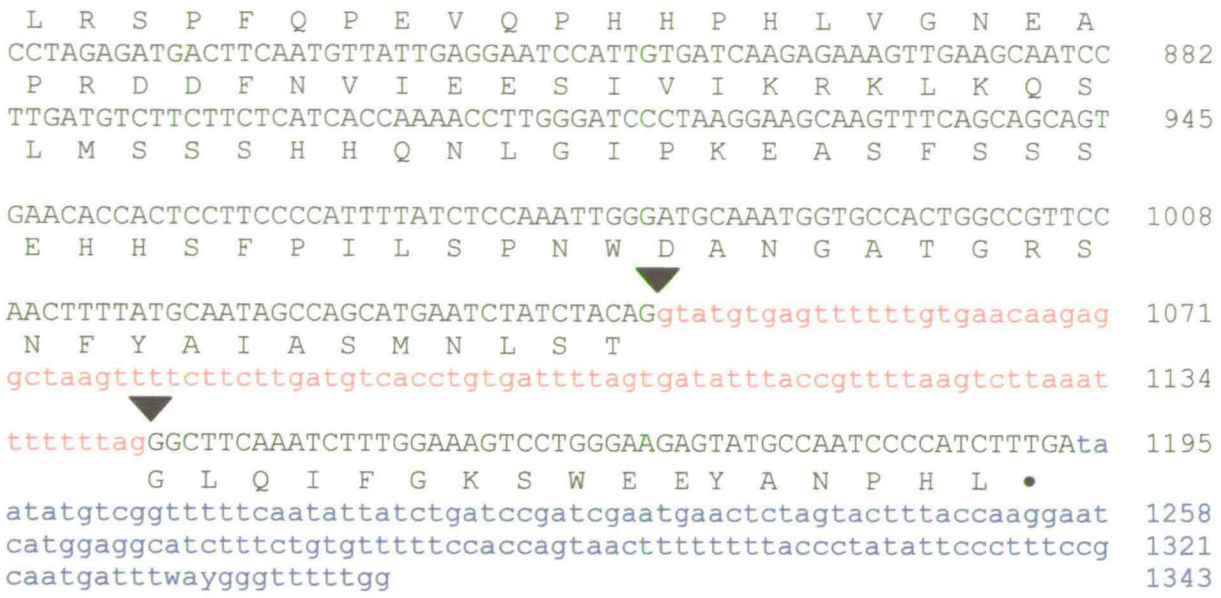
Lupinus nanus 2 (LEGCYC1A)

```
                atcttttaatcatttcagtacccttgggtcaacaacatgaatcaaa -1236
tctgtgtcaagttaatttcttctgcaaaatgagaccagacccccaccttaggttatagcaaca -1173
aaatttcacatgtattgatattaatattaattaatacatcatgtactttaagctacttttttat -1110
tggggctagagaacctacttttatttttttaaaacttcatttccttagaattctatgcaaagt -1047
aagaatagccctaaccgtatcacgctcatgtacaaaaggatgtattattaagtattaaccatc  -984
ttcaatgaatgaagcacacaatatatcaattatgcattattattactttcaaaattatgcaca  -921
aatattttaattttcagagattatttttgaataatttttattataccttaatcttttgatgtt  -858
ttcttaaaattaatactcacttttaaaatagagataaccaaagtgaaaacagtttctaccaat  -795
taattaaaaattttctcgtagacrtaaaaaaaattataatttaaagaaatcataacccccaa  -732
atttgttatcgatataaaaaacaagtcaaaaactatatcatcacaaatatccttttggtacct  -669
ggaacactgttttcaccctacttttatarccccttatggaaaagttycttatttttttggata  -516
aattagattaaaaaatataattggttattaggtaattcttctataactctctctctttcatct  -453
cctccaaaaaaattatagagtgtacataaatatgaaggtctatagattcaataatggaaagtg  -390
tgaaagcaaaactcatttccatcattggcctaaatgaaatcaaccctctcatcactttctcaa  -327
accactttccttattgtacttactagttcccttccacacacaaagagatttctataaaaagaa  -264
atctagttcattgttcatagtaagatatagatagattcctcttcattcttcatcactcaaaaa  -201
aaaagctagggcttttagyccataatcttcaaatgttcccttctacttacatatcctcaggcc  -138
cttacccttatttctcttcttcttcytcaccataccatccttttgctttctttaaccctgaaa   -75
attcttcttcaaacaacacctttttctcatgatctactttctttttccctataacatacaaccta   -12

ctcatcattatcATGTTCCCTTCTACTTACATATCCTCAGGCCCTTACCCTTATTTCTCTTCT    51
            M  F  P  S  T  Y  I  S  S  G  P  Y  P  Y  F  S  S
TCTTCTTCACCATACCATCCTTTTGCTTTCTTTAACCCTGAAAATTCTTCTTCAAACAACACC   114
 S  S  S  P  Y  H  P  F  A  F  F  N  P  E  N  S  S  S  N  N  T
TTTTCTCATGATCTACTTTCTTTTCCCTATAACATACAACCTACTCATCATTATCATGCTCCA   167
 F  S  H  D  L  L  S  F  P  Y  N  I  Q  P  T  H  H  Y  H  A  P
ACACAAGAAACTCTTTCCAATTTTGCAGATTATGCTGCTTCAGCTGCAATGTTTAAGACTGAT   230
 T  Q  E  T  L  S  N  F  A  D  Y  A  A  S  A  A  M  F  K  T  D
GTTAGTGGTAATTCCAATTTTGGTTTCTCCAATTTTCTTGCTAAGAAACCTGCTTCTAAGAAA   293
 V  S  G  N  S  N  F  G  F  S  N  F  L  A  K  K  P  A  S  K  K
GACAGGCATAGCAAGATCCACACATCACAGGGTTTGAGAGATAGGAGGGTGAGATTATCAATC   356
 D  R  H  S  K  I  H  T  S  Q  G  L  R  D  R  R  V  R  L  S  I
GAGATCGCGCGAAAGTTCTTCGATCTTCAAGATATGTTAGGGTTTGACAAGGCTAGTAACACA   419
 E  I  A  R  K  F  F  D  L  Q  D  M  L  G  F  D  K  A  S  N  T
CTTGAGTGGCTATTCAACAAGTCAAAGAAAGCAATGAAAGAATTAGCTAGAAGCAAAATCAGT   482
 L  E  W  L  F  N  K  S  K  K  A  M  K  E  L  A  R  S  K  I  S
AGCAGTGGTGTTGTTGCAAATAGCTTCTCCTCTTCGGATTCGGAGTTTGAAGTGGTTTCGATG   545
 S  S  G  V  V  A  N  S  F  S  S  S  D  S  E  F  E  V  V  S  M
ATAAACCCAGATTCAATTGATGCTACTCCAGAAGGGGTAGTGGTAGATTCAAAAGATAGGAAG   608
 I  N  P  D  S  I  D  A  T  P  E  G  V  V  V  D  S  K  D  R  K
CTGAAAAGGGCAAAGATTAAGGAATCAAGGGAAAAAGCTAGAGCTAGAGCAAGGGAAAGGACT   671
 L  K  R  A  K  I  K  E  S  R  E  K  A  R  A  R  A  R  E  R  T
AATAAAAAGATGTTAAGTAGCATGAAGAAAAAGTATCCTGCAATTGAAAACCCTCAAATGTTT   734
 N  K  K  M  L  S  S  M  K  K  K  Y  P  A  I  E  N  P  Q  M  F
AACATATTGAGGCTACCTTTTCATCATCCTGAGAATTTGGCGAAATCGCCTAATAATAAGTCG   797
 N  I  L  R  L  P  F  H  H  P  E  N  L  A  K  S  P  N  N  K  S
```

```
ATTCTATCTCATCATCATAACCCTCATCTTGTGTGTAGTGAAACTCCTAGAGATGATTTCAAT   860
 I  L  S  H  H  H  N  P  H  L  V  C  S  E  T  P  R  D  D  F  N
CTTTTTGAGGAGTCCATTGTGATCAAGAGAAAATTGAAGCAAAGCCATGCTATCCCTAAGGAA   923
 L  F  E  E  S  I  V  I  K  R  K  L  K  Q  S  H  A  I  P  K  E
TCAAATTTCAATAACAATACTGAACACCACTCCTTTCCCATTTTATCTCCAAATTTGGATGCT   986
 S  N  F  N  N  N  T  E  H  H  S  F  P  I  L  S  P  N  L  D  A

AATAATGGTGCCAATGGCAGATCCAATTTTTGTGCAGTTACCAACATGAATCTATCAACAGgt   1049
 N  N  G  A  N  G  R  S  N  F  C  A  V  T  N  M  N  L  S  T
atgtgaataatttttcataaacaaagagtttaacttaattttgatttttttgtgatatttact   1112

ttttaattttttttagGGCTTCAAATCTTTGGAAAGTCTTGGGAGGAGT                 1160
                 G  L  Q  I  F  G  K  S  W  E  E
```

Cadia 2 (LEGCYC1A)

```
                                                          tagcggccgc   -315
ggattcgcccttaaaaagggctcgagcggccgcccgggcaggacaatcatggaaagtgtgaag   -252
ccatcccagttccatcattggcctaaatgaaatcctctctctcatctcagtttctcaaaccac   -189
tttctttttgagttataggacttagtaactagtccacttccaactgaaaagatttgtataaaa   -126
aggtacctttcagagctgaggagatagataccttagcagtgtgtggtgtggtcttcaatcctc    -63
atcccacaacagctatttttttttttccaactgaaattaattaattaattccaaaatttgcag      0

ATGTTCCCTTCAACTTACAGCTCCTCAGGCCCTTATCCGTACCTCCCTTCATCTTCTTCATCA     63
 M  F  P  S  T  Y  S  S  S  G  P  Y  P  Y  L  P  S  S  S  S  S
TACCATCCTTTTACTTTCCTTAACCCTGAAAATGCTTCTGCAAACAACACCTTTTCCCATGAT   126
 Y  H  P  F  T  F  L  N  P  E  N  A  S  A  N  N  T  F  S  H  D
CCACTTTGTGTTCCCTACATACCTTCTACTCATCATGGTCCAGTCCCAGAAACACTAACCAAT   189
 P  L  C  V  P  Y  I  P  S  T  H  H  G  P  V  P  E  T  L  T  N
TTGGCAGTTGCAGACTGTTCTGCAGCAGCTGCAATGTTCAAAAACGATGTCAGTGGTGTTAAT   252
 L  A  V  A  D  C  S  A  A  A  A  M  F  K  N  D  V  S  G  V  N
TATGGCTTCTCCAATTTTCTTACAAAGAAACCGCCTGCAAAAAAAGATAGACACAGTAAGATT   315
 Y  G  F  S  N  F  L  T  K  K  P  P  A  K  K  D  R  H  S  K  I
CACACATCTCAGGGTTTGAGGGACCGCAGGGTGAGATTGTCCATTGAGATTGCACGCAAGTTC   378
 H  T  S  Q  G  L  R  D  R  R  V  R  L  S  I  E  I  A  R  K  F
TTTGATCTTCAAGACATGTTAGGGTTTGACAAAGCCAGCAACACCCTTGAGTGGCTCTTCAAC   441
 F  D  L  Q  D  M  L  G  F  D  K  A  S  N  T  L  E  W  L  F  N
AAATCAAAGAAAGCAATGAAAGAGCTAGCTCGAAGCAAGCAAAGTAGCAGTGGTGCTGCCAAT   504
 K  S  K  K  A  M  K  E  L  A  R  S  K  Q  S  S  S  G  A  A  N
AGCTTTTCCTCCTCTACGGAGTGTGAAGTGGTTTCAGTGATCAACCAACACCTCACTGATCCA   567
 S  F  S  S  S  T  E  C  E  V  V  S  V  I  N  Q  H  L  T  D  P
GAAGGGGTAGTAGTAGAATCAAAAGAGAGGAAGCTGAAAAGAGCAAAGATGAAGGAATCAAGG   630
 E  G  V  V  V  E  S  K  E  R  K  L  K  R  A  K  M  K  E  S  R
GAAAAAGCAAGGGCAAGAGCAAGGGAAACGCCTAGTAACAAAATGAGCAACACAAGTGGCACT   693
 E  K  A  R  A  R  A  R  E  T  P  S  N  K  M  S  N  T  S  G  T
GGAAAAGTGCAAGACTTGAAGAAAAAGTGCCCTGTAACTGAAAACCCTCAAATCCAGCACCAA   756
 G  K  V  Q  D  L  K  K  K  C  P  V  T  E  N  P  Q  I  Q  H  Q
TTGAGATCACCCTTTCAGCCTGAGGTTCAACCTCATCACCCTCACCTTGTTGGTAATGAAGCG   819
```


L R S P F Q P E V Q P H H P H L V G N E A
CCTAGAGATGACTTCAATGTTATTGAGGAATCCATTGTGATCAAGAGAAAGTTGAAGCAATCC 882
P R D D F N V I E E S I V I K R K L K Q S
TTGATGTCTTCTTCTCATCACCAAAACCTTGGGATCCCTAAGGAAGCAAGTTTCAGCAGCAGT 945
L M S S S H H Q N L G I P K E A S F S S S
GAACACCACTCCTTCCCCATTTTATCTCCAAATTGGGATGCAAATGGTGCCACTGGCCGTTCC 1008
E H H S F P I L S P N W D A N G A T G R S
AACTTTTATGCAATAGCCAGCATGAATCTATCTACAGgtatgtgagtttttgtgaacaagag 1071
N F Y A I A S M N L S T
gctaagttttcttcttgatgtcacctgtgattttagtgatatttaccgttttaagtcttaaat 1134
ttttttagGGCTTCAAATCTTTGGAAAGTCCTGGGAAGAGTATGCCAATCCCCATCTTTGAta 1195
G L Q I F G K S W E E Y A N P H L •
atatgtcggtttttcaatattatctgatccgatcgaatgaactctagtactttaccaaggaat 1258
catggaggcatctttctgtgtttttccaccagtaacttttttttaccctatattccctttccg 1321
caatgatttwaygggtttttgg 1343

APPENDIX 6

Partial genomic sequences of *Cadia purpurea* LEGCYC2 and *Lupinus nanus* LEGCYC1A*, predicted intron region (Hebsgaard *et al.*, 1996) for each locus highlighted in red. The predicted amino acid translation is given below, with the partial TCP domain and R domain underlined. In addition, nucleotide sequences of *C. purpurea* and *L. nanus* actin homologues and *C. purpurea* histone 4 homologue are given.

Cadia purpurea LEGCYC2, genomic DNA, partial codons

```
CAGGGTGAGACTGTCAAGTGAAATAGCCCGCAAGTTCTTTGATCTTCAGGACATGCTAGAGTT   63
  R  V  R  L  S  S  E  I  A  R  K  F  F  D  L  Q  D  M  L  E  F
TGACAAACCTAGCAATACCCTTGAGTGGCTCTTCACCAAGTCTGAGAATGCAATCAAAGAACT  126
  D  K  P  S  N  T  L  E  W  L  F  T  K  S  E  N  A  I  K  E  L
GGCTAGAAGTAAGCATAGCAGCTGCAACTGCAATGAGGGTGACAAGTGCTCCTGTGACCAGCC  189
  A  R  S  K  H  S  S  C  N  C  N  E  G  D  K  C  S  C  D  Q  P
ACATGAGGTAGACACATCAAATGAGAAATCATTGGCAGGCAGTGGTGGTGATGGTTCTAAAGG  252
  H  E  V  D  T  S  N  E  K  S  L  A  G  S  G  G  D  G  S  K  G
GAGGAAGTTGAAATGGGCACAGGGAGAAGATGTTTGTGTTCAGACAAAAAAGGAGTCACGGGA  315
  R  K  L  K  W  A  Q  G  E  D  V  C  V  Q  T  K  K  E  S  R  E
AAAGGCAAGAGCAAGAGCAAGAGAAAGGACTTGTTACAAGATGTGCAACACTGGGAGGGTGCA  378
  K  A  R  A  R  A  R  E  R  T  C  Y  K  M  C  N  T  G  R  V  Q
AGACTTGGAGAAGTGCCCTGCAACTGCAAACCCTCAAATACTGCACCAATTGAGGTCATCCAT  441
  D  L  E  K  C  P  A  T  A  N  P  Q  I  L  H  Q  L  R  S  S  I
TCAGCCTGAGCATGAGGTTTGTGCAAGATGGCCTCATCGGATGGGTCAACCTTACCCTTACCC  504
  Q  P  E  H  E  V  C  A  R  W  P  H  R  M  G  Q  P  Y  P  Y  P
TCACCAAGGTAGTGAAGCACCCAGAGAAGGCTTTAATGTCATTGAGGAATCTATTATGATAAA  567
  H  Q  G  S  E  A  P  R  E  G  F  N  V  I  E  E  S  I  M  I  K
AAGGAGTATGAAGCCATCTTTGATGTCTTCTTCTCATAGCCAAGACATGGTGATCCCTAAGGA  630
  R  S  M  K  P  S  L  M  S  S  S  H  S  Q  D  M  V  I  P  K  E
AGCAAGTTTCAACAACAATGACTACCATTCATTCCCCTATTCCACTCCAAATTGGGATACTAA  693
  A  S  F  N  N  N  D  Y  H  S  F  P  Y  S  T  P  N  W  D  T  N
TGGGAACTCGAACTTTTGTGGAATAGCCACCATGAATCTATCTAAATTTTTCGTGAACCAGTT  756
  G  N  S  N  F  C  G  I  A  T  M  N  L  S  K  F  F  V  N  Q  L
Ggtaagtattcttctcaaatcacttgaggtttttttaaacttttttaaagaaatttagtgatttg  819
ggctcctgatttgtagaGCTTCAAATCTTT                                     849
                  L  Q  I  F
```

Lupinus nanus LEGCYC1A*, genomic DNA, partial codons

```
AAAGCTAGCAAAACTCTTGAGTGGCTCTTCAACAAGTCCAAGAAAGCAATGAAGGACCTTGCT   63
 K  A  S  K  T  L  E  W  L  F  N  K  S  K  K  A  M  K  D  L  A
AGAAGCAACCATCACAGTAGCAATGGTTTTGCCAATAGCTTCTCCTCCTCTTCTTCTTCTCT  126
 R  S  N  H  H  S  S  N  G  F  A  N  S  F  S  S  S  S  S  S  S
TCTTCTTCAGATTCGGAGCGTGAAGTGGTTTCAATTATCAAACAAGATGCCACTAATCCACAA  189
 S  S  S  D  S  E  R  E  V  V  S  I  I  K  Q  D  A  T  N  P  Q
```

```
GTGGTAGTTTTAGATTCAAAAGAAAGGAAGGTGAAAAGGGCAAGGATGAAGGAATCAAGGGAA   252
 V  V  V  L  D  S  K  E  R  K  V  K  R  A  R  M  K  E  S  R  E
AAAGCAAGGGCAAGAGCTAGAGAAAGGACTAGTAACAAGATGTGCAAAAAAAAGTGTCCTATA   315
 K  A  R  A  R  A  R  E  R  T  S  N  K  M  C  K  K  K  C  P  I
ACTGATAACCCTCAAATGCTGCATCAATTAAGGTCACCCTTTGGTCATCCCGAGGATTCAGCA   378
 T  D  N  P  Q  M  L  H  Q  L  R  S  P  F  G  H  P  E  D  S  A
AGATCACCTGATAATAGGTCGATTCCATCTCATCATCACCATCACCAGCACCGTCATCTTACG   441
 R  S  P  D  N  R  S  I  P  S  H  H  H  H  H  Q  H  R  H  L  T
GGTAACCAAGTTGCTCGAGATGACTTCAACGTCATCGAAGAGTCCATTGTGATCAAGCGAAAA   504
 G  N  Q  V  A  R  D  D  F  N  V  I  E  E  S  I  V  I  K  R  K
ATGAAGCAATCAATGTTATCCTCTTCTCATCATCATCAAAACCATATGATCCCTAAGGAAGCA   567
 M  K  Q  S  M  L  S  S  S  H  H  H  Q  N  H  M  I  P  K  E  A
AGTTCCAACATCAACACTGAACACCATTCCTTCCCAATTTTATCTCCAAATTGGGATGCTAAT   630
 S  S  N  I  N  T  E  H  H  S  F  P  I  L  S  P  N  W  D  A  N
AATAATGGTGCCACAAGCCGTACCAACTTTTGTGCTGgtatgtgaaattttcatgaacaagtt   693
 N  N  G  A  T  S  R  T  N  F  C  A
aaggaactaagttttcattttaattatcaatcaaatgtggaatcacctttgatttttttgttat   756
atttatctgaatttttttagGGCTTCAAATCTTT                                 790
                     G  L  Q  I  F
```

Cadia purpurea ACTIN, cDNA, partial codons

```
TGTTTCCTAGCATTGTTGGTCGTCCACGTCACACTGGTGTGATGGTTGGCATGGGyCAAAArG
ATGCATATGTTGGkGATGAAGCTCAGTCCAAGmGwGGTATmyTrACTCTGAAATATCCCATTG
ArCATGGTATTGTGAGyAACTGGGATGACATGGAGAAGATCTGGCATCACACCTTCTACAATG
AACTCCGTGTGGCCCCkGAGGAGCAyCCrGTTCTGCTCACTGAAGCACCTCTCAACCCAAAGG
CTAATCGTGAGAAAATGACCCAAATyATGTTTGAGACCTTCAACACmCCTGCTATGTATGTTG
CCATyCAGGCTGTTyTrTCmCTGTATGCCAGTGGCCGTACAACTGGTATyGTCCTGGACTCTG
GAGATGGTGTGAGCCACACTGTmCCCATyTATGAGGGGTATGCCCTCCCTCATGCCATCCTCC
GTCTTGACTTAGCAGGGCGTGACCTCACTGATACTT
```

Lupinus nanus ACTIN, cDNA, partial codons*

```
CTAACATTGTGGGTCGTCCACGTCACACAGGTGTGATGGTTGGwATGGGwCAAAAGGATGCAT
ATGTTGGTGATGAAGCTCAATCAAAGmGwGGTATATTGACTTTRAAATAyCCAATTGArCATG
GTATTGTGAGyAATTGGGATGACATGGAGAAAATCTGGCATCACACATTyTACAATGAACTTC
GTGTGGCTCCAGAAGAACATCCAGTTCTACTCACTGAAGCCTCTCTTAACCCAAAGGCTAATC
GTGAGAAAATGACTCAAATTATGTTTGAGACTTTCAACACCCCTGCTATGTAGTGCCAATTNA
GCCNGTTTAGyCCCTCTAGCCANTNGTNCCNNNANTNGGATTNNTTNNGAANNCGGNNAANGN
NNNGNNCNANNNGNNCCNAATTNNNNAGGGNTNGNCCNCCCNNNANNCNNNNNTCCGNNNNAA
CTNACCNGGNNTNGCCTNGACTGACTACTT
```

*(sequence poor after 270 bp)

Cadia purpurea Histone H4 homologue, complete codons

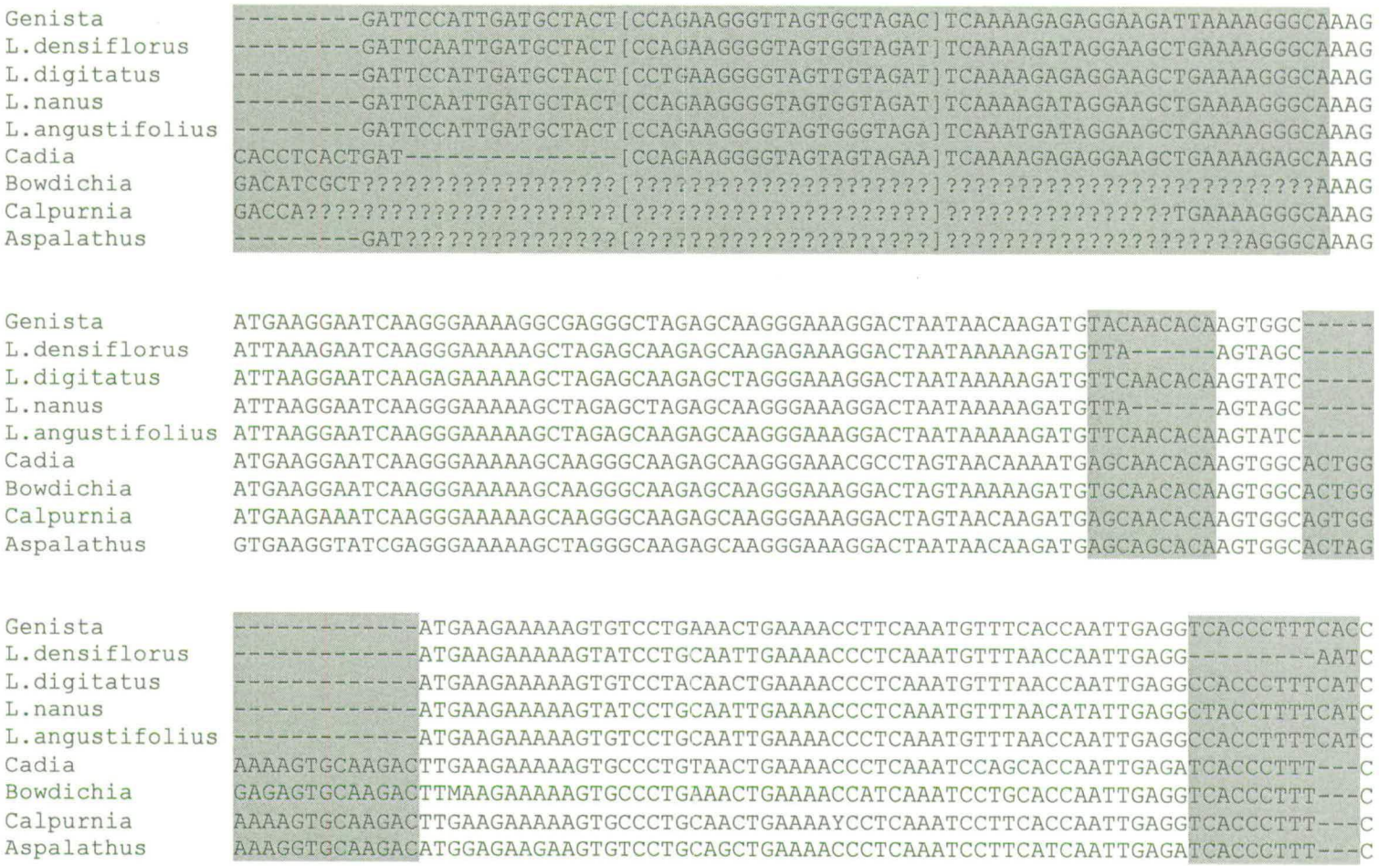
```
CCATGTCTGGAAGAGGAAAGGGAGGGAAAGGTCTGGGAAAGGGAGGAGCAAAACGTCACCGTA
AGGTTCTGAGGGATAACATCCAGGGAATCACGAAGCCTGCGATTCGGCGTCTTGCTCGGCGAG
GGGGTGTAAAGCGTATCAGCGGTTTGATATACGAAGAGACACGTGGTGTCCTCAAGATCTTCC
TGGAGAACGTTATTCGCGATGCTGTCACCTACACTGAGCACGCTCGCCGCAAAACTGTCACTG
CCATGGACGTTGTCTACGCACTCAAGAGGCAGGGTCGTACTCTTTACGGTTTCGGCGGTTAGA
TAA
```

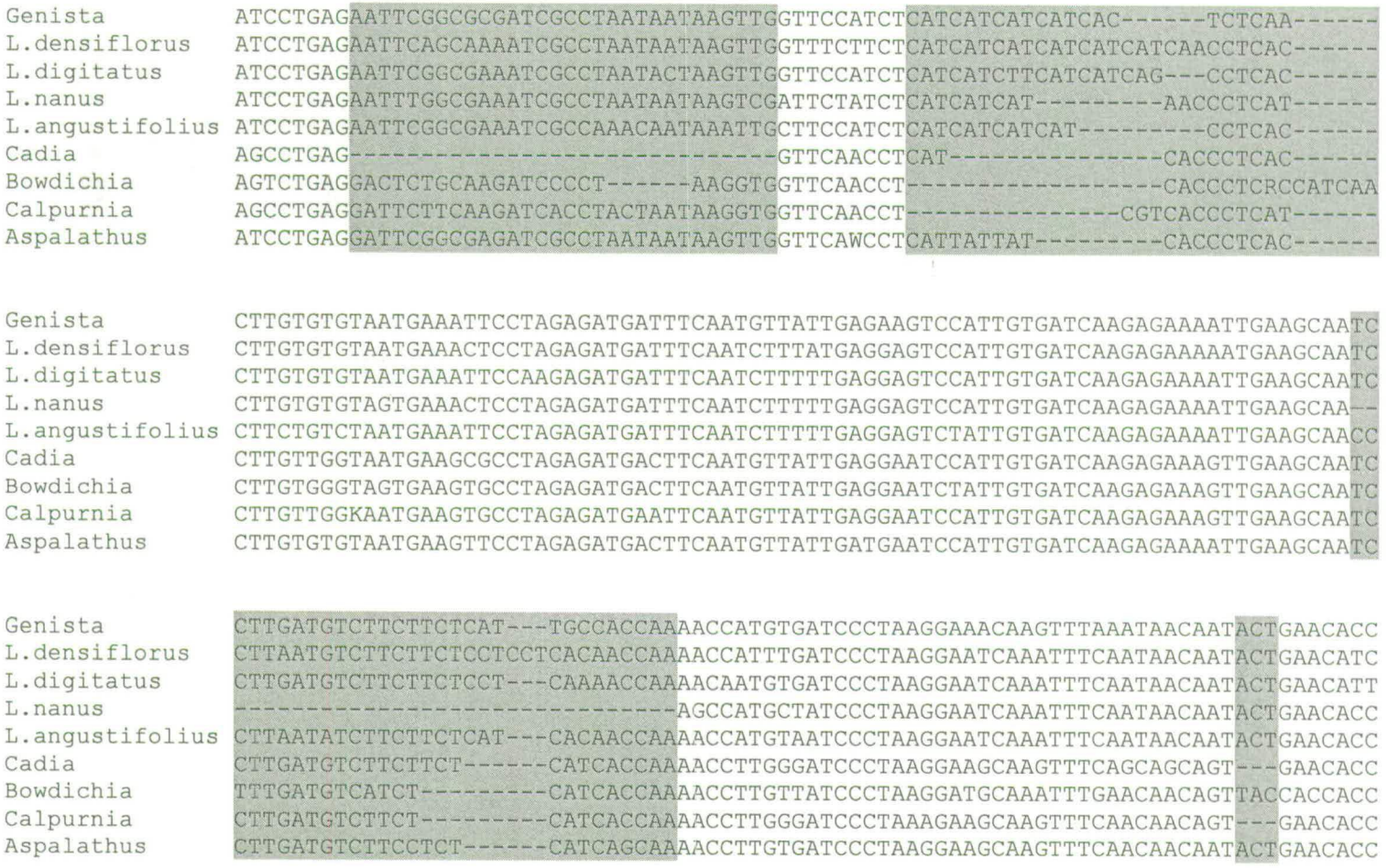
APPENDIX 7: Alignment of LEGCYC1A and LEGCYC1B nucleotide sequences from genistoid taxa. Regions in grey were excluded from the PAML analyses.

LEGCYC1A

```
Genista          AACACA---TTTTCCCATGATCCACTTTCTGTTCCCTACAACATACCAACTACTCATCATCATCATCATGCTCCAATCCC
L.densiflorus    AACACT---TTTTCTCATGATCTACTTTCTTTTCCATATAACATACCAACTACTCATCATTAT---CATGCTCCAATCCC
L.digitatus      AACACA---TTTTCCCATGATCCACTTTCTTTTCCTTACAACATGCCAACCACTCATCATTTT---CATGCTCCAATTCC
L.nanus          AACACC---TTTTCTCATGATCTACTTTCTTTTCCCTATAACATACAACCTACTCATCATTAT---CATGCTCCAACACA
L.angustifolius  AACACA---TTTTCTCATGATCCATTTTCTTTTCCTTACAACATGCCAAATACTCATCATTAT---CATGCTCCAAACCC
Cadia            AACACC---TTTTCCCATGATCCACTTTGTGTTCCCTAC---ATACCTTCTACTCATCAT---------GGTCCAGTCCC
Bowdichia        AACACC---TTCCTTCATGATCCACTTTCTGTTCCCTAC---ATACCCACTACTCATCAT---------TCCCCAATCCC
Calpurnia        AACACC---TTTTCCCATGATCCACTTTCTGTTCCCTAC---ATACCCTCTACTCATCAT---------GGTCCAATCCC
Aspalathus       AACACCACATTTCCCCATGATCCACTTTCTGTTCCTTAC---ATACCAACTCCTCATCAT---------------------

Genista          AGACACACTTTCCAATTTTGCAGAT---------TATGCT---GCTTCAGCTGCAATGTTCAAAAGTGATGATAGT---G
L.densiflorus    AGAAACAGTTTCCAATTTTGCTGAT---------TGTGCTGCTGCTTCAGCTGTAATGTTCAAAAATGATGTTAGT---G
L.digitatus      AGAAACAGTGGCCAATTTTGTAGAT---------TGTGGT---GGTTCAGCTGCAATGTTTAAAAATGATGTTAGT---G
L.nanus          AGAAACTCTTTCCAATTTTGCAGAT---------TATGCT---GCTTCAGCTGCAATGTTTAAGACTGATGTTAGT---G
L.angustifolius  AGAAACAGTTGCCAATTTTGCAGAT---------TGTGGT---GCTTCAGCTGCAATGTTCAAAAATGATGTTAGT---G
Cadia            AGAAACACTAACCAATTTGGCAGTTGCAGAC---TGTTCT---GCAGCAGCTGCAATGTTCAAAAACGATGTCAGT---G
Bowdichia        AGAAACACTAACCAATTTGGCAGTTGCAGAC---TGTGCT---GCTGCAGCTGCAATGTTCAAAAATGATGTCAGTGGGG
Calpurnia        AGAAACACTAACCAATTTGGCAGTTGCAGAC---TGTGCT---GCTGCAGCTGCAATGTTCAAAAACGATGTCAGT---G
Aspalathus       -GAAACACTAGCCAATTTTGCAGTTGCAGAAAATTGTGCT---GCTGCAGCTGCAATGTTCAGAAATGATGTCAGT---G

Genista          GT---TCCAATTTTGGCTTCTCCAATTTGCTCACCAAGAAACCTGCTCCAAAGAAAGACAGGCACAGCAAGATCCACACA
L.densiflorus    GT---TCCAATTTTGGGTTATCCAATTTTCTGGCCAAGAAACCTGCTTCAAAGAAAGACAGGCATAGCAAGATCCATACA
L.digitatus      GT---TCCAATTTTGGCTTCTCCAATTTTATGGCCAAGAAACCTGCTCCAAAGAAAGACAGGCATAGCAAGATCTATACA
L.nanus          GTAATTCCAATTTTGGTTTCTCCAATTTTCTTGCTAAGAAACCTGCTTCTAAGAAAGACAGGCATAGCAAGATCCACACA
L.angustifolius  GT---TCCAATTTTGGCTTCTCCAATTTTATGGCCAAGAAACCTGCTCCAAAGAAAGACAGGCATAGCAAGATCTATACA
Cadia            GT---GTTAATTATGGCTTCTCCAATTTTCTTACAAAGAAACCGCCTGCAAAAAAAAGATAGACACAGTAAGATTCACACA
Bowdichia        GT---GCTCATTATGGCATCTCCAATTTGCTTACCAAAAAACCA---ACCAAGAAAGATAGGCACAGCAAGATTCACACA
Calpurnia        GT---GTCCATTATGGCTTCTCCAATTTTCTTACAAAGAAACCACCTGCAAAGAAAGATAGGCACAGTAAGATTCACACA
Aspalathus       GT---TCCCAATATGGCATCTCAAATTTTCTTACCAAGAAACCTGCTGCAAAGAAAGACAGGCACAGCAAGATCCACACA
```

```
Genista          TCTCAGGGTTTGAGGGACCGCAGGGTGAGATTGTCGATCGATATCTCGCGAAAGTTCTTCGATCTTCAAGACATGTTAGG
L.densiflorus    TCACAGGGTTTGAGAGATAGGAGGGTGAGATTATCGATCGAGATTGCGCGAAAATTCTTTGATCTTCAAGATATGTTAGG
L.digitatus      TCTCAGGGTTTGAGGGACAGAAGGGTGAGATTGTCGATCGAGATCGCGCGAAAGTTCTTTGATCTTCAAGATATGTTAGG
L.nanus          TCACAGGGTTTGAGAGATAGGAGGGTGAGATTATCAATCGAGATCGCGCGAAAGTTCTTCGATCTTCAAGATATGTTAGG
L.angustifolius  TCTCAGGGTTTGAGGGACAGGAGGGTGAGATTGTCGATCGAGATCGCGCGAAAGTTCTTTGATCTTCAAGATATGTTAGG
Cadia            TCTCAGGGTTTGAGGGACCGCAGGGTGAGATTGTCCATTGAGATTGCACGCAAGTTCTTTGATCTTCAAGACATGTTAGG
Bowdichia        TCTCAGGGCTTGAGGGACCGAAGGGTAAGATTGTCCATCGACATTGCGCGCAAGTTCTTCGATCTTCAAGACATGTTAGG
Calpurnia        TCTCAGGGTTTGAGGGACCGCAGGGTGAGATTGTCCATTGAGATTGCACGCAAGTTCTTTGATCTTCAAGACATGTTAGG
Aspalathus       TCTCAGGGTCTGAGGGACCGGAGGGTGAGATTGTCCATCGAGATCGCGCGCAAGTTCTTCGATCTTCAAGACATGCTAGG

Genista          GTTTGACAAGGCCAGCAACACACTTGAGTGGCTCTTCAACAAGTCCAAGAAAGCGATGAAAGAGTTAGCTCAAAGTAAAA
L.densiflorus    GTTTGACAAGGCTAGTAACACACTTGAGTGGCTATTCAACAAGTCCAAGAAAGCAATGAAAGAATTAGCTAGAAGCAAAA
L.digitatus      GTTTGACAAGGCCAGTAACACACTTGAGTGGCTCTTTAACAAGTCAAAGAAAGCAATGAAAGAATTAGCTAGAAGCAAAA
L.nanus          GTTTGACAAGGCTAGTAACACACTTGAGTGGCTATTCAACAAGTCAAAGAAAGCAATGAAAGAATTAGCTAGAAGCAAAA
L.angustifolius  GTTTGACAAGGCCAGTAACACACTTGAGTGGCTATTCAACAAGTCAAAGAAAGCAATGAAGGAATTAGCTAGAAGCAAAA
Cadia            GTTTGACAAAGCCAGCAACACCCTTGAGTGGCTCTTCAACAAATCAAAGAAAGCAATGAAAGAGCTAGCTCGAAGCAAGC
Bowdichia        GTTTGACAAAGGCAGCAGCACCCTTGAGTGGCTCTTCAAAAAGTCAAAGAAAGCAATTAAAGAGCTTGCTAGAAGCAAGA
Calpurnia        GTTTGACAAAGCCAGCAACACCCTTGAGTGGCTATTCAACAAGTCAAAGAAAGCAATGAAAGACCTAGCTCGAAGCAAGC
Aspalathus       GTTTGACAAAGCCAGCAACACCCTTGAGTGGCTCTTCAACAAGTCCAAGAACGCAATGAAAGAGCTAGCTCGAAGCAAGC

Genista          ACAGTGGCAGTGGTGTTGTTGCCAATGGCTTCTCC---TCTTCGGATTCGGAGTGTGAAGTCGTTTCAATGATAAACCAA
L.densiflorus    ACAGTAGCAGTGGTGTTGTTGCAAATAGCTTTTCC---TCTTCGGATTCGGAGTTTGAAGTGGTTTCAATGATAAACCAA
L.digitatus      ACAGTAGCAGTGGTGTTGTTGCAAATAGCTTCTCC---TCTTCGGATTCGGAGTGTGAAGTGGTTTCAATGATAAACCAA
L.nanus          TCAGTAGCAGTGGTGTTGTTGCAAATAGCTTCTCC---TCTTCGGATTCGGAGTTTGAAGTGGTTTCGATGATAAACCCA
L.angustifolius  ACAGTAGCAGTGGTGTTGTTGCAAATAGCTTCTCC---TCTTCGGATTCGGAGTGTGAAGTCGTTTCAATGATAAACCAA
Cadia            AAAGTAGCAGTGGTGCT---GCCAATAGCTTTTCCTCCTCT------ACGGAGTGTGAAGTGGTTTCAGTGATCAACCAA
Bowdichia        ACAGTAGCAGTGGTGCT---GCCAATAGCTTCTCCTCCTCT------TCGGAGTGTGAAGTGGTTTCAGGGATCAACCAA
Calpurnia        AAAGTAGCAGTGGTGCT---GCCAATAGCTTCTCCTCCTCT------TCGGAGTGTGAAGTGGTTTCAGTGATCAACCAA
Aspalathus       ACAGTAGCGGTGGTGGTGGTACCAATAGCTCCTCCTCCTCT------TCGGAATGCGAAGTGGTTTCGGTAAACAACCTA
```


Genista ---------GATTCCATTGATGCTACT[CCAGAAGGGTTAGTGCTAGAC]TCAAAAGAGAGGAAGATTAAAAGGGCAAAG
L.densiflorus ---------GATTCAATTGATGCTACT[CCAGAAGGGGTAGTGGTAGAT]TCAAAAGATAGGAAGCTGAAAAGGGCAAAG
L.digitatus ---------GATTCCATTGATGCTACT[CCTGAAGGGGTAGTTGTAGAT]TCAAAAGAGAGGAAGCTGAAAAGGGCAAAG
L.nanus ---------GATTCAATTGATGCTACT[CCAGAAGGGGTAGTGGTAGAT]TCAAAAGATAGGAAGCTGAAAAGGGCAAAG
L.angustifolius ---------GATTCCATTGATGCTACT[CCAGAAGGGGTAGTGGGTAGA]TCAAATGATAGGAAGCTGAAAAGGGCAAAG
Cadia CACCTCACTGAT----------[CCAGAAGGGGTAGTAGTAGAA]TCAAAAGAGAGGAAGCTGAAAAGAGCAAAG
Bowdichia GACATCGCT??????????????????[??????????????????]??????????????????????????AAAG
Calpurnia GACCA??????????????????????[??????????????????]??????????????TGAAAAGGGCAAAG
Aspalathus ---------GAT????????????[??????????????????]??????????????????AGGGCAAAG
Genista ATGAAGGAATCAAGGGAAAAGGCGAGGGCTAGAGCAAGGGAAAGGACTAATAACAAGATGTACAACACAAGTGGC-----
L.densiflorus ATTAAAGAATCAAGGGAAAAAGCTAGAGCAAGAGCAAGAGAAAGGACTAATAAAAAGATGTTA------AGTAGC-----
L.digitatus ATTAAGGAATCAAGAGAAAAAGCTAGAGCAAGAGCTAGGGAAAGGACTAATAAAAAGATGTTCAACACAAGTATC-----
L.nanus ATTAAGGAATCAAGGGAAAAAGCTAGAGCTAGAGCAAGGGAAAGGACTAATAAAAAGATGTTA------AGTAGC-----
L.angustifolius ATTAAGGAATCAAGGGAAAAAGCTAGAGCAAGAGCAAGGGAAAGGACTAATAAAAAGATGTTCAACACAAGTATC-----
Cadia ATGAAGGAATCAAGGGAAAAAGCAAGGGCAAGAGCAAGGGAAACGCCTAGTAACAAAATGAGCAACACAAGTGGCACTGG
Bowdichia ATGAAGGAATCAAGGGAAAAAGCAAGGGCAAGAGCAAGGGAAAGGACTAGTAAAAAGATGTGCAACACAAGTGGCACTGG
Calpurnia ATGAAGAAATCAAGGGAAAAAGCAAGGGCAAGAGCAAGGGAAAGGACTAGTAACAAGATGAGCAACACAAGTGGCAGTGG
Aspalathus GTGAAGGTATCGAGGGAAAAAGCTAGGGCAAGAGCAAGGGAAAGGACTAATAACAAGATGAGCAGCACAAGTGGCACTAG
Genista ----------ATGAAGAAAAAGTGTCCTGAAACTGAAAACCTTCAAATGTTTCACCAATTGAGGTCACCCTTTCACC
L.densiflorus ----------ATGAAGAAAAAGTATCCTGCAATTGAAAACCCTCAAATGTTTAACCAATTGAGG---------AATC
L.digitatus ----------ATGAAGAAAAAGTGTCCTACAACTGAAAACCCTCAAATGTTTAACCAATTGAGGCCACCCTTTCATC
L.nanus ----------ATGAAGAAAAAGTATCCTGCAATTGAAAACCCTCAAATGTTTAACATATTGAGGCTACCTTTTCATC
L.angustifolius ----------ATGAAGAAAAAGTGTCCTGCAATTGAAAACCCTCAAATGTTTAACCAATTGAGGCCACCTTTTCATC
Cadia AAAAGTGCAAGACTTGAAGAAAAAGTGCCCTGTAACTGAAAACCCTCAAATCCAGCACCAATTGAGATCACCCTTT---C
Bowdichia GAGAGTGCAAGACTTMAAGAAAAAGTGCCCTGAAACTGAAAACCATCAAATCCTGCACCAATTGAGGTCACCCTTT---C
Calpurnia AAAAGTGCAAGACTTGAAGAAAAAGTGCCCTGCAACTGAAAAYCCTCAAATCCTTCACCAATTGAGGTCACCCTTT---C
Aspalathus AAAGGTGCAAGACATGGAGAAGAAGTGTCCTGCAGCTGAAAACCCTCAAATCCTTCATCAATTGAGATCACCCTTT---C


Genista ATCCTGAGAATTCGGCGCGATCGCCTAATAATAAGTTGGTTCCATCTCATCATCATCATCAC------TCTCAA------
L.densiflorus ATCCTGAGAATTCAGCAAAATCGCCTAATAATAAGTTGGTTTCTTCTCATCATCATCATCATCATCAACCTCAC------
L.digitatus ATCCTGAGAATTCGGCGAAATCGCCTAATACTAAGTTGGTTCCATCTCATCATCTTCATCATCAG---CCTCAC------
L.nanus ATCCTGAGAATTTGGCGAAATCGCCTAATAATAAGTCGATTCTATCTCATCATCAT---------AACCCTCAT------
L.angustifolius ATCCTGAGAATTCGGCGAAATCGCCAAACAATAAATTGCTTCCATCTCATCATCATCAT---------CCTCAC------
Cadia AGCCTGAG----------GTTCAACCTCAT----------CACCCTCAC------
Bowdichia AGTCTGAGGACTCTGCAAGATCCCCT------AAGGTGGTTCAACCT----------CACCCTCRCCATCAA
Calpurnia AGCCTGAGGATTCTTCAAGATCACCTACTAATAAGGTGGTTCAACCT----------CGTCACCCTCAT------
Aspalathus ATCCTGAGGATTCGGCGAGATCGCCTAATAATAAGTTGGTTCAWCCTCATTATTAT---------CACCCTCAC------
Genista CTTGTGTGTAATGAAATTCCTAGAGATGATTTCAATGTTATTGAGAAGTCCATTGTGATCAAGAGAAAATTGAAGCAATC
L.densiflorus CTTGTGTGTAATGAAACTCCTAGAGATGATTTCAATCTTTATGAGGAGTCCATTGTGATCAAGAGAAAAATGAAGCAATC
L.digitatus CTTGTGTGTAATGAAATTCCAAGAGATGATTTCAATCTTTTTGAGGAGTCCATTGTGATCAAGAGAAAATTGAAGCAATC
L.nanus CTTGTGTGTAGTGAAACTCCTAGAGATGATTTCAATCTTTTTGAGGAGTCCATTGTGATCAAGAGAAAATTGAAGCAA--
L.angustifolius CTTCTGTCTAATGAAATTCCTAGAGATGATTTCAATCTTTTTGAGGAGTCTATTGTGATCAAGAGAAAATTGAAGCAACC
Cadia CTTGTTGGTAATGAAGCGCCTAGAGATGACTTCAATGTTATTGAGGAATCCATTGTGATCAAGAGAAAGTTGAAGCAATC
Bowdichia CTTGTGGGTAGTGAAGTGCCTAGAGATGACTTCAATGTTATTGAGGAATCTATTGTGATCAAGAGAAAGTTGAAGCAATC
Calpurnia CTTGTTGGKAATGAAGTGCCTAGAGATGAATTCAATGTTATTGAGGAATCCATTGTGATCAAGAGAAAGTTGAAGCAATC
Aspalathus CTTGTGTGTAATGAAGTTCCTAGAGATGACTTCAATGTTATTGATGAATCCATTGTGATCAAGAGAAAATTGAAGCAATC
Genista CTTGATGTCTTCTTCTCAT---TGCCACCAAAACCATGTGATCCCTAAGGAAACAAGTTTAAATAACAATACTGAACACC
L.densiflorus CTTAATGTCTTCTTCTCCTCCTCACAACCAAAACCATTTGATCCCTAAGGAATCAAATTTCAATAACAATACTGAACATC
L.digitatus CTTGATGTCTTCTTCTCCT---CAAAACCAAAACAATGTGATCCCTAAGGAATCAAATTTCAATAACAATACTGAACATT
L.nanus ----------AGCCATGCTATCCCTAAGGAATCAAATTTCAATAACAATACTGAACACC
L.angustifolius CTTAATATCTTCTTCTCAT---CACAACCAAAACCATGTAATCCCTAAGGAATCAAATTTCAATAACAATACTGAACACC
Cadia CTTGATGTCTTCTTCT------CATCACCAAAACCTTGGGATCCCTAAGGAAGCAAGTTTCAGCAGCAGT---GAACACC
Bowdichia TTTGATGTCATCT---------CATCACCAAAACCTTGTTATCCCTAAGGATGCAAATTTGAACAACAGTTACCACCACC
Calpurnia CTTGATGTCTTCT---------CATCACCAAAACCTTGGGATCCCTAAAGAAGCAAGTTTCAACAACAGT---GAACACC
Aspalathus CTTGATGTCTTCCTCT------CATCAGCAAAACCTTGTGATCCCTAAGGAAGCAAGTTTCAACAACAATACTGAACACC

```
Genista          ACTCCTTCCCCATTTTATCTCCAAATTGGGATGCTAATAATGGTGACAATGGCAAATCCAACTTTTGTGCAATAGCCAGC
L.densiflorus    ACTCCTTCCCTATTTTATCTCCAAATTTGGATGCTAATAATGGTGCCAATGGCAGATCCAACTTTTGTGCAGTAACCAAC
L.digitatus      ACTCCTTCCCTATTTTATCTCCAAATTTGGATGCTAATAATGGTGCCAATGGCAGATCCAACTTCTGTGCAGTAACCAAC
L.nanus          ACTCCTTTCCCATTTTATCTCCAAATTTGGATGCTAATAATGGTGCCAATGGCAGATCCAATTTTTGTGCAGTTACCAAC
L.angustifolius ACTCCTTCCCCATTTTATCTCCAAATTTGGATGCTAATAATGGTGCCAATAGCAGATCCAACTTTTGTTCAATAACCAAC
Cadia            ACTCCTTCCCCATTTTATCTCCAAATTGGGATGCAAAT---GGTGCCACTGGCCGTTCCAACTTTTATGCAATAGCCAGC
Bowdichia        ACTCCTTCCCCAATTGTTCTCCAAATTGGGATGCTGAT---GGTGCCACTGGCCGTTCCAACTTTTGTGCAATAGCCAGC
Calpurnia        ACTCCTTCCCCATTTTATCTCCAAATTGGGATGCTAAT---GGTGCCACTGGCTGTTCCAACTTTTGTGCAATCGCCAGT
Aspalathus       ACTCTTTCCCAATTTTATCTCCAAATTGGGATGCAAAT---GGTGCCACAGGCCGATCCAACTTTTGTGCAATAGCCAGC
```

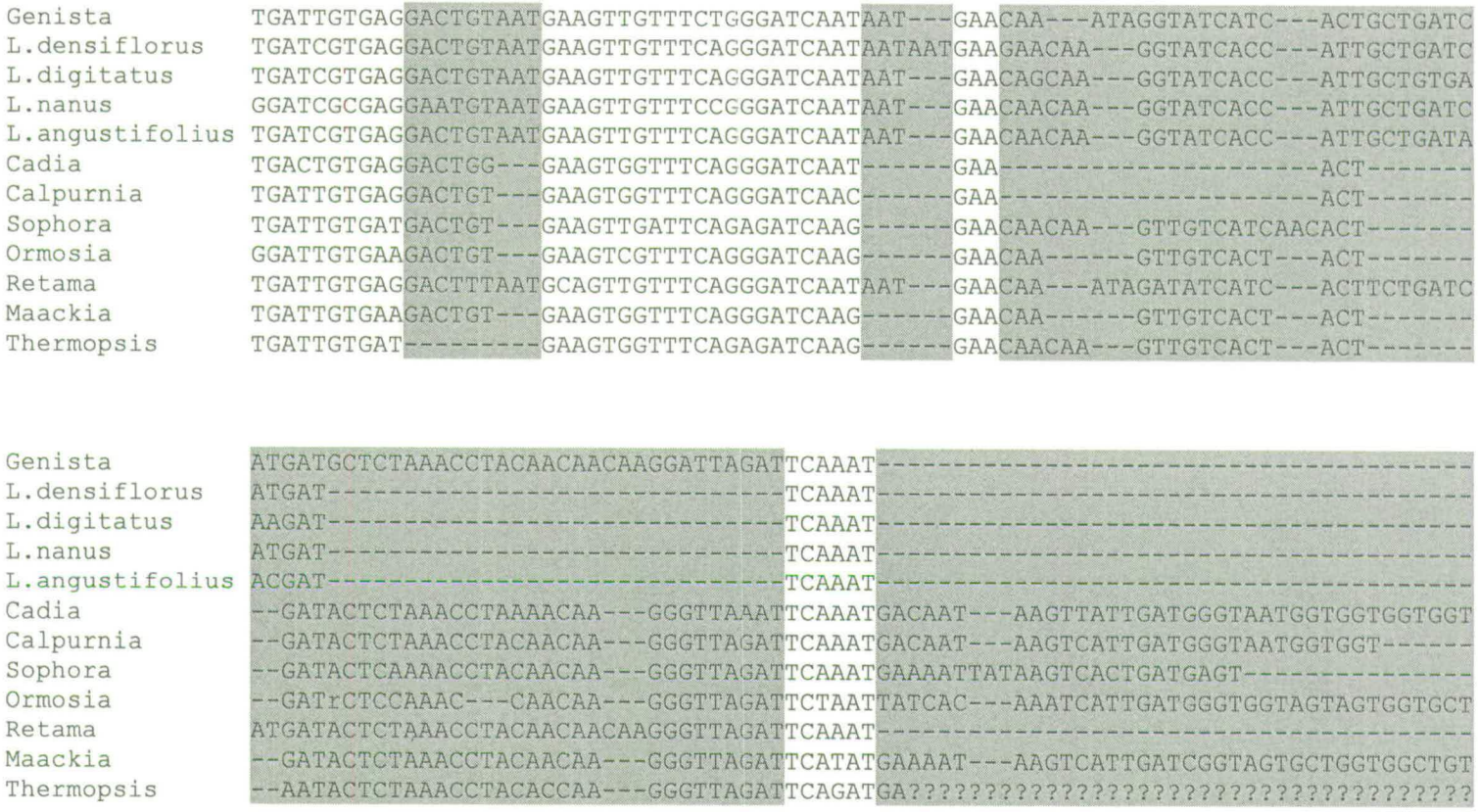
```
Genista          ATGAATCTATCTACAG
L.densiflorus    ATGAATCTATCAACAG
L.digitatus      ATGAATCTATCTACAG
L.nanus          ATGAATCTATCAACAG
L.angustifolius ATGAATCTATCTACAG
Cadia            ATGAATCTATCTACAG
Bowdichia        TTGAATCTTTCTACAG
Calpurnia        ATGAATTTATCTACAG
Aspalathus       ATGAATCTATCTACAG
```

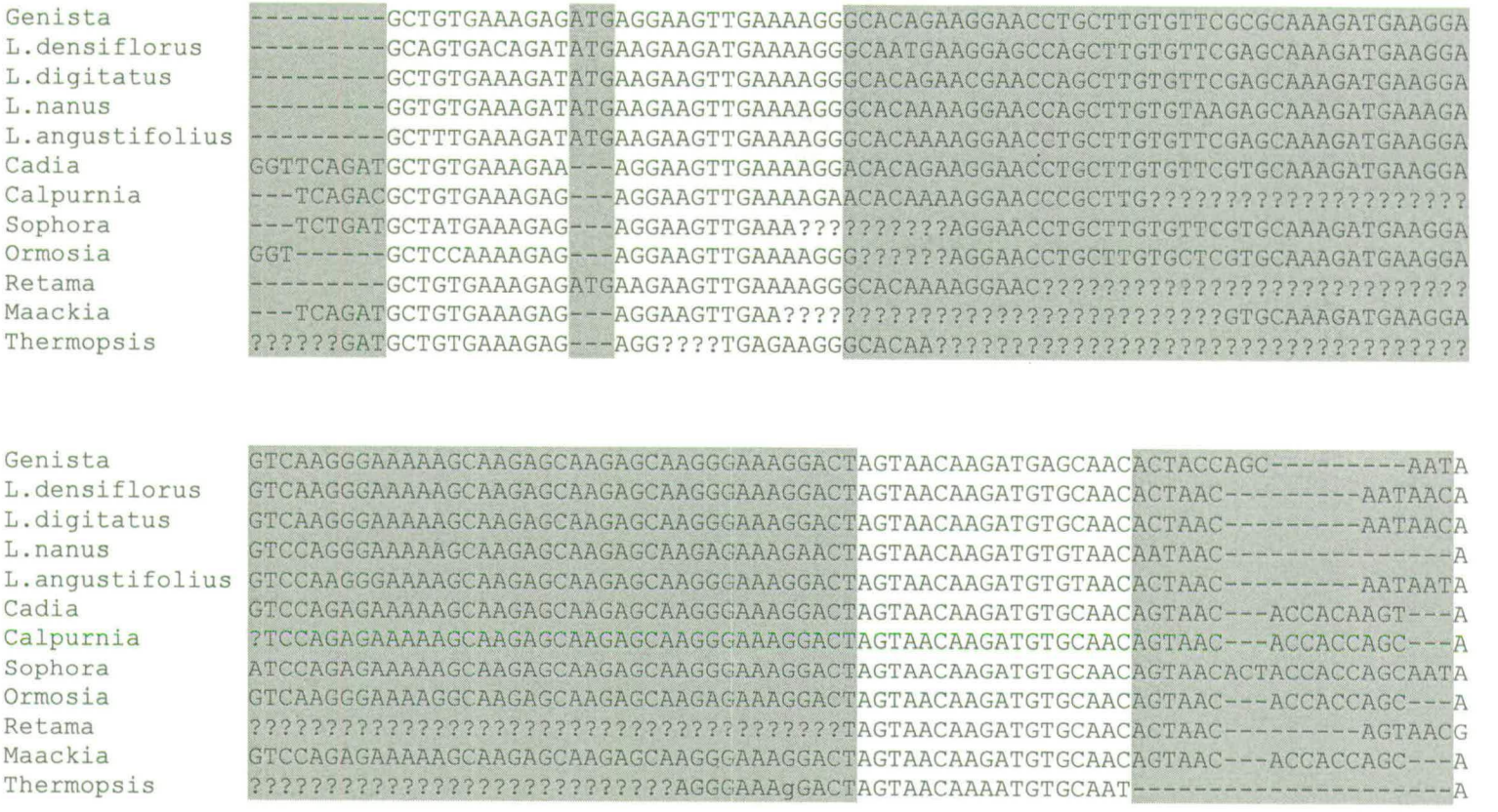
LEGCYC1B

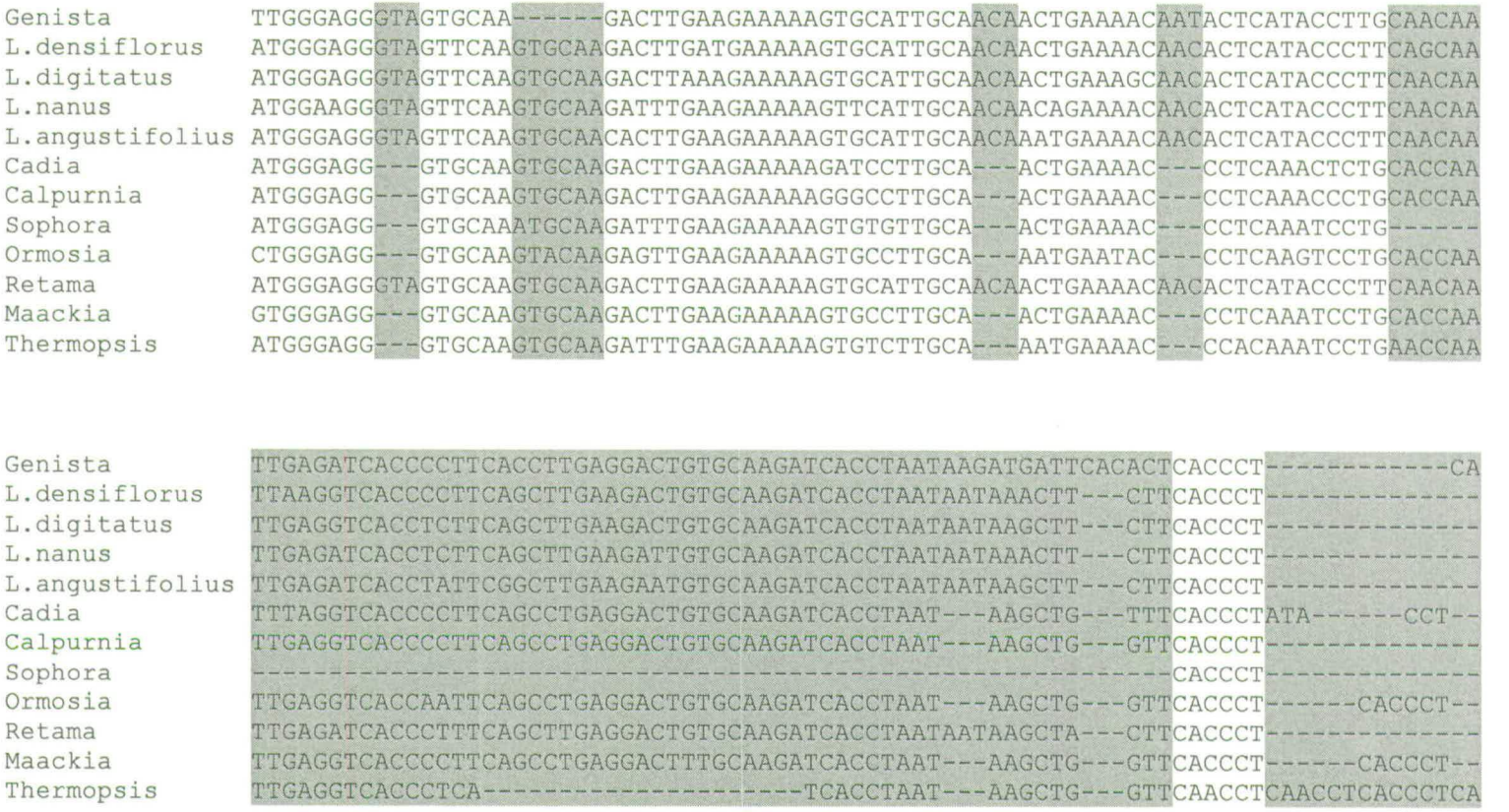
```
Genista          ACCCTT---CTTCATGATCCACTTGCT---CATGTT---CCCTACAACATACCAACT---CATCATCATATTCATAACAC
L.densiflorus    AACCTT---CTTCATGATCCACTTGTT---CATGTT---CCCTACAACTTACCAACTATTCATCATCATATTCATAACAC
L.digitatus      ACCCTT---CTTCATGATCCACTTGTT---CATGTT---CCCTACAACTTACCAACT---CATCATAATATTCATAACAC
L.nanus          AACCTT---CTTCATGATCCACTTGTT---CATGTT---CCTTACAACTTACCAAGT---CATCATCATATTCATAACAC
L.angustifolius  ACCCTT---CTTCATGATCCACTTGTT---CATGTT---CCCTACAACTTACCAACT---CATCATCATATTCATAATAC
Cadia            ACCTCT---CTTCATGATCCACTTGCT------GTT---CCATAC---ATACCAACT---CATCAT---------AACAC
Calpurnia        ACCCTT---CTTCATGATCCACTTGCT------GTT---CCATAC---ATACCAACT---CATCAT---------AACAC
Sophora          ACCCTT---CTTCATGATCCACTTTCTGTACATGTTAATCCCTAC---ATACCAACT---CATCAT---------AACTC
Ormosia          ACCCTT---CTTCATGATCCACTTGCT------GTT---CCCTAC---ATACCAACT---CATCAT---------AACAC
Retama           ACCCTT---CTTCATGATCCACTTGCT---CATGTT---CCCTACAACATACCAACT---CATCATCATAkTCATAACAC
Maackia          ACCCTT---CTTCATGATCCACTTGCT------GTT---CCCTAC---ATACCAACT---CATCAT---------AACAC
Thermopsis       ACCCTTAATCTTCATGATCCACTTCTTGTTAATGTT---CCCTAC---ACACACACT---CATCAA----------

Genista          ACCAATAATCCAA GAAACACTGACC---AATTTGGCTGTTTCTGAT---------GCTCATGCTGCTGCAATGCCGAAA
L.densiflorus    ACCTATAATCCAT GAAACACTGACC---AATTTGGCTGTTTCTGAT------GCTGCTGCTGCTACA---ATGCCCAAA
L.digitatus      ACCTATAATCCAA GAAACACTGACC---AATTTGGCTGTTTCTGAT---------GCCGCTGCTACA---ATGCCGAAA
L.nanus          ACCTATAATCCAA GAAACACTGACC---AATTTGGCTGTTTCTGAT----------GCTGCTACA---ATGCCGAAA
L.angustifolius  ACCTATAATCCAA GAAACACTGACC---AATTTGGCTGTTTCTGAT---------GTTGCTGCTACA---ATGCCGAAA
Cadia            TCCA---ATCCCA GAAACACTGACA---AATTTGGCAGTTTCTGATGACTGTGGT---GCTGCTTCA---ATGCCCAAA
Calpurnia        TCCA---ATCCCA GAAACACTGACC---AATTTGGTAGTTTCTGATAACTGTGGTGCTGCT------GCAATGCCCAAA
Sophora          TCAA---ATCCCA GGAACACTTACC---AATTTGGCTGTTTCTGATAACTGTGGTGGTGCTGCTGCTACAATGCACAAA
Ormosia          TCAA---ATCCCA GAAACACTGACC---AATTTGGCAGTTTCAGATAACTGT---GCTGCTGCTGCTGCAATGCCTAAG
Retama           ACCAATAATCCAA GAAACACTGACC---AATTTGGCTGTTTCTGAT---------GCTCATGCTGCTGCAATGCCGAAA
Maackia          TCCA---ATCCCA GAAACACTGACC---AATTTGGCTGTTTCTGATAACTGTGGT------GCTGCTGCAATGCCCAAA
Thermopsis       -------ATTCCA GAAACACTGACCAATAATTTGGCTCTTTCTGAT----------------ACTCCAATGCCCAAA
```

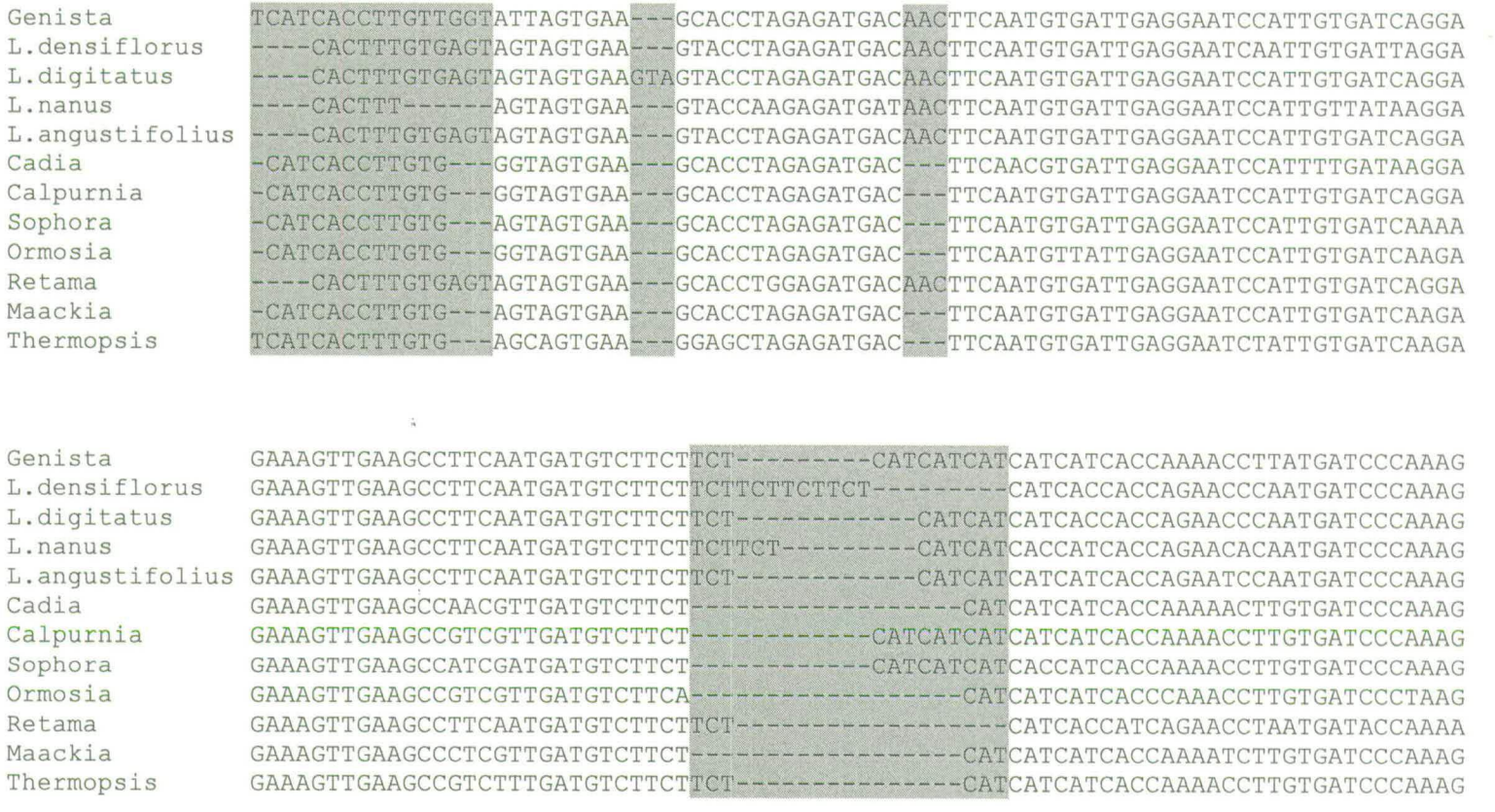
```
Genista          CAAGACCCGATTAT?AATGGCGGTGGTGTT------CATCATCACTATGGACTTTCTTCTTTGCTCACAAAGAAACCAGC
L.densiflorus    CAAGACCCTATTATGAGTGGTGGTGCT---------CATCATCACTATGGCCTTTCTTGTCTGCTCACAAAGAAACCAGC
L.digitatus      CAAGACCCTATTATGAGTGGTGGTGGTGGTGTT---CATCATCACTATGGGCTTTCTTCTCTGCTTACAAAGAAACCAGC
L.nanus          CAAGACCCTATTATGAGTGGTGGTGGTGGTGGTGTTCATCATCACTATGGGCTTTCTTCTCTGCTCACAAAGAAACCAGC
L.angustifolius  CAAGACCCTATTATGAGTGGTGGTGGTGGTGCT---CATCATCACTATGGGCTTTCTTCTCTGCTCACAAAGAAACCAGC
Cadia            CAAGACACT------AGTGGTGCT-------------------CACTATGGCCTTTCTTGTTTGCTTACAAAGAAACCAGC
Calpurnia        CAAGACCCT------AGTGGTGCT-------------------CACTATGGCCTTTCTTGTTTGCTCACAAAGAAACCAGC
Sophora          CAAGACCCT------AGTGGTGGTGCT----------------AACTATGGCTTTTCTAGTTTGATCACAAAGAAACCAGC
Ormosia          CAAGACTCC------ACTGGTGCT-------------------CACTATGGCATTTCyAGTTTGCTCACAAAGAAACCAGC
Retama           CAAGACCTGATTATAAGTGGCGGTGGTGGTGTT---CATCATCACTATGGACTTTCTTCTTTGCTCACAAAGAAACCAGC
Maackia          CAAGACCCT------AGTGCTGCT-------------------CACTATGGCCTTTCTTGTTTGCTCACAAAGAAACCAGC
Thermopsis       CAAGACCCT------AATGTTTCTTCT----------------CACTATGGCATTTCTTGTTTGCTTACAAAGAAGCCAGC

Genista          TAAGAAAGACAGGCACAGCAAGATTTACACCTCTCAGGGCTTGAGGGACAGGAGGGTGAGGCTGTCGATCGAGATCGCAC
L.densiflorus    TAAAAAAGATAGGCACAGCAAGATTTACACCTCTCAGGGCTTGAGGGATCGGAGGGTGAGGCTTTCGATCGAGATCGCTC
L.digitatus      CAAAAAAGATAGGCACAGTAAGATTTACACCTCTCAGGGCCTGAGGGATCGGAGGGTGAGGCTTTCGATCGAGATCGCAC
L.nanus          CAAAAAGGATAGGCACAGCAAGATTTACACCTCTCAGGGCTTGAGGGATCGGAGGGTGAGGCTTTCGATTGAGATCGCGC
L.angustifolius  CAAAAAAGATAGGCACAGTAAGATTTACACCTCTCAGGGCTTGAGGGATCGGAGGGTGAGGCTTTCGATCGAGATCGCAC
Cadia            CAAGAAAGATAGGCACAGCAAGATTTACACCTCCCAGGGCTTGAGGGACCGCAGGGTGAGGTTGTCCATTGAGATCGCCC
Calpurnia        CAAGAAAGATAGGCACAGCAAGATTTACACCTCTCAGGGCTTGAGGGACCGTAGGGTGAGGTTGTCCATTGATATCGCCC
Sophora          CAAGAAAGACAGGCATAGCAAGATTTACACTTCTCAAGGCTTGAGGGACCGGAGGGTGAGGTTGTCGATCGAGATCGCAC
Ormosia          AAAGAAAGATAGGCACAGCAAGATTTACACCTCTCAGGGTTTGAGGGACCGCAGGGTGAGGTTGTCCATCGAGATTGCCC
Retama           CAAGAAAGATAGGCACAGCAAGATTTATACCTCTCAAGGCTTGAGGGACCGCAGGGTGAGGCTGTCGATCGAGATTGCGC
Maackia          CAAGAAAGACAGGCATAGCAAGATTTACACCTCTCAGGGCTTGAGGGACCGTAGGGTGAGGTTGTCCATCGAGATCGCCC
Thermopsis       TAAGAAAGACAGGCATAGCAAGATATACACTTCTCAAGGCTTGAGAGACCGTAGGGTGAGGTTATCGATCGAGATCGCGC
```

```
Genista          GGAAGTTCTTCGATCTACAAGACATGCTAGGGTTTGACAAAGCAAGCAACACCCTTGAGTGGCTCTTCAACAAGTCCAAG
L.densiflorus    GAAAGTTCTTCGATCTACAAGACATGCTAGGGTTTGACAAAGCAAGCAACACCCTTGAGTGGCTCTTCAACAAGTCCAAG
L.digitatus      GAAAGTTCTTCGATCTACAAGACATGCTAGGGTTTGACAAAGCAAGCAACACCCTTGAGTGGCTCTTCAACAAGTCCAAG
L.nanus          GAAAGTTCTTCGATCTACAAGACATGCTAGGGTTTGACAAAGCAAGCAACACCCTTGAGTGGCTCTTCAACAAGTCCAAG
L.angustifolius GAAAGTTCTTCGATCTACAAGATATGCTAGGGTTTGACAAAGCAAGCAACACCCTTGAGTGGCTCTTCAACAAATCCAAG
Cadia            GCAAGTTCTTTGATCTACAAGACATGCTAGGGTTTGACAAAGCCAGTAACACTCTTGAGTGGCTCTTCAACAAGTCCAAG
Calpurnia        GCAAGTTCTTTGATCTACAAGACATGCTAGGGTTTGACAAAGCCAGTAACACCCTTGAGTGGCTCTTCAACAAGTCCAAG
Sophora          GAAAGTTCTTTGATCTACAAGACATGCTAGGGTTTGACAAAGCAAGTAACACCCTTGAGTGGCTCTTCAACAAGTCCAAG
Ormosia          GCAAGTTCTTTGATCTTCAAGACATGCTAGGTTTTGACAAAGCCAGCAACACCCTTGAGTGGCTCTTCAACAAGTCCAAG
Retama           GAAAGTTCTTCGATCTACAAGACATGCTAGGGTTTGACAAAGCAAGCAACACCCTTGAGTGGCTCTTCAACAAGTCCAAG
Maackia          GCAAGTTCTTTGATCTACAAGACATGCTAGGGTTTGACAAAGCCAGTAACACCCTTGAGTGGCTCTTCAACAAGTCCAAG
Thermopsis       GAAAGTTCTTTGATCTACAAGACATGTTAGGGTTTGACAAAGCAAGTAACACCCTTGAGTGGCTCTTCAACAAGTCCAAG

Genista          AAAGCAATTAAGAAGCTAGCTAGAAGCAACAACAGCAATATCAGT---GAAGGTGATGCTAAGAGCTTATCCTCTTCTTC
L.densiflorus    AGAGCAATTAAGGAGCTAGCTAGAAGCAAGAAAAGCAATGGCAAT---GAAGGTGATGCTAATAACTTATCCTCATCTTC
L.digitatus      AGAGCAATTAAGGAGCTAGCTAGAAGCAAGAAAACCAATGGCAGT---GAAGGTGATGCTAATAGCTTCTCATCATCTTC
L.nanus          AGAGCAATTAAGGACCTAGCTAGAAGCAAGAAAAACAATGGTAGT---GAAGGTGATGCTAATAGTTTATCCTCATCTTC
L.angustifolius AGAGCAATTAAGGAGCTAGCTAGAAGCAAGAAAAACAATGGCAGT---GAAGGTGATGCTAATAGCTTCTCCTCATCTTC
Cadia            AAAGCAATTAAAGATCTAGCCAGAAGCAAGCACAGCAAC---AGT---GAAGGT---GCCAAGAGCTTCGCCTCATCTTC
Calpurnia        AAAGCAATTAAAGATCTAGCCAGAAGCAAACACAGC---ATCAGT---GAAGGT---GCAAAGAGCTTCGCCTCATCTTC
Sophora          AAAGCAATTAAGGATCTAGCTAGAAGCAAGAACAGCAATATCAGT---GATGGT---GCTAAGAGTTTCTCCTCATCTTC
Ormosia          AAAGCAATTAAAGAGCTAGCTCGAAGCAAGCACAGCAAC---AGy---GAAGGT---GCCAAGAGCTTCTCCTCATCTTC
Retama           AAAGCAATAAAGGAGCTAGCTAGAAGCAAGAACAGCAATATCAGT---GAAGGTGATGCTAAGAGCTTCTCCTCTTCTTC
Maackia          AAAGCAATTAAAGAGCTAGCTAGAAGCAAGCACAGC---ATCAGC---GAAGGT---GCAAAGAGCTTCTCCTCATCTTC
Thermopsis       AAAGCAATTAAAGATCTAGCTAGAACCAAACACAACATT------GAAGAAGGT---TCCAAAAGGTTC---TCTTCTTC
```


Genista TGATTGTGAGGACTGTAATGAAGTTGTTTCTGGGATCAATAAT---GAACAA---ATAGGTATCATC---ACTGCTGATC
L.densiflorus TGATCGTGAGGACTGTAATGAAGTTGTTTCAGGGATCAATAATAATGAAGAACAA---GGTATCACC---ATTGCTGATC
L.digitatus TGATCGTGAGGACTGTAATGAAGTTGTTTCAGGGATCAATAAT---GAACAGCAA---GGTATCACC---ATTGCTGTGA
L.nanus GGATCGCGAGGAATGTAATGAAGTTGTTTCCGGGATCAATAAT---GAACAACAA---GGTATCACC---ATTGCTGATC
L.angustifolius TGATCGTGAGGACTGTAATGAAGTTGTTTCAGGGATCAATAAT---GAACAACAA---GGTATCACC---ATTGCTGATA
Cadia TGACTGTGAGGACTGG---GAAGTGGTTTCAGGGATCAAT------GAA----------ACT-------
Calpurnia TGATTGTGAGGACTGT---GAAGTGGTTTCAGGGATCAAC------GAA----------ACT-------
Sophora TGATTGTGATGACTGT---GAAGTTGATTCAGAGATCAAG------GAACAACAA---GTTGTCATCAACACT-------
Ormosia GGATTGTGAAGACTGT---GAAGTCGTTTCAGGGATCAAG------GAACAA------GTTGTCACT---ACT-------
Retama TGATTGTGAGGACTTTAATGCAGTTGTTTCAGGGATCAATAAT---GAACAA---ATAGATATCATC---ACTTCTGATC
Maackia TGATTGTGAAGACTGT---GAAGTGGTTTCAGGGATCAAG------GAACAA------GTTGTCACT---ACT-------
Thermopsis TGATTGTGAT---------GAAGTGGTTTCAGAGATCAAG------GAACAACAA---GTTGTCACT---ACT-------
Genista ATGATGCTCTAAACCTACAACAACAAGGATTAGATTCAAAT----------
L.densiflorus ATGAT----------TCAAAT----------
L.digitatus AAGAT----------TCAAAT----------
L.nanus ATGAT----------TCAAAT----------
L.angustifolius ACGAT----------TCAAAT----------
Cadia --GATACTCTAAACCTAAAACAA---GGGTTAAATTCAAATGACAAT---AAGTTATTGATGGGTAATGGTGGTGGTGGT
Calpurnia --GATACTCTAAACCTACAACAA---GGGTTAGATTCAAATGACAAT---AAGTCATTGATGGGTAATGGTGGT------
Sophora --GATACTCAAAACCTACAACAA---GGGTTAGATTCAAATGAAAATTATAAGTCACTGATGAGT----------
Ormosia --GATrCTCCAAAC---CAACAA---GGGTTAGATTCTAATTATCAC---AAATCATTGATGGGTGGTAGTAGTGGTGCT
Retama ATGATACTCTAAACCTACAACAACAAGGGTTAGATTCAAAT----------
Maackia --GATACTCTAAACCTACAACAA---GGGTTAGATTCATATGAAAAT---AAGTCATTGATCGGTAGTGCTGGTGGCTGT
Thermopsis --AATACTCTAAACCTACACCAA---GGGTTAGATTCAGATGA??????????????????????????????????


Genista ---------GCTGTGAAAGAGATGAGGAAGTTGAAAAGGGCACAGAAGGAACCTGCTTGTGTTCGCGCAAAGATGAAGGA
L.densiflorus ---------GCAGTGACAGATATGAAGAAGATGAAAAGGGCAATGAAGGAGCCAGCTTGTGTTCGAGCAAAGATGAAGGA
L.digitatus ---------GCTGTGAAAGATATGAAGAAGTTGAAAAGGGCACAGAACGAACCAGCTTGTGTTCGAGCAAAGATGAAGGA
L.nanus ---------GGTGTGAAAGATATGAAGAAGTTGAAAAGGGCACAAAAGGAACCAGCTTGTGTAAGAGCAAAGATGAAAGA
L.angustifolius ---------GCTTTGAAAGATATGAAGAAGTTGAAAAGGGCACAAAAGGAACCTGCTTGTGTTCGAGCAAAGATGAAGGA
Cadia GGTTCAGATGCTGTGAAAGAA---AGGAAGTTGAAAAGGACACAGAAGGAACCTGCTTGTGTTCGTGCAAAGATGAAGGA
Calpurnia ---TCAGACGCTGTGAAAGAG---AGGAAGTTGAAAAGAACACAAAAGGAACCCGCTTG??????????????????????
Sophora ---TCTGATGCTATGAAAGAG---AGGAAGTTGAAA???????????AGGAACCTGCTTGTGTTCGTGCAAAGATGAAGGA
Ormosia GGT------GCTCCAAAAGAG---AGGAAGTTGAAAAGGG???????AGGAACCTGCTTGTGCTCGTGCAAAGATGAAGGA
Retama ---------GCTGTGAAAGAGATGAAGAAGTTGAAAAGGGCACAAAAGGAAC??????????????????????????????
Maackia ---TCAGATGCTGTGAAAGAG---AGGAAGTTGAA????????????????????????????????GTGCAAAGATGAAGGA
Thermopsis ??????GATGCTGTGAAAGAG---AGG????TGAGAAGGGCACAA????????????????????????????????????
Genista GTCAAGGGAAAAAGCAAGAGCAAGAGCAAGGGAAAGGACTAGTAACAAGATGAGCAACACTACCAGC---------AATA
L.densiflorus GTCAAGGGAAAAAGCAAGAGCAAGAGCAAGGGAAAGGACTAGTAACAAGATGTGCAACACTAAC---------AATAACA
L.digitatus GTCAAGGGAAAAAGCAAGAGCAAGAGCAAGGGAAAGGACTAGTAACAAGATGTGCAACACTAAC---------AATAACA
L.nanus GTCCAGGGAAAAAGCAAGAGCAAGAGCAAGAGAAAGAACTAGTAACAAGATGTGTAACAATAAC---------------A
L.angustifolius GTCCAAGGGAAAAGCAAGAGCAAGAGCAAGGGAAAGGACTAGTAACAAGATGTGTAACACTAAC---------AATAATA
Cadia GTCCAGAGAAAAAGCAAGAGCAAGAGCAAGGGAAAGGACTAGTAACAAGATGTGCAACAGTAAC---ACCACAAGT---A
Calpurnia ?TCCAGAGAAAAAGCAAGAGCAAGAGCAAGGGAAAGGACTAGTAACAAGATGTGCAACAGTAAC---ACCACCAGC---A
Sophora ATCCAGAGAAAAAGCAAGAGCAAGAGCAAGGGAAAGGACTAGTAACAAGATGTGCAACAGTAACACTACCACCAGCAATA
Ormosia GTCAAGGGAAAAGGCAAGAGCAAGAGCAAGAGAAAGGACTAGTAACAAGATGTGCAACAGTAAC---ACCACCAGC---A
Retama ??????????????????????????????????????TAGTAACAAGATGTGCAACACTAAC---------AGTAACG
Maackia GTCCAGAGAAAAAGCAAGAGCAAGAGCAAGGGAAAGGACTAGTAACAAGATGTGCAACAGTAAC---ACCACCAGC---A
Thermopsis ??????????????????????????????AGGGAAAgGACTAGTAACAAAATGTGCAAT----------A


Genista TTGGGAGGGTAGTGCAA------GACTTGAAGAAAAAGTGCATTGCAACAACTGAAAACAATACTCATACCTTGCAACAA
L.densiflorus ATGGGAGGGTAGTTCAAGTGCAAGACTTGATGAAAAAGTGCATTGCAACAACTGAAAACAACACTCATACCCTTCAGCAA
L.digitatus ATGGGAGGGTAGTTCAAGTGCAAGACTTAAAGAAAAAGTGCATTGCAACAACTGAAAGCAACACTCATACCCTTCAACAA
L.nanus ATGGAAGGGTAGTTCAAGTGCAAGATTTGAAGAAAAAGTTCATTGCAACAACAGAAAACAACACTCATACCCTTCAACAA
L.angustifolius ATGGGAGGGTAGTTCAAGTGCAACACTTGAAGAAAAAGTGCATTGCAACAAATGAAAACAACACTCATACCCTTCAACAA
Cadia ATGGGAGG---GTGCAAGTGCAAGACTTGAAGAAAAAGATCCTTGCA---ACTGAAAAC---CCTCAAACTCTGCACCAA
Calpurnia ATGGGAGG---GTGCAAGTGCAAGACTTGAAGAAAAAGGGCCTTGCA---ACTGAAAAC---CCTCAAACCCTGCACCAA
Sophora ATGGGAGG---GTGCAAATGCAAGATTTGAAGAAAAAGTGTGTTGCA---ACTGAAAAC---CCTCAAATCCTG------
Ormosia CTGGGAGG---GTGCAAGTACAAGAGTTGAAGAAAAAGTGCCTTGCA---AATGAATAC---CCTCAAGTCCTGCACCAA
Retama ATGGGAGGGTAGTGCAAGTGCAAGACTTGAAGAAAAAGTGCATTGCAACAACTGAAAACAACACTCATACCCTTCAACAA
Maackia GTGGGAGG---GTGCAAGTGCAAGACTTGAAGAAAAAGTGCCTTGCA---ACTGAAAAC---CCTCAAATCCTGCACCAA
Thermopsis ATGGGAGG---GTGCAAGTGCAAGATTTGAAGAAAAAGTGTCTTGCA---AATGAAAAC---CCACAAATCCTGAACCAA
Genista TTGAGATCACCCCTTCACCTTGAGGACTGTGCAAGATCACCTAATAAGATGATTCACACTCACCCT----------CA
L.densiflorus TTAAGGTCACCCCTTCAGCTTGAAGACTGTGCAAGATCACCTAATAATAAACTT---CTTCACCCT----------
L.digitatus TTGAGGTCACCTCTTCAGCTTGAAGACTGTGCAAGATCACCTAATAATAAGCTT---CTTCACCCT----------
L.nanus TTGAGATCACCTCTTCAGCTTGAAGATTGTGCAAGATCACCTAATAATAAACTT---CTTCACCCT----------
L.angustifolius TTGAGATCACCTATTCGGCTTGAAGAATGTGCAAGATCACCTAATAATAAGCTT---CTTCACCCT----------
Cadia TTTAGGTCACCCCTTCAGCCTGAGGACTGTGCAAGATCACCTAAT---AAGCTG---TTTCACCCTATA------CCT--
Calpurnia TTGAGGTCACCCCTTCAGCCTGAGGACTGTGCAAGATCACCTAAT---AAGCTG---GTTCACCCT----------
Sophora ----------CACCCT----------
Ormosia TTGAGGTCACCAATTCAGCCTGAGGACTGTGCAAGATCACCTAAT---AAGCTG---GTTCACCCT------CACCCT--
Retama TTGAGATCACCCTTTCAGCTTGAGGACTGTGCAAGATCACCTAATAATAAGCTA---CTTCACCCT----------
Maackia TTGAGGTCACCCCTTCAGCCTGAGGACTTTGCAAGATCACCTAAT---AAGCTG---GTTCACCCT------CACCCT--
Thermopsis TTGAGGTCACCCTCA----------TCACCTAAT---AAGCTG---GTTCAACCTCAACCTCACCCTCA


Genista TCATCACCTTGTTGGTATTAGTGAA---GCACCTAGAGATGACAACTTCAATGTGATTGAGGAATCCATTGTGATCAGGA
L.densiflorus ----CACTTTGTGAGTAGTAGTGAA---GTACCTAGAGATGACAACTTCAATGTGATTGAGGAATCAATTGTGATTAGGA
L.digitatus ----CACTTTGTGAGTAGTAGTGAAGTAGTACCTAGAGATGACAACTTCAATGTGATTGAGGAATCCATTGTGATCAGGA
L.nanus ----CACTTT------AGTAGTGAA---GTACCAAGAGATGATAACTTCAATGTGATTGAGGAATCCATTGTTATAAGGA
L.angustifolius ----CACTTTGTGAGTAGTAGTGAA---GTACCTAGAGATGACAACTTCAATGTGATTGAGGAATCCATTGTGATCAGGA
Cadia -CATCACCTTGTG---GGTAGTGAA---GCACCTAGAGATGAC---TTCAACGTGATTGAGGAATCCATTTTGATAAGGA
Calpurnia -CATCACCTTGTG---GGTAGTGAA---GCACCTAGAGATGAC---TTCAATGTGATTGAGGAATCCATTGTGATCAGGA
Sophora -CATCACCTTGTG---AGTAGTGAA---GCACCTAGAGATGAC---TTCAATGTGATTGAGGAATCCATTGTGATCAAAA
Ormosia -CATCACCTTGTG---GGTAGTGAA---GCACCTAGAGATGAC---TTCAATGTTATTGAGGAATCCATTGTGATCAAGA
Retama ----CACTTTGTGAGTAGTAGTGAA---GCACCTGGAGATGACAACTTCAATGTGATTGAGGAATCCATTGTGATCAGGA
Maackia -CATCACCTTGTG---AGTAGTGAA---GCACCTAGAGATGAC---TTCAATGTGATTGAGGAATCCATTGTGATCAAGA
Thermopsis TCATCACTTTGTG---AGCAGTGAA---GGAGCTAGAGATGAC---TTCAATGTGATTGAGGAATCTATTGTGATCAAGA
Genista GAAAGTTGAAGCCTTCAATGATGTCTTCTTCT---------CATCATCATCATCATCACCAAAACCTTATGATCCCAAAG
L.densiflorus GAAAGTTGAAGCCTTCAATGATGTCTTCTTCTTCTTCTTCT---------CATCACCACCAGAACCCAATGATCCCAAAG
L.digitatus GAAAGTTGAAGCCTTCAATGATGTCTTCTTCT----------CATCATCATCACCACCAGAACCCAATGATCCCAAAG
L.nanus GAAAGTTGAAGCCTTCAATGATGTCTTCTTCTTCT---------CATCATCACCATCACCAGAACACAATGATCCCAAAG
L.angustifolius GAAAGTTGAAGCCTTCAATGATGTCTTCTTCT----------CATCATCATCATCACCAGAATCCAATGATCCCAAAG
Cadia GAAAGTTGAAGCCAACGTTGATGTCTTCT----------CATCATCATCACCAAAAACTTGTGATCCCAAAG
Calpurnia GAAAGTTGAAGCCGTCGTTGATGTCTTCT----------CATCATCATCATCATCACCAAAACCTTGTGATCCCAAAG
Sophora GAAAGTTGAAGCCATCGATGATGTCTTCT----------CATCATCATCACCATCACCAAAACCTTGTGATCCCAAAG
Ormosia GAAAGTTGAAGCCGTCGTTGATGTCTTCA----------CATCATCATCACCCAAACCTTGTGATCCCTAAG
Retama GAAAGTTGAAGCCTTCAATGATGTCTTCTTCT----------CATCACCATCAGAACCTAATGATACCAAAA
Maackia GAAAGTTGAAGCCCTCGTTGATGTCTTCT----------CATCATCATCACCAAAATCTTGTGATCCCAAAG
Thermopsis GAAAGTTGAAGCCGTCTTTGATGTCTTCTTCT----------CATCATCATCACCAAAACCTTGTGATCCCAAAG

```
Genista          GAAGCAAGTTTCAACAACAACAACAAC----------ACT---GACTACAACTCCTTCACCAATTTGTCTCCAAATTG
L.densiflorus    GAAACAAGTTTCAACAACAACAACAAC----------ACT---GACTACAACTCCTTCACCAATTTGTCTCCAAATTG
L.digitatus      GAAGCAAGTTTCAACAACAAC----------------ACT---GGCTACAACTCCTTCACCAATTTGTCTCCAAATTG
L.nanus          GAAGCAAGTTTCAACAACAACAACAAC----------AAT---GATTACAACTCCTTCACCAACTTGTCTCCAAATTG
L.angustifolius  GAAGCAAGTTTCAACAACAAC----------------ACT---GACTACAACACCTTCACCAATTTGTCTCCCAATTG
Cadia            GAAGCTAGTTTCAAC--------------------AGCAAT---GACTACCACTCCTTCCCCAATTTGTCTCCAAATTG
Calpurnia        GAAGCAAGTTTCAACAAC----------------AATACT---GACTACCACTCCTTCCCCAATTTGTCTCCAAATTG
Sophora          GAAGCAAGTTTCAACAACAAC----------AATACT---GAATACCACTCCTTCCCCAATTTGTCTCCAAATTG
Ormosia          GAATCAAGTTTCAACAAC----------------AGTACT---GACTACCACTCCTTCCCCAATTTGTCTCCAAATTG
Retama           GAAGCAAGTTTCAACAACAACAACAACAACAACAACACT---GACTACAACTCCTTCACCAATTTGTCTmCAAATTG
Maackia          GAAGCAAGTTTCAACAAC----------------AGTACT---GACTACCACTCCTTCCCCAATTTGTCTCCAAATTG
Thermopsis       GAAGCAAGTTTCAATAAC----------AGCAGTACTACTGAATACCACCCCTTCCCCAATTTGTCTCCAAATTG
```

```
Genista          GAATAATGCTAGTAATGGTGGCAGTGATATTAATGGCAGATCCAACTTTTGTACAATAGCCAGCATGAATCTCTCTACA
L.densiflorus    GGATAAT------AATGGTGGAAATGGTATTAATGGCAGATCCAACATTTGTACAATAGCCAGCATGAATCTCTCTACA
L.digitatus      GGAAAATGCTAATAGTGGTGGCAATGGTATTAATGGCAGATCCAACTTTTGTACAATAGCTAGCATGAATCTCTCCACA
L.nanus          GGAT---------AATGGTGGAAATGGTATTAATAGCAGATCCAACTTTTGTACAATAGCCAGCATGAATCTCTCTACA
L.angustifolius  GGAT---GCT---AATGGTGGCAATGGTATTAATGGCAGATCCAACTTTTGTACAATAGCCAGCATGAATCTCTCTACA
Cadia            GGAT---GCTAATAATGGTACCAATGCC---ACTGGCCGCGCCAACTTTTGTACAATAGCCAGCATGAATCTATCTACA
Calpurnia        GGAT---GCTAATAATGGTACC---------ACTGGTCGCTCCAACTTTTGTACAATATCCAGCATGAATCTATCTACA
Sophora          GGAT---GCTAATAATGCTGCCAGT------ACTAGCCGCTCCAACTTTTGTGCAATAGCCAGCATGAATCTATCTACA
Ormosia          GGAT---GCCAATAATGGTGCCAGTGCC---ACTGGCCGCTCCAACTTTTGTGCAATAGCCAGCATGAATCTATCTACA
Retama           GGATAATGCTAATAATGGTGGCAACGGTATTAATGGCAGATCCAACTTTTGTACAATAGCCAGCATGAATCTCTCTACA
Maackia          GGAT---GCTAATAATGGTGCCAGT---------GGCCGCTCCAACTTTTGTGCAATAGCCAGCATGAATCTATCTACA
Thermopsis       GGAT---GCTAATAATGGTACCAATACC---ACTGGCCGCTCCAACTTTTGTGCAATAGCTAGCATGAATCTATCTACA
```

APPENDIX 8: PUBLICATION

Citerne, H.L., Luo, D., Pennington, R.T., Coen, E., and Cronk, Q.C.B. 2003. A phylogenomic investigation of *CYCLOIDEA*-like TCP genes in the Leguminosae. *Plant Physiol.* 131: 1042-1053.

A Phylogenomic Investigation of *CYCLOIDEA*-Like TCP Genes in the Leguminosae[1]

Hélène L. Citerne*, Da Luo, R. Toby Pennington, Enrico Coen, and Quentin C.B. Cronk[2]

Royal Botanic Garden Edinburgh, 20a Inverleith Row, Edinburgh EH3 5LR, United Kingdom (H.C., R.T.P., Q.C.B.C.); Institute of Cell and Molecular Biology, University of Edinburgh, Edinburgh EH9 3JR, United Kingdom (H.C., Q.C.B.C.); Shanghai Institute of Plant Physiology, Chinese Academy of Sciences, 300 Fenglin Road, Shanghai 200032, China (D.L.); and Genetics Department, John Innes Centre, Colney Lane, Norwich NR4 7UH, United Kingdom (E.C.)

Numerous TCP genes (transcription factors with a TCP domain) occur in legumes. Genes of this class in Arabidopsis (*TCP1*) and snapdragon (*Antirrhinum majus*; *CYCLOIDEA*) have been shown to be asymmetrically expressed in developing floral primordia, and in snapdragon, they are required for floral zygomorphy (bilaterally symmetrical flowers). These genes are therefore particularly interesting in Leguminosae, a family that is thought to have evolved zygomorphy independently from other zygomorphic angiosperm lineages. Using a phylogenomic approach, we show that homologs of *TCP1/CYCLOIDEA* occur in legumes and may be divided into two main classes (*LEGCYC* group I and II), apparently the result of an early duplication, and each class is characterized by a typical amino acid signature in the TCP domain. Furthermore, group I genes in legumes may be divided into two subclasses (*LEGCYC* IA and IB), apparently the result of a duplication near the base of the papilionoid legumes or below. Most papilionoid legumes investigated have all three genes present (*LEGCYC* IA, IB, and II), inviting further work to investigate possible functional difference between the three types. However, within these three major gene groups, the precise relationships of the paralogs between species are difficult to determine probably because of a complex history of duplication and loss with lineage sorting or heterotachy (within-site rate variation) due to functional differentiation. The results illustrate both the potential and the difficulties of orthology determination in variable gene families, on which the phylogenomic approach to formulating hypotheses of function depends.

The considerable advances in plant developmental genetics from a few model species have provided a starting point for studying plant morphological diversity and evolution at the molecular level. Genes that control development have been implicated in the evolution of novel phenotypes (for review, see Baum, 1998; Doebley and Lukens, 1998; McSteen and Hake, 1998; Cronk, 2001; Shepard and Purugganan, 2002). There is now a growing interest in expanding this knowledge to other species less amenable to genetic studies but displaying patterns of morphological variation that could be accounted for by changes in the expression of developmental genes.

Comparative expression studies rely on a phylogenetic framework to help identify candidate genes (Eisen, 1998). This approach has been used to find putative orthologs of MADS-box genes in non-model species of basal eudicots (Kramer and Irish, 1999). We present here a study of the evolution of putative homologs of the floral symmetry gene *CYCLOIDEA* (*CYC*) in legumes, with particular emphasis on the subfamily Papilionoideae. Using relatively wide sampling within Leguminosae is potentially a useful way of identifying the different subgroups within a gene family, as represented in legumes.

In snapdragon (*Antirrhinum majus* L. [Lamiales, Veronicaceae]), floral dorsal identity is controlled by two closely related nuclear genes *CYC* and *DICHOTOMA* (*DICH*; Luo et al., 1996, 1999; Almeida et al., 1997). In floral meristems, *CYC* and *DICH* have overlapping expression patterns on the adaxial side, but they have diverged so that *CYC* is expressed slightly later in development than *DICH* but has a greater effect on phenotype. These two genes belong to a gene family of putative transcription factors characterized by a basic helix-loop-helix domain referred to as the TCP domain (Cubas et al., 1999a). In Arabidopsis, 24 members have been identified. A subclass of this gene family, to which *CYC/DICH* and the maize (*Zea mays*) architecture gene *TEOSINTE BRANCHED 1* belong, also has a highly conserved Arg-rich R domain (Cubas, 2002). CYC-like genes have been implicated in the control of floral symmetry in other species in the Lamiales, such as *Linaria vulgaris* Miller (Veronicaceae; Cubas et al., 1999b). The homolog of CYC in Arabidopsis, TCP1, has recently been shown to be expressed transiently at the adaxial base of floral and axillary meristems (Cubas

[1] This work was supported by The Carnegie Trust for the Universities of Scotland and by the Systematics Association.

[2] Present address: Botanical Garden and Centre for Plant Research, University of British Columbia, 6804 Southwest Marine Drive, Vancouver, BC, Canada V6T 1Z4.

* Corresponding author; e-mail h.citerne@rbge.org.uk; fax 44-131-248-2901.

Article, publication date, and citation information can be found at www.plantphysiol.org/cgi/doi/10.1104/pp.102.016311.

et al., 2001). This suggests that asymmetric expression of *CYC*-like genes may predate the divergence of the Asteridae (e.g. snapdragon) and the Rosidae (e.g. Arabidopsis and Leguminosae). Such asymmetrically expressed genes may have been recruited repeatedly for the evolution of zygomorphy in separate lineages.

The Leguminosae is one such plant family where zygomorphy is believed to have evolved separately from the Lamiales (Stebbins, 1974; Donoghue et al., 1998). With approximately 18,000 species, it is one of the most species-rich angiosperm families, with the greatest number of species (approximately 12,000) found in the subfamily Papilionoideae. Papilionoids are characterized by highly zygomorphic flowers, with an enlarged dorsal (standard) petal, and lateral (wings) and ventral (keel) petals surrounding the reproductive organs. This highly specialized floral form, an adaptation to bee pollination, contrasts with that of the other two subfamilies Caesalpinioideae and Mimosoideae. Mimosoid flowers are typically actinomorphic, with reduced outer whorls, whereas Caesalpinioideae display more variation in floral morphology ranging from near radial symmetry to zygomorphy. Current molecular evidence suggests that mimosoids and papilionoids have evolved from different lineages of a paraphyletic caesalpinioid group (Doyle et al., 1997; Bruneau et al., 2001; Kajita et al., 2001; Fig. 1).

Within the Papilionoideae, a few taxa with atypical near radial symmetry have traditionally been considered basal members of this subfamily, even transitional between caesalpinioids and papilionoids (Polhill, 1981). However, recent molecular evidence suggests that these unusual taxa are derived from typical papilionoids (Pennington et al., 2000). These putative reversals from zygomorphy to actinomorphy provide a framework for studying the control of floral symmetry in legumes.

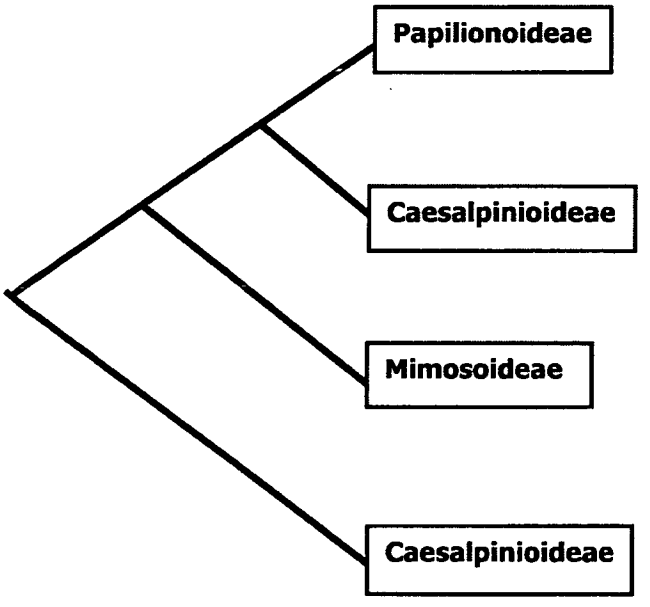


Figure 1. Phylogenetic relationship of the three legume subfamilies based on current molecular evidence, with Mimosoideae and Papilionoideae derived from a paraphyletic Caesalpinoideae (Doyle et al., 1997; Bruneau et al., 2001; Kajita et al., 2001).

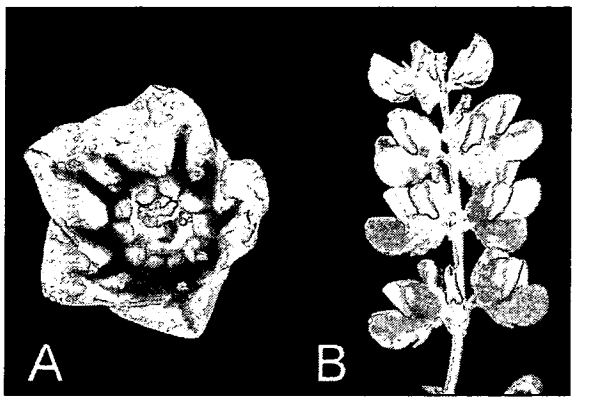


Figure 2. a, Flower of *Cadia purpurea*, a near actinomorphic papilionoid legume. b, Inflorescence of *Lupinus nanus* bearing highly zygomorphic flowers typical of the Papilionoideae.

In the model legumes *Lotus japonicus*, soybean (*Glycine max*), and pea (*Pisum sativum*), *CYC*-like genes have been isolated, and in the case of *L. japonicus*, two genes have been found to be asymmetrically expressed in the developing flower (D. Luo, unpublished data). This study aims to expand these findings to other taxa from other major papilionoid groups such as the dalbergioid and genistoid clades as well as basal lineages (Pennington et al., 2001) where most of the morphological variation lies. This study comprises species with unusual flower morphology, such as *Acosmium subelegans* (Mohl.) Yakovlev and *Cadia purpurea* (Picc.) Aiton from the Genistoid clade, and *Swartzia jorori* Harms from the basal papilionoid group (Polhill, 1981; Pennington et al., 2001). *C. purpurea*, in particular, has open, near radial flowers, with equal free stamens arranged in a ring (Fig. 2a). This contrasts with typical papilionoids from the Genistoid group such as *Lupinus* (*Lupinus nanus*; Fig. 2b). Inclusion of legumes with unusual floral morphology is likely to be useful in studies of the origin of derived modifications in floral symmetry.

As functional gene studies expand from model organisms to related species, it becomes necessary to identify the functional counterparts of genes well-characterized in model species. The phylogenomic method proposes that orthology (i.e. common descent) is a likely predictor of functional equivalence (Eisen, 1998). Modern phylogenetic techniques now often permit robust determination of orthology relations of genes. We have thus taken a phylogenetic approach to investigate orthologs of *CYC* in legumes, with sampling that ensures coverage of all the main clades of papilionoid legumes (Fig. 3).

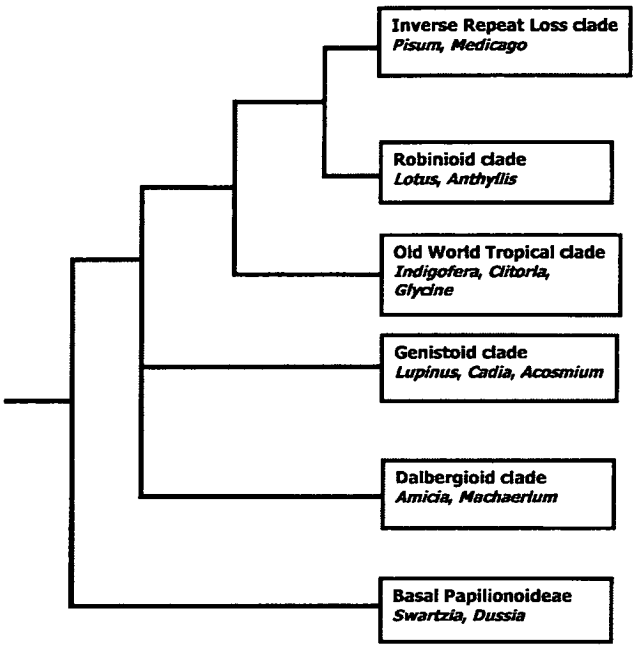


Figure 3. Schematic representation of the relationship of some of the major groups in the Papilionoideae as defined by current molecular evidence (Doyle et al., 1997; Hu et al., 2000; Kajita et al., 2001; Pennington et al., 2001), with representative taxa used in the *LEGCYC* analyses.

RESULTS

Legume *CYC* Sequence Characterization

Thirty-eight sequences with a TCP and R domain were amplified using primers LEGCYC/F1 and R1 in 16 different taxa. Sequence number per taxon ranged from one to four, with only one sequence isolated from non-papilionoid taxa. However, basal papilionoid taxa, such as *S. jorori* and *Dussia macroprophyllata* Harms, had multiple copies comparable in number with more derived papilionoid species (see Table I for summary and GenBank accession no.). No evident sequence modifications (e.g. premature stop codons) were observed in papilionoids with unusual floral morphology.

Fragment length ranged from 274 bp (Pisum 1) to 427 bp (Clitoria 1), with a mean length of 333.81 (± 40.2) bp. These fragments were also highly variable in sequence (at the amino acid and nucleotide level), with numerous substitutions and indel events in the region between the TCP and R domain. As a result, unambiguous sequence alignment for all legume *CYC*-like sequences was only possible in the TCP and R domains.

Position of Legume CYC-Like Sequences in the TCP Gene Family

TCP domains of seven legume CYC-like protein sequences from two species, *C. purpurea* and *L. japonicus*, were analyzed in the context of the TCP gene family. Analysis of the TCP domain peptide matrix using protein distance, parsimony, maximum likelihood (ML), and Bayesian methods resulted in congruent trees with strong support values for the major groups. Figure 4 shows the protein ML unrooted phylogram, with support values obtained by Bayesian analysis of the data. The 50% majority rule (MR) protein distance and maximum parsimony trees are also shown for comparison (Figs. 5 and 6, respectively). All analyses strongly suggest that the TCP gene

Table I. *List of sequences obtained with primers LEGCYC-F1 and R1, and corresponding GenBank accession number*

Sequence	GenBank Accession No.	Sequence	GenBank Accession No.
Ceratonia 1	AY225810	Lupinus sp. 1	AY225832
Dialium 1	AY225811	Lupinus sp. 2	AY225834
Zapoteca 1	AY225812	Lupinus sp. 3	AY225833
Pisum 1	AY225813	Lupinus sp. 4	AY225835
Anthyllis 1	AY225814	Lupinus nanus 1	AY225836
Anthyllis 2	AY225815	Lupinus nanus 2	AY225837
Anthyllis 3	AY225816	Lupinus nanus 3	AY225838
Lotus berthelotii 1	AY225817	Lupinus angustifolius 1	AY225839
Lotus berthelotii 2	AY225818	Lupinus angustifolius 2	AY225840
Indigofera 1	AY225819	Machaerium 1	AY225841
Indigofera 2	AY225820	Machaerium 2	AY225842
Indigofera 3	AY225821	Amicia 1	AY225843
Clitoria 1	AY225822	Amicia 2	AY225844
Clitoria 2	AY225823	Dussia 1	AY225845
Clitoria 3	AY225824	Dussia 2	AY225846
Cadia 1	AY225825	Dussia 3	AY225847
Cadia 2	AY225826	Swartzia 1	AY225848
Cadia 3	AY225827	Swartzia 2	AY225849
Cadia 4	AY225828	Swartzia 3	AY225850
Acosmium 1	AY225829		
Acosmium 2	AY225830		
Acosmium 3	AY225831		

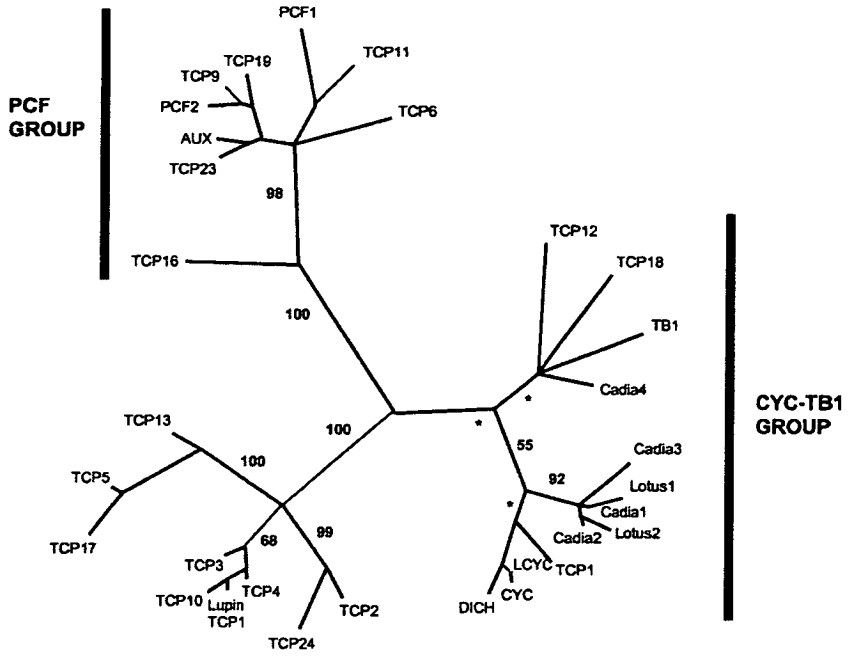


Figure 4. Unrooted phylogram of protein ML analysis using TREEPUZZLE v5.0 (Schmidt et al., 2000) of the TCP domain data set including representative legume sequences. Support values were obtained using MrBayes (Huelsenbeck and Ronquist, 2001); asterisks indicate that a clade was recovered in <50% of Bayesian trees. Results support a LEGCYC clade (excluding Cadia 4) as sister to the CYC/TCP1 clade. All TCP genes unless otherwise indicated, Arabidopsis; PCF, rice; TB1, maize; LCYC, *L. vulgaris*; CYC and DICH, snapdragon; AUX, cotton.

family can be divided into two main groups: the PCF group (recovered in every analyses with 100% support values) and a second group containing CYC/TB1 and, among others, the five Arabidopsis genes (TCP1, TCP12, TCP18, TCP2, and TCP24) with an R domain. These results confirm the conclusions of Cubas (2002), but with greater sampling and more comprehensive phylogenetic analysis. Within the latter group, CYC/TB1 genes form a separate group from another well-supported clade (in all analyses) of yet uncharacterized proteins. Although unrooted trees are difficult to interpret evolutionarily, because the point of origin is uncertain, these trees strongly suggest that the legume sequences here are the best candidates for CYC/TCP1 orthologs.

All analyses suggest that the legume CYC (LEGCYC) sequences from *C. purpurea* and *L. japonicus* (with the exception of Cadia 4) form a strongly supported group (found in 92% of Bayesian trees). This monophyletic group (here called LEGCYC) is sister to the CYC-TCP1 clade in the ML, Bayesian (Fig. 4) and distance (Fig. 5) trees. LEGCYC genes are therefore putative orthologs of CYC and TCP1. Cadia 4 is recovered in ML (Fig. 4) and distance (Fig. 5) analyses in the clade containing TB1, TCP12, and TCP18. The parsimony analysis is not informative because the relationship between the LEGCYC clade, Cadia 4, the CYC/LCYC/DICH clade, TCP1, TCP12, TCP18, and TB1 collapses in a 50% MR consensus tree (Fig. 6).

Evolution of *LEGCYC* Genes: Partial TCP and R Nucleotide Analyses

To recover major groups within the *LEGCYC* genes, we analyzed a matrix of 29 legume nucleotide sequences, rooted using snapdragon *CYC* and *DICH*, chosen to represent the full range of papilionoid legume taxa and sequence variation. The legume sequences could only be aligned with the snapdragon sequences using the highly conserved TCP and R domains. Parsimony analysis of the 67 informative sites out of 145 in the partial TCP and R nucleotide sequences produced 168 trees with a minimal length of 278 steps (additional branch swapping did not recover any more maximum parsimony trees), a consistency index (CI) of 0.424 and a retention index (RI) of 0.636, indicating fairly high homoplasy (parallel evolution) in the data. A strict consensus tree (Fig. 7), rooted on snapdragon genes *CYC* and *DICH*, resolves only one large supported clade within the ingroup (corresponding to group II, see below). Otherwise, only the relationship between sequences from different species of the same genus (e.g. *Lupinus* spp.) or related genera (e.g. *Anthyllis* and *Lotus* spp.) were supported in this analysis.

Model-based methods, such as Bayesian inference, are less sensitive to long-branch attraction and may therefore be better alternatives for analyzing homoplastic data. Bayesian analysis (Fig. 8) recovered two groups of legume sequences with support values (called here group I and group II). Group II had very high (97%) Bayesian support, whereas group I had weak support of 52%. Both groups include species from basal as well as more derived papilionoids and would appear to represent an early duplication event. However, relationships between sequences other than from closely related species or genera (e.g. *Lupinus* spp.) were difficult to interpret.

Therefore although parsimony analysis of this small data set did not resolve relationships between

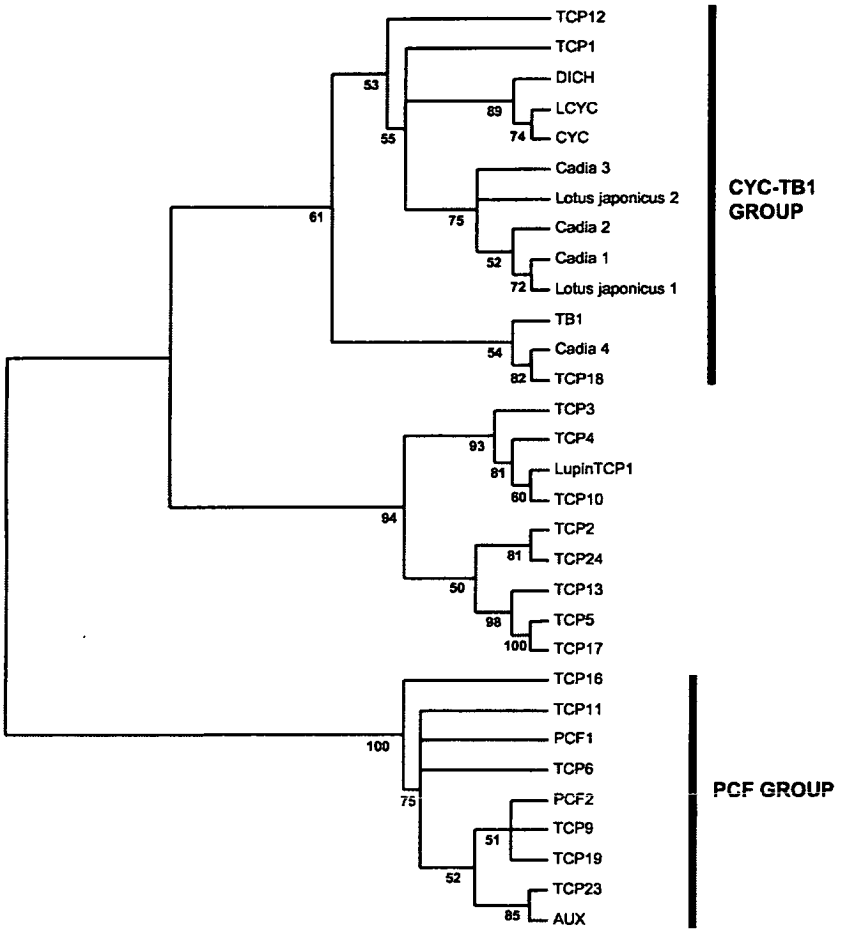


Figure 5. Fifty percent MR consensus tree of the protein distance analysis using the PAM-Dayhoff model of protein substitution (PROTDIST; Felsenstein, 1993) of the TCP domain. Values >50% of the 100 jackknife replicates are given at branch nodes. Taxa as in legend to Figure 4.

LEGCYC genes well, Bayesian analysis gave a more fully resolved tree. The poor performance of parsimony analysis was probably due to high homoplasy in the data set coupled with the low number of informative characters with consequent low phylogenetic signal.

Evolution of LEGCYC Genes: Inclusion of Sequence Data between the TCP and R Domains

The region between the TCP and R domains was then added to the initial data set, together with additional legume sequences. Due to the high length and sequence variability of this region, it could not be aligned with nonlegume sequences, and so all analyses are unrooted. Furthermore, because of length variability, alignment was difficult even within legumes. For this reason some of the positions in which the alignment was ambiguous were excluded from the analysis (300 aligned positions). Eight *LEGCYC* sequences were excluded altogether from this analysis for the same reason. The remaining 38 sequences covered 292 unambiguously aligned characters, which required the insertion of 34 gaps of 1- to 6-bp triplets for alignment.

Parsimony analysis of the resulting 153 parsimony informative characters from the extended data set resulted in a single most parsimonious tree of 748 steps, with CI = 0.452 and RI = 0.601. The tree recovered two clades (groups I and II from the previous analyses) with a bootstrap value of 65%, although sequence relationship within these groups had little bootstrap support with the exception of sequences from closely related taxa (Fig. 9). The topology of the ML tree and the 50% MR consensus tree from the Bayesian analysis was identical, with only three nodes collapsing in the Bayesian consensus tree. The topology of those trees was also similar to the tree from the parsimony analysis, but the level of support for the nodes (estimated by Bayesian inference) was much higher in the model-based analysis. For instance, group I and II were recovered in the Bayesian analysis with high support (Fig. 10). Comparison of the partial TCP domains of amino acid sequences from group I and II showed that there were five synapomorphies, which suggests these clades are genuine (Fig. 11). These groupings were also supported by considerable differences in the variable region, such as presence or absence of motifs, which could not be included in the analysis.

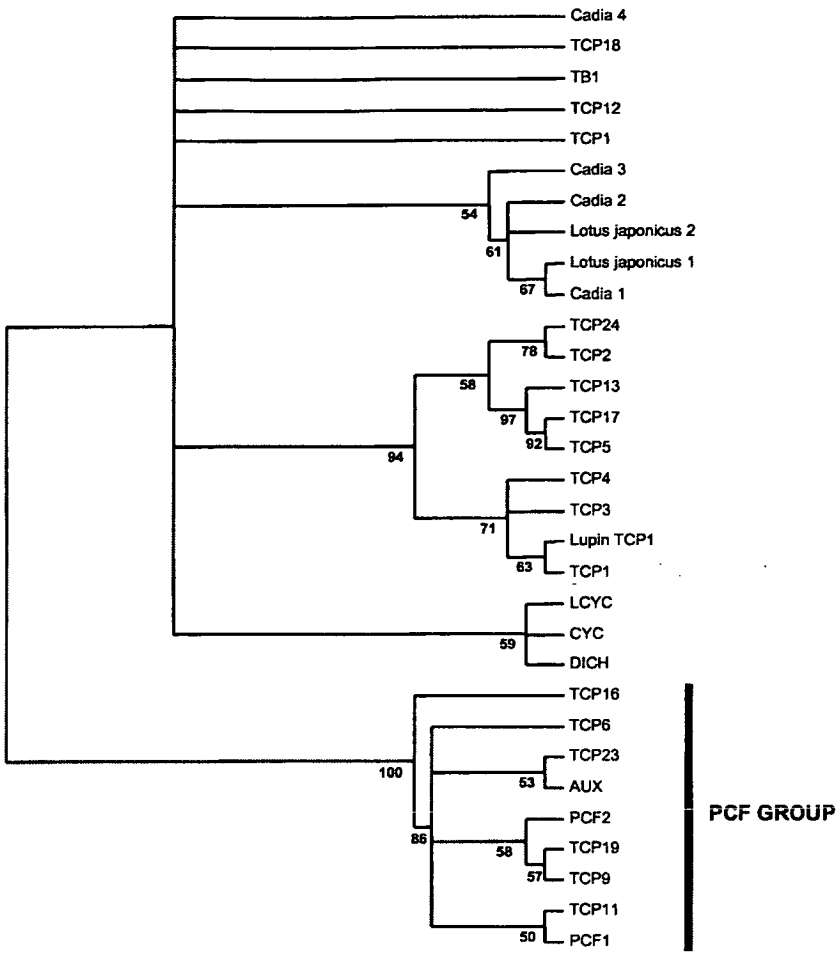


Figure 6. Fifty percent MR consensus tree of protein maximum parsimony analysis (PROTPARS; Felsenstein, 1993) of the TCP domain. Support values above 50% from the 100 jackknife replicates are shown. Maximum parsimony fails to resolve groups recovered in protein, ML, Bayesian, and protein distance analyses. Although it does not contradict any of the results from other methods, it offers no support for a CYC/TB1 clade.

Within group I, two sequences from most taxa were found. These segregated into two clades (A and B, see Fig. 9), which for the most part contained one sequence per taxon, with a few exceptions (for example Machaerium 1 and 2). Clade A contained one *LEGCYC* sequence from representatives from both the genistoid (*Lupinus* spp., *Cadia* sp., and *Acosmium* spp.) and robinioid (*Lotus* spp. and *Anthyllis* sp.) clades, whereas clade B contained another *LEGCYC* sequence from these taxa. Although these clades have no bootstrap support in the parsimony analysis, they were found the ML tree and in most Bayesian trees. This suggests a putative orthology relationship between sequences within these clades (IA and IB) and a further conserved duplication in *LEGCYC* sequences (*LEGCYC* IA and IB) of possible functional significance.

DISCUSSION

Presence of TCP1/CYC Orthologs in Leguminosae

In the TCP gene family analyses, evidence from sequence similarity (PROTDIST) and evolution (ML and Bayesian analyses) strongly suggest that the legume *CYC*-like sequences examined here are homologous to the floral symmetry genes in snapdragon, *CYC* and *DICH*, and to the adaxially expressed floral gene *TCP1* in Arabidopsis. Within this legume clade, a lower estimate of three *CYC*-like copies were found within the Papilionoideae, in species ranging from the basal-most clade (*S. jorori*) to higher papilionoids (e.g. the robinioid *A. hermannia*). Because of their apparent orthology with snapdragon *CYC*, these genes are candidates for floral developmental genes in the Leguminosae. However, these analyses, many of which lead to poorly resolved trees, highlight some of the difficulties in making detailed orthology statements within gene families and *CYC*-like genes in particular.

Complex Evolution of CYC-Like Genes in the Leguminosae

No simple pattern of gene evolution tracking organismal phylogeny within the legume *CYC* family was recovered in the phylogenetic analyses. Possible confounding factors such as intermediate levels of concerted evolution, variation in the rate of sequence

evolution, and independent gene loss and duplication events, which render the interpretation of gene trees difficult (Doyle, 1994), cannot be ruled out here.

Different levels of variation in different parts of the sequences also made analysis difficult. The highly conserved TCP and R boxes were alignable but contained little phylogenetically informative information, whereas the variable region contained much variation but was difficult to align. Furthermore, the variation in the TCP and R domains was mainly at the synonymous third codon position and had a high degree of homoplastic variation (accounting for two-thirds of the steps required). High levels of homoplasy, possibly resulting in long-branch attraction and therefore artificial groupings, is suggested by the low support values of the trees from this analysis and the collapse of many nodes in the maximum parsimony strict consensus trees. Also, because the analysis includes clades between which functional differentiation may exist, particular amino acid positions may be subject to different selection pressure in different parts of the tree. This within-site rate variation, or heterotachy (Lopez et al., 2002), is also likely to make phylogenetic reconstruction more difficult.

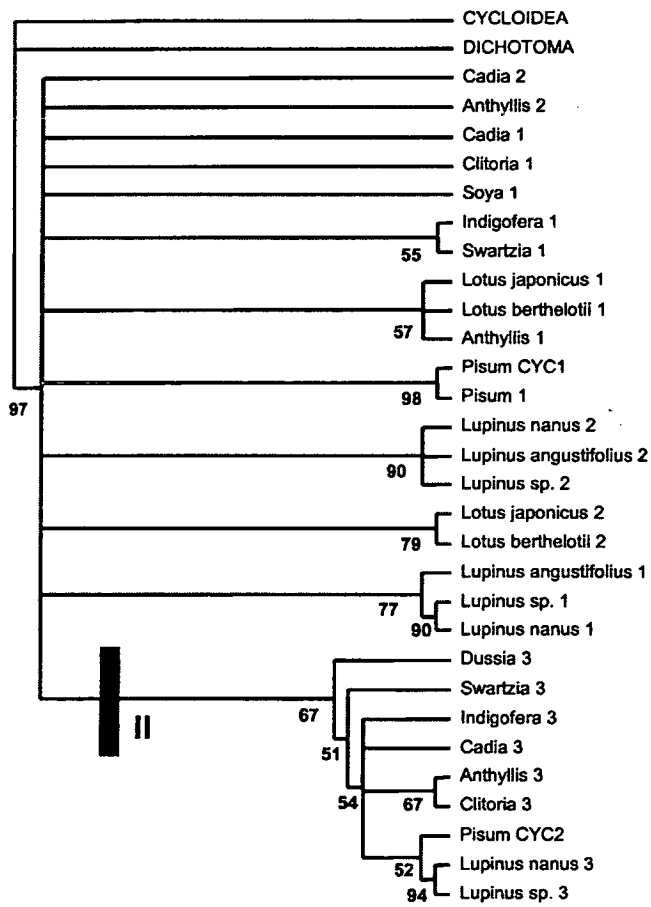


Figure 7. Maximum parsimony analysis of the legume partial TCP and R domain nucleotide sequences. Strict consensus of 168 most parsimonious trees (CI = 0.424, RI = 0.636), with bootstrap values shown, rooted on snapdragon *CYC* and *DICH*.

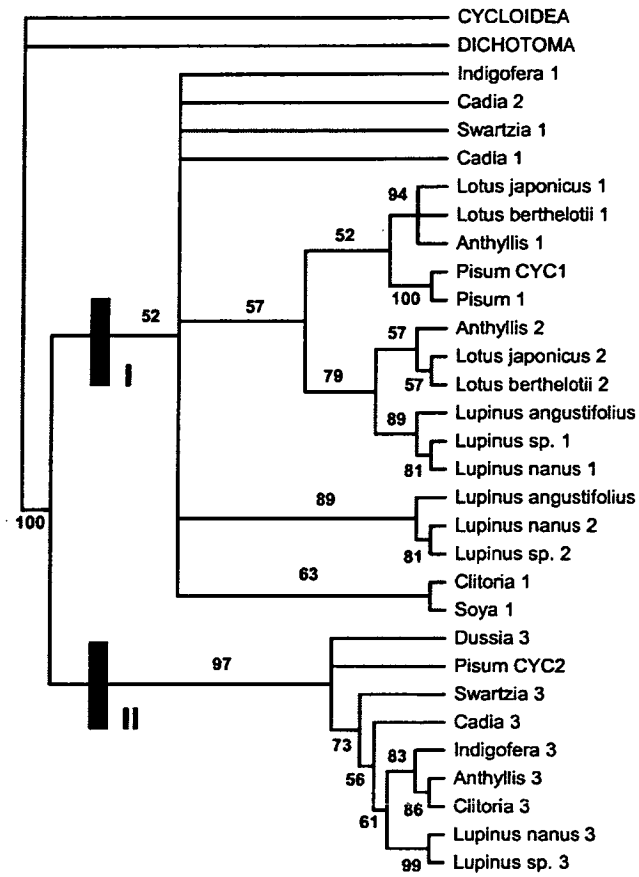


Figure 8. Bayesian analysis MR tree of the legume TCP and R nucleotide sequences allowing for codon-specific nucleotide substitution, rooted on snapdragon *CYC* and *DICH*. Major clades I and II within *LEGCYC* are indicated with high Bayesian support.

Two Major Subgroups (I and II) of Legume *CYC*-Like Genes Represent a Probable Early Duplication

Despite the problematic nature of the data, certain patterns do emerge from the analyses. Results of the rooted Bayesian analysis suggests that *LEGCYC* genes can be divided into two main groups (referred to as I and II), which are characterized by different amino acid signatures. The results of the unrooted legume analyses of the extended dataset are also consistent with the two-group hypothesis, and these groups, although only moderately supported by maximum parsimony, are strongly supported by Bayesian inference. Taxa ranging from the basal-most papilionoids to highly derived species (from the "inverse repeat loss" clade, e.g. pea) have both groups of genes suggesting that these genes probably diverged after a duplication event before the evolution of the Papilionoideae. In addition to the putative amino acid synapomorphies in the TCP domain (Fig. 11),

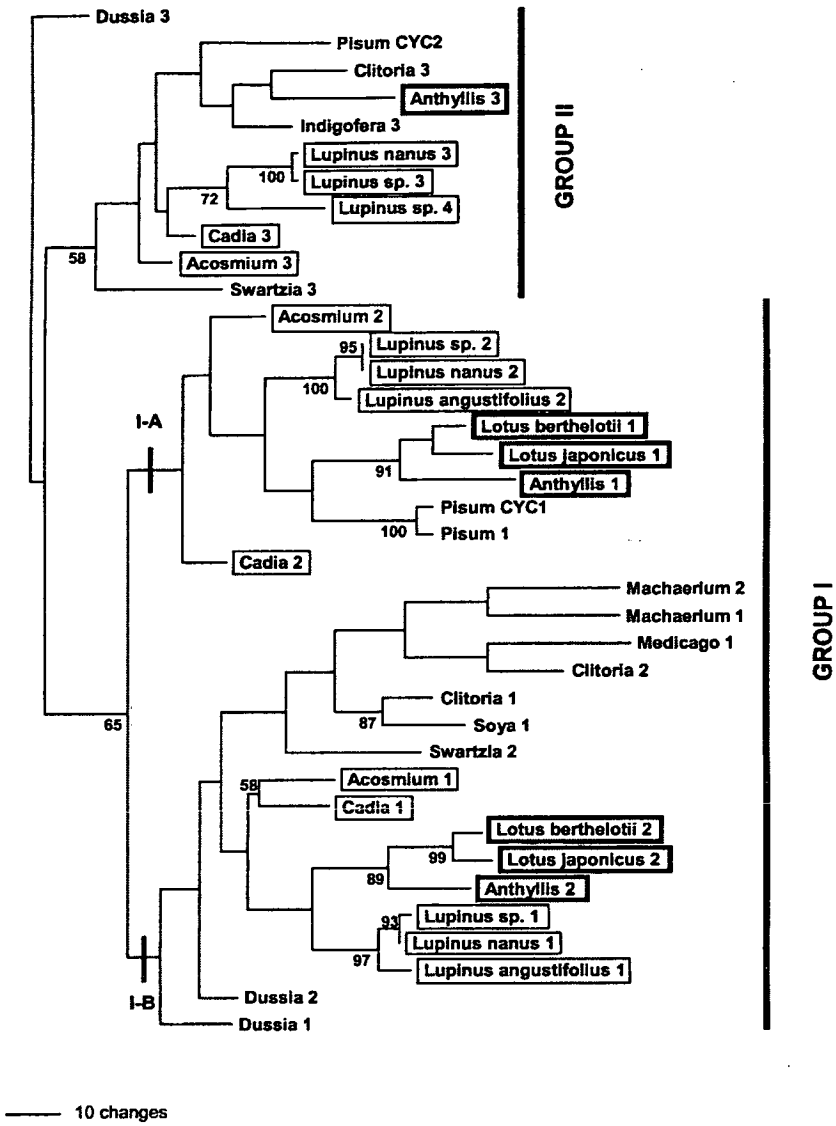


Figure 9. Unrooted phylogram of single most parsimonious tree (748 steps, CI = 0.452, RI = 0.601) from the maximum parsimony analysis of 38 partial legume CYC-like sequences including some sequence data (292 characters, 153 parsimony informative) from the hypervariable region between the TCP and R domains. Bootstrap values (below in bold) are given for branches with >50% support. Major groups recovered in previous analyses (group I and group II) are shown. Clades containing *Lupinus* spp. and *Lotus* spp. sequences are highlighted (I-A and I-B) suggesting putative duplication events.

these groups are also distinguished by specific motifs in the otherwise variable region between the TCP and R domains.

Evidence for Two Subgroups (IA and IB) of Group I *LEGCYC* Sequences

Within group I, one other major duplication event appears to have occurred, giving rise to two subgroups IA and IB. We recovered genes belonging to both clades in a wide range of the species sampled here, implying that this duplication occurred at least early in the diversification of the papilionoids.

However, the relationships between sequences within these groups appear complex and require further investigation. Even though our sampling is fairly extensive compared with many studies of developmental gene phylogeny, further sampling may help resolve relationships within and between gene copies. However, these results are in agreement with a trend of independent duplications, and possible losses, with rapid gene evolution outside of the conserved TCP and R domains, previously documented in *CYC*-like genes families from other plant groups (e.g. Gesneriaceae; Citerne et al., 2000).

The Limitations and Potential of Phylogenomics

The lack of resolution resulting from problematic analyses (particularly using parsimony) highlights the limitations of phylogenomics, at least in rapidly evolving genes with high levels of homoplasy and in gene families where functional differentiation may lead to high levels of heterotachy (within-site rate variation). These problems may lead to difficulties in robust orthology estimation and hence functional prediction. In this study, Bayesian inference gives better resolution than parsimony; with the large

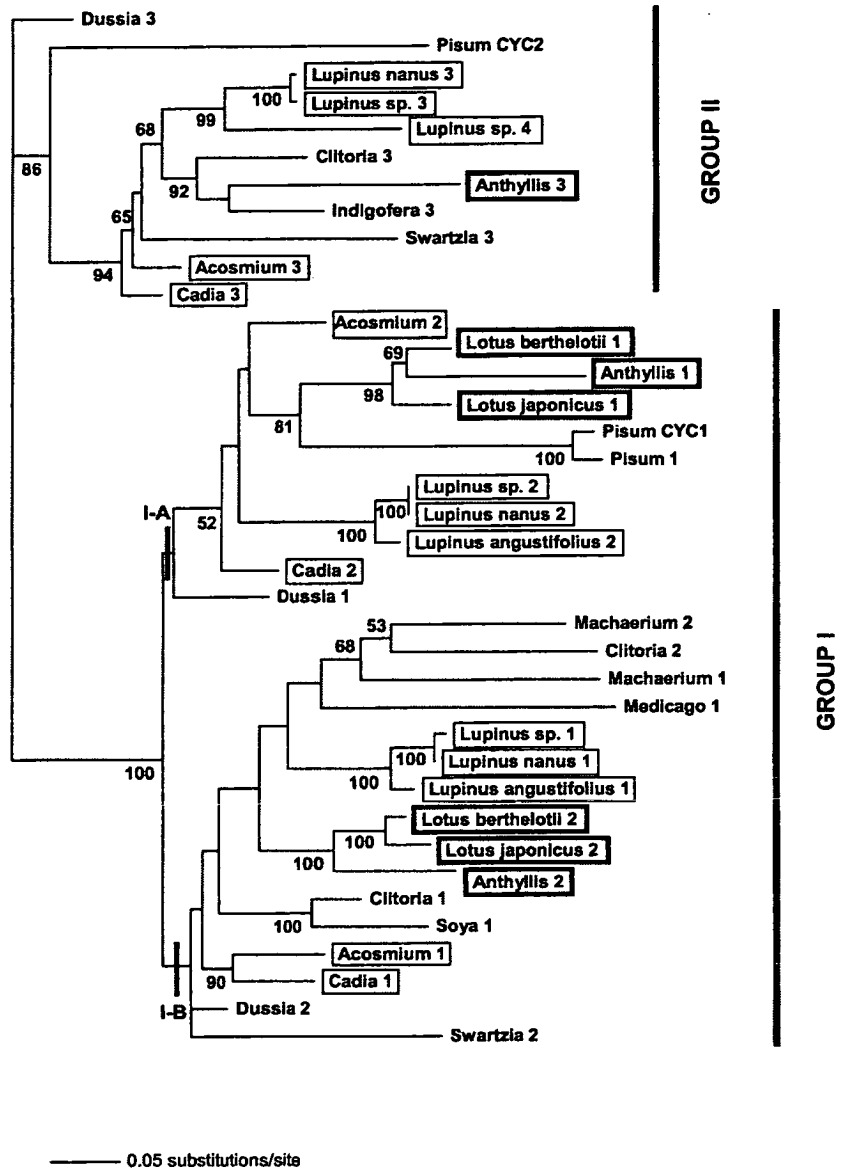


Figure 10. Unrooted phylogram of the ML analysis (using the GTR + I + G model of nucleotide substitution) of partial legume *CYC* sequences. Support values at each node were obtained by Bayesian analysis of the data set and represent the frequency of each node in the MR consensus tree. The two main groups of *LEGCYC* (I and II) are highlighted, and one putative duplication event in group I is marked by A and B.

amount of homoplasy in these data it is likely that model-based methods such as Bayesian inference will outperform parsimony.

The recognition of a major legume *CYC*-like (*LEGCYC*) group in this study does however suggest likely candidate genes for functional equivalents of *CYC/TCP1*. Furthermore, within this group of legume *CYC* candidates, further subgroups are recognized in this study (*LEGCYC* IA, IB, and II), inviting investigation of possible functional differences between these. Thus even where phylogenetic analyses are difficult, partial resolution may still enable hypotheses to be generated. Although we recognize the limitation of phylogenomics, we still regard this approach as extremely promising even with relatively intractable gene families.

MATERIALS AND METHODS

Molecular Methods: DNA Extraction, PCR, Cloning, and Sequencing

For each species, genomic DNA was extracted from either fresh or silica dried leaf material following a modification of the cetyl-trimethyl-ammonium bromide procedure of Doyle and Doyle (1987). Previously extracted DNA was available for *Dialium guinanense* (R.T. Pennington, Royal Botanic Garden Edinburgh), pea (line 399; J. Hofer, John Innes Centre), and *Lupinus angustifolius* cv Merrit (S. Barker, University of Western Australia, Perth).

The region delimited by the conserved TCP and R domains was amplified using primers LEGCYC/F1, 5′-TCA GGG SYT GAG GGA CCG-3′, and LEGCYC/R1, 5′-TCC CTT GCT CTT GCT CTT GC-3′. These primers were designed based on available sequences of CYC-like genes from *Lotus japonicus* and soybean (*Glycine max*; D. Luo, unpublished data), compared with nucleotide sequences of the TCP and R domains from snapdragon (*Antirrhinum majus*; *CYC*, Y16313; and *DICH*, AF199465), Arabidopsis (*TCP1*, AC002130; *TCP12*, AC011914; and *TCP18*, AP001303) and maize (*Zea mays* subsp. *mays*; *TB1*, AF340199). PCR amplifications were carried out using *Taq* and reagents

```
                  A                 C          DE
             W S  D      N          R   D   T  NR MED
GROUP I     RVRLSIEIARKFFDLQDMLGFDKASNTLEWLFNKSKKAIKEL
                 *             *   *           **
GROUP II    RVRLSSEIARKFFDLQDMLEFDKPSNTLEWLFTKSENAIKEL
                   NDV        E  DV          LA  DT
                    Q            QY           N   S
```

Figure 11. Comparison of the partial TCP domain amino acid sequence from group I and II *CYC*-like sequences in legumes. Asterisk highlights group-specific changes; asterisks above and below bold sequences are amino acid differences found less frequently in these groups.

(Bioline, London) in a 50-μL mix containing 2.5 μL of 50 mM $MgCl_2$, 5 μL of a 2 mM dNTP mix, 2.5 μL of each primer (10 μm; MWG Biotech, Gersberg, Germany), 1 unit of BIOTAQ, and 10 to 20 ng of DNA. Conditions consisted of an initial denaturation step at 94°C (3 min), followed by 30 cycles of denaturation at 94°C (1 min), annealing at 50°C to 55°C (30 s), and extension at 72°C (30 s), followed by a final extension step at 72°C (5 min). PCR products were purified using the QIAquick PCR Purification Kit (Qiagen Ltd, Dorking, Surrey, UK) and then cloned using TOPO-TA Cloning Kit for Sequencing (Invitrogen, Carlsbad, CA). Dye-terminator cycle sequencing was carried out using Thermosequenase II (Amersham Biosciences UK, Little Chalfont, Buckinghamshire, UK). Samples were analyzed on an ABI 377 Prism Automatic DNA Sequencer (Applied Biosystems, Foster City, CA). In taxa of particular interest (*Cadia purpurea* and *Lupinus nanus*), 36 to 39 clones were sequenced, respectively. In addition, the entire open reading frame of two gene pairs in *C. purpurea* and *L. nanus* was sequenced by genome walking (modified from Siebert et al., 1995).

Phylogenetic Analysis: Taxon and Sequence Selection

CYC-like genes from legumes were placed in the context of the TCP gene family, represented by certain key sequences from *L. japonicus* and *C. purpurea* (Lotus japonicus 1 and 2, Cadia 1–4; Table I). To simplify the analysis, certain Arabidopsis TCP genes belonging to the PCF group (Cubas, 2002) were excluded (*TCP7, TCP8, TCP14, TCP15, TCP20, TCP21*, and *TCP22* following the nomenclature of Cubas [2002]), whereas other sequences of particular interest were added: *Gossypium hirsutum AUXIN* (AF165924), *Lupinus albus 'TCP1 '* (AJ426419), *Linaria vulgaris LCYC* (AF161252), and snapdragon *DICH* (AF199465). The 58 amino acids of the TCP domain were aligned manually. The matrix of 31 sequences was analyzed using not only protein distance methods similar to that of Cubas (2002), but also maximum parsimony, ML, and Bayesian methods (see below).

Results from these analyses guided the choice of sequences sampled to investigate the evolution of *CYC*-like genes in the legume family, using nucleotides of the TCP and R domains, with *CYC*, *DICH*, and *TCP1* as outgroups. Twenty-nine taxa were sampled to represent the phylogenetic range of the papilionoids.

For the detailed analysis within the legumes including the nucleotide region between the TCP and R domains, a larger number of species was used, with representatives from the three subfamilies Caesalpinioideae, Mimosoideae, and Papilionoideae (Table II). Particular emphasis was placed on sampling representatives from all major papilionoid groups defined by current molecular phylogenetic evidence (Doyle et al., 1997; Hu et al., 2000; Kajita et al., 2001; Lavin et al., 2001; Pennington et al., 2001; M. Wojciechowski, M. Lavin, and M. Sanderson, unpublished data; Fig. 3, names of groups follow [Pennington et al., 2001]). All legume sequences obtained with primers LEGCYC/F1-R1, with the exception of Cadia 4, were selected as the ingroup. Additional legume sequences from separate studies were included in this analysis: *L. japonicus* (Lotus japonicus 1, Lotus japonicus 2),

Table II. *Species used in survey of CYC-like genes using primers LEGCYC-F1 and R1*

Relationship of major Papilionoid clades (from Doyle et al., 1997; Hu et al., 2000; Kajita et al., 2001; Pennington et al., 2001) given in figure 3. –, XXX.

Subfamily	Clade	Taxon	Source[a]	Location
Caesalpinioideae	.	*Ceratonia oroethauma* (Hillc.) Lewis & Verdc.	1996 0942A	Oman
		Dialium guianense (Aubl.) Sandw.	R.T. Pennington 639	Napo, Ecuador
Mimosoideae		*Zapoteca tetragona* (Willd.) H.M. Hernandez	1999 1149	Guatemala
Papilionoideae	Inverse Repeat Loss clade	Pea (*Pisum sativum*) line 399	–	cultivated, John Innes Centre, Norwich, UK
	Robinioid clade	*Anthyllis hermanniae* L.	1975 1501	Mediterranean
		Lotus berthelotii Masf.	1978 0702B	Canary Islands
	Old World Tropical clade	*Indigofera pendula* Franch.	1991 0547A	China
		Clitoria sp.	R.T. Pennington 990	San Martín, Peru
	Genistoid clade	*Cadia purpurea* (Picc.) Aiton	1994 2001A	Yemen
		Acosmium subelegans (Mohl.) Yakovlev	Bridgewater 358	Mato Grosso do Sul, Brazil
		Lupinus sp.	R.T. Pennington 815	Piura, Peru
		L. nanus Doug. ex Benth.	–	commercial seed (Sutton Seeds, Paignton, Devon, UK)
		Lupinus angustifolius L. cv Merrit	–	cultivated, University of Western Australia, Perth
	Dalbergioid clade	*Machaerium scleroxylon* Tul.	1999 0888A	Brazil
		Amicia glandulosa Kunth	R.T. Pennington 654	Loja, Ecuador
	Basal Papilionoideae	*Dussia macroprophyllata* Harms	1995 1539A	Heredia, Costa Rica
		Swartzia jorori Harms	R.T. Pennington 938	Santa Cruz, Bolivia

[a]Source number refers to either RBGE living collection number (e.g. 1996 0942A) or collector's voucher number from wild collections (e.g. R.T. Pennington 639). All herbanium vouchers at RGBE.

soybean (Soya 1), pea (Pisum CYC1, Pisum CYC2; D. Luo, personal communication), and *Medicago truncatula* (Medicago 1, BG455508). Snapdragon CYC and DICH and Arabidopsis TCP1 were chosen as outgroups in the partial TCP and R domains nucleotide sequence analysis.

DNA Sequence Alignment

Unambiguous alignment of all 54 legume *CYC*-like DNA sequences from 25 taxa was only possible in the TCP and R domains and reduced the matrix to 145 nucleotide characters. However, by excluding certain problematic sequences, it was possible to align certain parts of the variable region between these two conserved domains as protein sequences that were then analyzed as nucleotide sequences. Protein sequences were aligned using ClustalX (Thompson et al., 1997), followed by manual adjustments taking both amino acids and nucleotides into consideration.

Phylogenetic Analysis

Protein Methods

Protein distance analysis was carried out using program from the PHYLIP package (Felsenstein, 1993). One hundred half-deletion jackknife data sets were obtained with SEQBOOT, distance matrices were calculated with PROTDIST using the PAM-Dayhoff model of amino acid substitution, neighbor-joining trees were obtained with NEIGHBOR, and a consensus tree was produced by CONSENSE. Branches with <50% support were collapsed. Protein ML analysis was also carried out using PHYLIP. The most parsimonious trees were calculated with PROTPARS (Felsenstein, 1993), with support values obtained by 100 half-deletion jackknife replicates as described above. A 50% MR consensus tree was obtained with CONSENSE, collapsing branches with <50% jackknife support. Protein ML analysis was carried out using TREEPUZZLE v5 (Schmidt et al., 2002) with the BLOSUM 62 model of substitution (Henikoff and Henikoff, 1992) allowing for two rates of heterogeneity (1 invariable + 1 variable). To provide support values, Bayesian analysis was carried out using MrBayes v2.01 (Huelsenbeck and Ronquist, 2001), using the PAM-Dayhoff amino acid substitution model with one million generations sampled every 100 generations with a burn-in of 100,000 generations.

DNA Methods

Maximum parsimony analysis was carried out using PAUP* 4.0b7 (Phylogenetic Analysis Using Parsimony, version 4.0b7, Sinaur Associates, Sunderland, MA). Heuristic searches with 1,000 random addition replicates (to avoid local optima) and tree bisection reconnection (TBR) branch swapping were conducted with steepest descent and multrees options selected. A maximum of 10 minimal length trees was retained per replicate, and a further heuristic search by TBR was carried out on the shortest trees. Branch support values were calculated by 1,000 bootstrap replicates with simple sequence addition and a maximum of 10 minimal length trees retained per replicate. This search method was carried out both for the TCP and R nucleotide matrices, as well as the matrix incorporating certain variable regions. Bayesian phylogenetic analysis of the TCP plus R data set was carried out using MrBayes v2.01 (Huelsenbeck and Ronquist, 2001), using a general time reversible (GTR) model and site-specific rates partitioned by codon. Chains were run for 600,000 generations (burn-in of 100,000 generations) sampled every 100 generations. Resultant trees were used to generate a 50% MR consensus tree in PAUP* v4.0b7.

ML analyses were carried out for the matrix incorporating the more variable regions. The best-fit model was GTR + I + G (GTR model estimating the proportion of invariable sites and γ-distribution; Rodriguez et al., 1990), selected as the best-fit model of nucleotide substitution by the Akaike Information Criterion using Modeltest v3.06 (Posada and Crandall, 1998). A heuristic ML analysis with TBR branch swapping was carried out using PAUP* v4.0b7 with the parameters defined from above.

Distribution of Materials

Upon request, all novel materials described in this publication will be made available in a timely manner for noncommercial research purposes, subject to the requisite permission from any third-party owners of all or parts of the material. Obtaining permissions will be the responsibility of the requestor.

ACKNOWLEDGMENTS

We thank the Royal Botanic Garden Edinburgh for use of laboratory and glasshouse facilities, the horticultural staff, and the laboratory staff (particularly Michelle Hollingsworth and Alex Ponge) for assistance. Julie Hofer (John Innes Centre) and Susan Barker (University of Western Australia) kindly made available DNA samples. We thank Debbie White (RGBE) for the photographs.

Received October 16, 2002; returned for revision November 20, 2002; accepted December 29, 2002.

LITERATURE CITED

Almeida J, Rocheta M, Galego L (1997) Genetic control of flower shape in *Antirrhinum*. Development **124:** 1387–1392

Baum DA (1998) The evolution of plant development. Curr Opin Plant Biol **1:** 79–86

Bruneau A, Forest F, Herendeen PS, Klitgaard BB, Lewis GP (2001) Phylogenetic relationships in the Caesalpinioideae (Leguminosae) as inferred from chloroplast *trn*L intron sequences. Syst Bot **26:** 487–514

Citerne HL, Moller M, Cronk QCB (2000) Diversity of cycloidea-like genes in Gesneriaceae in relation to floral symmetry. Ann Bot **86:** 167–176

Cronk QCB (2001) Plant evolution and development in a post-genomic context. Nat Rev Genet **2:** 607–619

Cubas P (2002) Role of TCP genes in the evolution of morphological characters in angiosperms. *In* QCB Cronk, RM Bateman, JA Hawkins, eds, Developmental Genetics and Plant Evolution. Taylor and Francis, London, pp 247–266

Cubas P, Coen E, Zapater JMM (2001) Ancient asymmetries in the evolution of flowers. Curr Biol **11:** 1050–1052

Cubas P, Lauter N, Doebley J, Coen E (1999a) The TCP domain: a motif found in proteins regulating plant growth and development. Plant J **18:** 2115–2222

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

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

Donoghue MJ, Ree RH, Baum DA (1998) Phylogeny and the evolution of flower symmetry in the Asteridae. Trends Plant Sci **3:** 311–317

Doyle JJ (1994) Evolution of a plant homeotic multigene family: towards connecting molecular systematics and molecular developmental genetics. Syst Biol **43:** 307–328

Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small amounts of fresh leaf tissue. Phytochem Bull **19:** 11–15

Doyle JJ, Doyle JL, Ballenger JA, Dickson EE, Kajita T, Ohashi H (1997) A phylogeny of the chloroplast gene *rbc*L in the Leguminosae: the taxonomic correlations and insights into the evolution of nodulation. Amer J Bot **84:** 541–554

Eisen JA (1998) Phylogenomics: improving functional predictions for uncharacterized genes by evolutionary analysis. Genome Res **8:** 163–167

Felsenstein J (1993) PHYLIP (Phylogeny Inference Package) version 3.5c. Department of Genetics, University of Seattle, Washington

Henikoff S, Henikoff JG (1992) Amino acid substitution matrices from protein blocks. Proc Natl Acad Sci USA **89:** 10915–10919

Hu J-M, Lavin M, Wojciechowski M, Sanderson MJ (2000) Phylogenetic systematics of the tribe Millettieae (Leguminosae) based on chloroplast *trn*K/*mat*K sequences, and its implications for evolutionary patterns in Papilionoideae. Am J Bot **87:** 418–430

Huelsenbeck JP, Ronquist F (2001) MRBAYES: Bayesian inference of phylogeny. Bioinformatics **17:** 754–755

Kajita T, Ohashi H, Tateishi Y, Bailey D, Doyle JJ (2001) *rbc*L legume phylogeny, with particular reference to Phaseoleae, Millitteae, and allies. Syst Bot **26:** 515–536

Kramer EM, Irish VF (1999) Evolution of genetic mechanisms controlling petal development. Nature **399:** 144–148

Lavin M, Pennington RT, Klitgaard BB, Sprent JI, de Lima HC, Gasson P (2001) The dalbergioid legumes (Fabaceae): delimitation of a pantropical monophyletic clade. Am J Bot **88:** 503–533

Lopez P, Casane D, Philippe H (2002) Heterotachy, an important process of protein evolution. Mol Biol Evol **19:** 1–7

Luo D, Carpenter R, Copsey L, Vincent C, Clark J, Coen E (1999) Control of organ asymmetry in flowers of *Antirrhinum*. Cell **99:** 367–376

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

McSteen P, Hake S (1998) Genetic control of plant development. Curr Opin Biotechnol **9:** 189–195

Pennington RT, Klitgaard BB, Ireland H, Lavin M (2000) New insights into floral evolution of basal Papilionoideae from molecular phylogenies. *In* PS Herendeen, A Bruneau, eds, Advances in Legume Systematics, Vol 9. Royal Botanic Gardens, Kew, UK, pp 233–248

Pennington RT, Lavin M, Ireland H, Klitgaard B, Preston J, Hu J-M (2001) Phylogenetic relationships of basal papilionoid legumes based upon sequences of the chloroplast *trn*L intron. Syst Bot **26:** 537–556

Polhill RM (1981) Papilionoideae. *In* RM Polhill, PH Raven eds, Advances in Legume Systematics, Vol 1. Royal Botanic Gardens, Kew, UK, pp 191–208

Posada D, Crandall KA (1998) Modeltest: testing the model of DNA substitution. Bioinformatics **14:** 817–818

Rodriguez F, Oliver JF, Marin A, Medina JR (1990) The general stochastic model of nucleotide substitutions. J Theor Biol **142:** 485–501

Schmidt HA, Strimmer K, Vingron M, von Haeseler A (2002) TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. Bioinformatics **18:** 502–504

Shepard KA, Purugganan MD (2002) The genetics of plant morphological evolution. Curr Opin Plant Biol **5:** 49–55

Siebert PD, Chenchick A, Kellogg DE, Lukyanov KA, Lukyanov SA (1995) An improved PCR method for walking in uncloned genomic DNA. Nucleic Acids Res **23:** 1087–1088

Stebbins GL (1974) Flowering Plants: Evolution above the Species Level. Harvard University Press

Thompson JF, Gibson F, Plewmiak F, Jenamougin F, Higgins DG (1997) The ClustalX window interface: flexible strategies for multiple sequence alignment aided by quality analysis tool. Nucleic Acids Res **25:** 4876–4882