

SHORT COMMUNICATION

The TCP domain: a motif found in proteins regulating plant growth and development

Pilar Cubas^{1,†}, Nick Lauter², John Doebley² and Enrico Coen^{1,*}

¹John Innes Centre, Colney Lane, Norwich NR4 7UH, UK, and

²Department of Plant Biology, University of Minnesota, St. Paul, Minnesota 55108, USA

Summary

The *cycloidea* (*cyc*) and *teosinte branched 1* (*tb1*) genes code for structurally related proteins implicated in the evolution of key morphological traits. However, the biochemical function of CYC and TB1 proteins remains to be demonstrated. To address this problem, we have analysed the predicted secondary structure of regions conserved between CYC and TB1, and looked for related proteins of known function. One of the conserved regions is predicted to form a non-canonical basic-Helix-Loop-Helix (bHLH) structure. This domain is also found in two rice DNA-binding proteins, PCF1 and PCF2, where it has been shown to be involved in DNA-binding and dimerization. This indicates that the conserved domain most probably defines a new family of transcription factors, which we have termed the TCP family after its first characterised members (TB1, CYC and PCFs). Other plant proteins of unknown function also belong to this family. We have studied two of these in *Arabidopsis* and have shown that they are expressed in rapidly growing floral primordia. This, together with the proposed involvement of *cyc* and *tb1* in influencing meristem growth, suggests that many members of the TCP family may affect cell division. Some of these genes may have been recruited during plant evolution to generate new morphological traits.

Introduction

The *cycloidea* (*cyc*) and *teosinte branched 1* (*tb1*) genes have been implicated in the evolution of key morphological traits. The *cyc* gene is involved in the control of floral symmetry, a character that has changed many times during plant evolution (Carpenter and Coen, 1990; Luo *et al.*, 1996;

Stebbins, 1974). The *tb1* gene controls developmental switches that contributed to the evolution of maize from its wild ancestor teosinte (Doebley *et al.*, 1995; Doebley *et al.*, 1997). Although both *cyc* and *tb1* have been isolated, the biochemical function of their encoded proteins is unclear (Doebley *et al.*, 1997; Luo *et al.*, 1996). To address this problem, we have analysed the predicted secondary structure of these and related proteins, and compared some of the gene expression patterns.

The *cyc* gene is required, together with a related gene, *dichotoma* (*dich*), to establish dorsoventral asymmetry of the *Antirrhinum* flower. In flowers mutant for both *cyc* and *dich*, differences between dorsal, lateral and ventral organs are eliminated, rendering the flower radially symmetrical. The *cyc* gene is expressed in the dorsal region of wild-type floral meristems, from very early through to later stages of development. The initial activity reduces the growth rate in the dorsal region of the wild-type meristem and controls primordium initiation. Late expression prevents the full development of the dorsal stamen and affects the asymmetry, size and cell types of the dorsal and lateral petals (Luo *et al.*, 1996).

The *tb1* gene affects the fate of maize axillary meristems; at lower nodes it prevents the outgrowth of buds and at upper nodes it promotes the development of female inflorescences (ears). In *tb1* mutants of maize, axillary buds of lower nodes grow out to give basal branches (tillers), and the buds of upper nodes give branches tipped with male inflorescences (tassels), a phenotype reminiscent of the ancestor of maize, teosinte (Doebley *et al.*, 1997).

Although the processes controlled by *cyc* and *tb1* appear to be unrelated, there are some common themes. First, both genes are involved in the development of axillary structures, either flowers or branches. Second, both genes affect petals and stamens, organs whose development is regulated by the B class of floral organ identity genes (Coen and Meyerowitz, 1991). Third, both genes have been proposed to function, at least in part, as modifiers of organ growth.

To understand the biological role of *cyc* and *tb1*, we have investigated the possible functions of their proteins. In particular, we have studied regions conserved between *cyc* and *tb1* that may act as functional domains. The predicted secondary structure of one of the regions is a basic-Helix-Loop-Helix (bHLH). This region is unrelated to the bHLH structure found in canonical bHLH transcription factors (Murre *et al.*, 1989), but is closely related to a bHLH domain

Received 14 December 1998; revised 24 February 1999; accepted 3 March 1999.

*For correspondence (fax +44 1603 456844; e-mail coen@bbsrc.ac.uk).

[†]Present address: Centro Nacional de Biotecnología, Universidad Autónoma de Madrid, Cantoblanco, 28047 Madrid, Spain.

found in two rice DNA binding proteins, PCF1 and PCF2 (Kosugi and Ohashi, 1997). Based on this homology we define a new class of proteins, the TCP family, that most likely act as transcription factors. We have also characterised two further members of the TCP family from *Arabidopsis* and have shown that they are expressed in floral organs undergoing rapid growth. Taken together with the phenotypic effects of *cyc* and *tb1*, this suggests that members of the TCP family may influence cell division and growth. Recruitment of some of these genes to play new roles during plant development may underlie some key morphological changes during angiosperm evolution.

Results

CYC and *TB1* contain a basic-Helix-Loop-Helix domain

To investigate the biochemical action of the *CYC* and *TB1* proteins, we analysed their predicted secondary structures. In particular, we studied the regions showing sequence conservation, as these might constitute functional domains.

The first conserved region is predicted to form a basic-Helix-Loop-Helix (bHLH, Figures 1a and 2a). This bHLH domain is defined by structural criteria and is unrelated in sequence to the bHLH domain found in MyoD, E12 and related proteins (Murre *et al.*, 1989). The basic region of the bHLH domain of both *CYC* and *TB1* is 21 residues long and includes a putative bipartite nuclear localisation signal (NLS) (Dingwall and Laskey, 1991; Doebley *et al.*, 1997; Luo *et al.*, 1996). The helical regions are amphipathic and comprise alternating conserved hydrophobic residues and partially conserved hydrophilic residues (Figure 1b). The second helix contains a LXXLL-motif (indicated in Figure 2a), which has been shown to mediate binding of transcriptional co-activators to liganded nuclear receptors in animals (Heery *et al.*, 1997). Helix II also contains three potential sites of phosphorylation (serine or threonine), two of them in conserved positions. The region linking the two helices has conserved glycine, aspartate and serine residues found with high frequency in loops (Lesczynski and Rose, 1986), as well as proline in the case of *CYC*. In addition to the bHLH domain, *CYC* and *TB1* have a second conserved region termed the R-domain (Figure 2b), which is rich in polar residues (arginine, lysine and glutamic acid). The R-domain is predicted to form a hydrophilic α -helix (not shown). Both the bHLH and R domains are also predicted to form coiled coils, similar to those formed by leucine zippers (Lupas *et al.*, 1991; Lupas, 1996).

The TCP domain

When *CYC* and *TB1* were first described, no sequence similarity to proteins of known biochemical function was

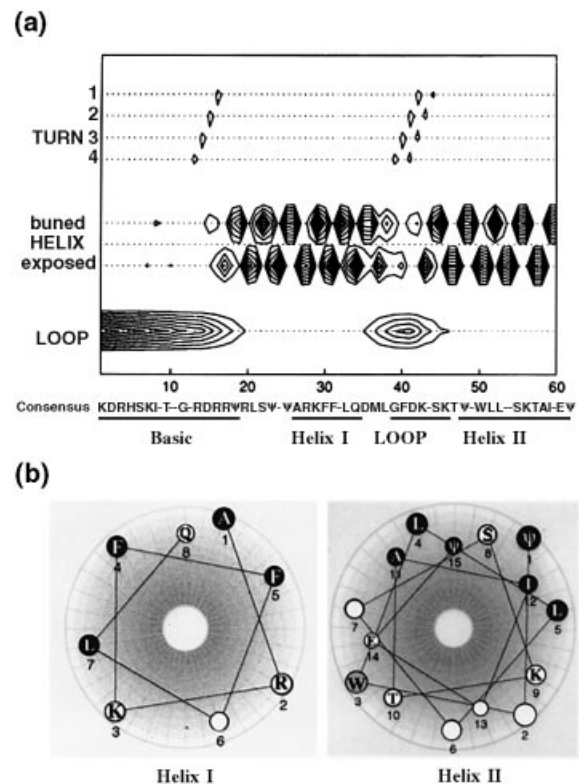


Figure 1. Predicted secondary structure of the *CYC*/*TB1* bHLH domain. (a) Plot representing secondary structure probabilities of the *CYC* bHLH domain based on DSM (Discrete State-Space Models) predictions. The consensus for *CYC* and *TB1* is shown below the plot. The probabilities of each residue being in each of the states (turn, helix or loop) are depicted using contour lines of constant probability, in increments of 0.1. The strand state is improbable for the *CYC* bHLH domain (not shown). (b) Helical wheels representing the residues from proposed helix I and helix II. Divergence angle between adjacent residues being 100°. Conserved hydrophobic residues are represented by white letters inside black circles. Conserved hydrophilic residues are represented by black letters inside white circles. Empty circles correspond to positions with non-conserved residues. Note that the conserved hydrophobic residues are at one side of the helix and conserved hydrophilic residues mainly at the opposite side. The grey circle corresponds to a tryptophan residue on the hydrophilic side. Ψ stands for any hydrophobic residue.

found. However, more recent database searches revealed two additional proteins with similarity to *CYC* and *TB1* in the bHLH region: PCF1 and PCF2. These proteins were isolated on the basis of their ability to bind specifically to promoter elements of the rice gene for the proliferating cell nuclear antigen, PCNA, a protein involved in DNA replication and cell cycle control (for a review see Zophonias and Hubscher, 1997). The PCF proteins may bind to DNA as homo- or heterodimers. A region containing the bHLH domain has been shown to be sufficient for DNA binding and necessary for dimerization (Kosugi and Ohashi, 1997). This suggests that *CYC* and *TB1* may also function as DNA-binding proteins and may interact with other proteins through their bHLH domain.

In addition to *CYC*, *TB1* and the PCFs, several sequences of unknown function from *Arabidopsis* and maize show

detectable in the pollen sacs where microspore mother cells were undergoing meiosis (Figure 4c,g). Transcript levels were similar in dorsal (adaxial), lateral and ventral (abaxial) stamens. By stage 10, when pollen grains were mature, transcripts were no longer detectable in stamens (Figure 4c). (4) In carpels, signal was detected in the placental tissue during stage 9 when ovule primordia were forming (Figure 4g). The expression pattern in roots was not analysed.

In summary, the *Arabidopsis* TCP2 and TCP3 genes are most strongly expressed during flower development; expression being highest in petal and stamen primordia, but also being detectable in sepals and carpels. Expression coincides with the stages when primordia are growing rapidly, consistent with a role for these genes in primordial growth, but by no means conclusive as many other functions may also be compatible with such expression

patterns. Unlike *cyc*, there is no distinction in expression levels between dorsal and ventral primordia.

Discussion

We have shown that CYC and TB1 belong to a family of proteins sharing a common region, the TCP domain, that is predicted to form a basic-Helix-Loop-Helix (bHLH) structure. This region is unrelated in sequence to the canonical bHLH domain found in transcription factors such as MyoD (Murre *et al.*, 1989). However, it is similar to the bHLH domain found in PCF1 and PCF2, plant DNA-binding proteins that most probably act as transcription factors (Kosugi and Ohashi, 1997). The main conserved features of the TCP domain are: two short stretches of residues in the basic region, hydrophobic residues along the apolar face of both α -helices, a tryptophan in helix II, and a helix-breaking glycine in the loop between the helices. So far, members of the TCP family have only been found in plants.

What is the biochemical function of the TCP domain? Important clues have been obtained from the analysis of PCF1 and PCF2. In these proteins, the basic region of the TCP domain is necessary for specific binding to promoter elements of the PCNA gene (Kosugi and Ohashi, 1997). Basic regions have also been shown to be involved in DNA binding in the case of bHLH, bHLHZ and bZIP proteins (Hurst, 1994; Littlewood and Evans, 1994). In these transcription factors, the basic domain adopts an α -helical structure that interacts with the major groove of the DNA. In contrast, the basic region of the TCP domain contains residues that prevent helix formation, suggesting that in this case the basic region binds to DNA through a different mechanism (Kosugi and Ohashi, 1997). It is possible that other TCP proteins bind DNA through their basic domain

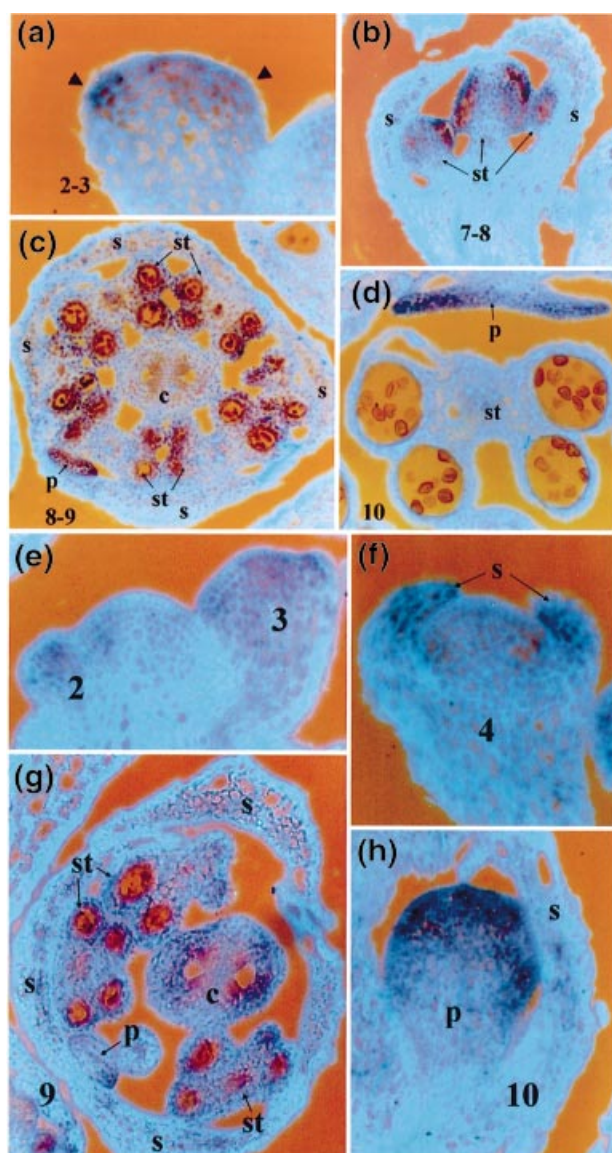


Figure 4. Expression patterns of TCP2 and TCP3 during *Arabidopsis* floral development.

RNA *in situ* hybridisation of wild-type *Arabidopsis* sections probed with TCP2 (a-d) or TCP3 (e-h). (a,b,e,f,h) show longitudinal sections; (c,d,g) are transverse sections. (a) Stage 2–3 floral meristem, when sepal primordia are yet not visible. TCP2 mRNA accumulates at the positions where adaxial (right) and abaxial (left) sepal primordia will form (arrowheads), although some signal is detected throughout. (b) Stage 7–8 floral meristem when petals are small but the stamens are growing rapidly. Signal accumulates in the anther region. (c) Stage 8–9 floral meristem. Signal accumulated in the pollen sacs where pollen grains are forming and in the growing petals. (d) Stage 10 flower. In the anthers, the pollen grains are mature and TCP2 mRNA is no longer detectable. The mRNA accumulates in the growing petals. (e) Young floral meristems of stages 2 and 3 when TCP3 signal is diffuse throughout. (f) Stage 4 the signal accumulated in the growing sepal primordia. (g) Stage 9 flower meristem. Notice that the expression pattern of TCP3 is similar to that of TCP2 (c), although TCP3 is expressed at higher levels in the carpel primordia. In the petal primordia the TCP3 signal is stronger at the margins. (h) Stage 10 petal primordia. TCP3 mRNA mainly accumulates at the margins of the growing petal primordia. The section is parallel to the plane of the petal primordium, i.e. diagonal to the flower bud. Controls using TCP2 or TCP3 sense probes gave no signal (not shown). s, sepal; p, petal; st, stamen; c, carpel. The numbers indicate developmental stages according to Smyth *et al.* (1990).

in a similar way. In addition, the basic region of the TCP domain may target these proteins into the nucleus as it contains a complete (CYC/TB1 subfamily) or partial (PCF subfamily) bipartite NLS (Dingwall and Laskey, 1991; Doebley *et al.*, 1997; Luo *et al.*, 1996). This would fit the observation that NLSs often overlap or flank nucleic acid-binding domains (LaCasse and Lefebvre, 1995; Littlewood and Evans, 1994)

The role of the HLH region of the PCFs and other TCP proteins is less clear. One possibility is that the amphipathic helices mediate protein–protein interactions through their hydrophobic surfaces. This would be similar to the proposed role of the amphipathic helix (K domain) of MADS box genes (Davies and Schwarz-Sommer, 1994; Shore and Sharrocks, 1995). Amphipathic helices also mediate homo- and heterodimerization in bZIP and bHLH proteins of the MyoD type (Landschulz *et al.*, 1988; Murre *et al.*, 1989). In the PCFs, a region containing the TCP domain is essential for homo- and heterodimerization, although the role of the helices has not been tested (Kosugi and Ohashi, 1997). Amphipathic helices might also be involved in interactions with non-TCP proteins. For instance, helix II of the TCP domain contains a conserved sequence that resembles the LXXLL motif shown to be involved in protein interactions with liganded nuclear receptors (Heery *et al.*, 1997). It is possible that TCP proteins interact with as yet unidentified plant nuclear receptors or other proteins through this sequence.

Most members of the TCP family contain up to three potential phosphorylation sites in serine and threonine in the basic domain and helix II. Phosphorylation has been shown to affect nuclear localisation, DNA binding and transcriptional activation of regulatory proteins (Hunter and Karin, 1992), raising the possibility that the activity of the TCP proteins might be regulated by a similar mechanism.

The TCP proteins fall into two subfamilies (one including CYC and TB1 and the other including the PCFs) based on features both within and outside the TCP domain. Within the TCP domain, each subfamily has a different linker for the bipartite NLS, a distinct residue composition in the loop and hydrophilic faces of the helices, and a different length for helix II. Outside the TCP domain, most members of the CYC/TB1 subfamily have an R-domain, predicted to form a coiled coil that may mediate protein–protein interactions (Lupas *et al.*, 1991), and all members of the PCF subfamily share short regions flanking the TCP domain. These differences between subfamilies may reflect differences in the DNA-binding specificities and/or protein–protein interactions.

Unlike the homeo-domain containing genes of meta-zoans, many of which occur in tandem clusters in the genome (Holland *et al.*, 1994), the *Arabidopsis* TCP genes are dispersed throughout the genome. In this respect they

resemble the MADS box family of higher plants, which are also dispersed (Hauge *et al.*, 1993; Rounsley *et al.*, 1995).

All the members of the TCP family investigated thus far function in processes related to cell proliferation. CYC retards growth of the dorsal region of young floral meristems, affecting the number of primordia initiated. Later, it arrests dorsal stamen development and promotes dorsal petal growth (Luo *et al.*, 1996). TB1 is involved in arresting growth of some axillary buds, repressing internode growth in branches, and arresting petal (lodicule) and stamen development in the female flowers (Doebley *et al.*, 1995; Doebley *et al.*, 1997). PCF1 and PCF2 most probably control the transcription of PCNA, a gene expressed only in meristematic tissue where it is involved in DNA replication and cell cycle control. Here we show that the expression patterns of two *Arabidopsis* members of the CYC/TB1 subfamily, TCP2 and TCP3, also correlate with actively dividing regions of the floral meristem, suggesting a possible involvement of these genes in regulating cell growth and division. However, such expression patterns may also be compatible with other biological roles. Further experiments, such as inactivation of the TCP2 and TCP3 genes, will be needed to determine whether these genes are indeed involved in regulating growth.

The TCP2 and TCP3 genes are most probably not orthologues of CYC or TB1, as CYC and TB1 are more closely related to each other than to either of these genes (Figure 2c). However, they do share with CYC and TB1 some features in their expression patterns: the TCP2 and TCP3 genes are upregulated in petal and stamen primordia, similar to *cyc*; the *tb1* gene is also likely to be expressed in these organ primordia as *tb1* mutants affect the development of petals and stamens. One distinctive feature, however, is that *cyc* is expressed only in dorsal primordia whereas TCP2 and TCP3 are expressed uniformly along the dorsoventral axis.

It is possible that many members of the TCP family function in proliferating tissues where they may act in combination with other proteins to influence cell division and growth, and perhaps recruitment of some of these regulatory genes for new developmental functions has been involved in generating key changes in plant morphology during angiosperm evolution.

Experimental procedures

Secondary structure analysis

The secondary structure prediction for CYC was obtained by probabilistic Discrete State-Space Models analysis, DSMs (Stultz *et al.*, 1993; White *et al.*, 1994), submitting the protein sequences to the Protein Sequence Analysis (PSA) server (<http://bmerc-www.bu.edu/psa/coment3.htm>). Helical wheels were plotted with PROTEAN (DNASTAR for Windows 3.10a) and refined by hand. The prediction of coiled-coil regions (Lupas *et al.*, 1991) was

obtained by submitting the protein sequences to the Network Protein Sequence analysis server (<http://pbil.ibcp.fr>).

Multiple sequence comparison

The TCP and R-domain alignments were constructed with MEGALIGN (DNASTAR for Windows 3.10a), using CLUSTALW (Thompson *et al.*, 1994) and PAM250 weighted distances and were manually refined. Phylogenetic analysis was performed with the PHYLIP package (Felsenstein, 1985). Two hundred bootstrap re-samplings of the original data were generated with the SEQBOOT program. Distance matrices were made for each bootstrap dataset using the PRODIST program with the Dayhoff PAM distance method, and 200 trees constructed from these by the Neighbor-joining method.

In situ hybridisations

Plant material was grown under long days and collected for *in situ* hybridisation as described by Ratcliffe *et al.* (1998). Digoxigenin labelling of RNA probes, tissue preparation and *in situ* hybridisation were done as described by Coen *et al.* (1990).

Sequencing

Plasmid clones of the R30409 and T45419 ESTs (corresponding to TCP2 and TCP3, respectively) were obtained from the ABRC at Ohio University. These clones were completely sequenced at the Advanced Genetic Analysis Center (University of Minnesota) using gene specific internal oligonucleotide primers.

Genetic mapping

Four TCP genes were placed on the *Arabidopsis* genetic map using the R30409 (TCP2) and T45419 (TCP3) EST clones as probes on DNA blots of a set of 99 recombinant inbred lines (ABRC Stock Number CS1899). For R30409, there was one strongly hybridising locus that corresponds to TCP2 and maps to chromosome 4 between markers g3845 and m600. For T45419, there was one strongly hybridising locus that corresponds to TCP3 and maps to chromosome 1 between markers m213 and nga128. T45419 also hybridised to two other loci more faintly: one between markers m283 and nga361 on chromosome 2 and the other between markers EW18E10 I and nga162 on chromosome 3. In addition, TCP1 (AC002130) maps in chromosome 1, TCP5 (AB008269), TCP6 (AB010072) and TCP7 (AB007648) in chromosome 5, and TCP8 (H36511) and TCP9 (AC003680) in chromosome 2. DNA isolation and DNA blot analysis were performed as described by Doebley and Stec (1993). Linkage maps were assembled using Mapmaker 2.0 (Lander *et al.*, 1987).

Acknowledgements

We thank Oliver Ratcliffe for providing the *Arabidopsis* material for *in situ* hybridisation. Thanks also to Rosemary Carpenter, Desmond Bradley, Oliver Ratcliffe, Utpal Nath and Sandra Doyle for comments on the manuscript and Coral Vincent and Theresa Warr for help with the final version of the manuscript.

References

- Carpenter, R. and Coen, E.S. (1990) Floral homeotic mutations produced by transposon-mutagenesis in *Antirrhinum majus*. *Genes Dev.* **4**, 1483–1493.
- Coen, E.S. and Meyerowitz, E.M. (1991) The war of the whorls: genetic interactions controlling flower development. *Nature*, **353**, 31–37.
- Coen, E.S., Romero, J.M., Doyle, S., Elliott, R., Murphy, G. and Carpenter, R. (1990) *floricaula*: a homeotic gene required for flower development in *Antirrhinum majus*. *Cell*, **63**, 1311–1322.
- Davies, B. and Schwarz-Sommer, Z. (1994) Control of floral organ identity by homeotic MADS-box transcription factors. In *Results and Problems in Cell Differentiation* (Nover, L., ed.). Berlin: Springer-Verlag, pp. 235–258.
- Dingwall, C. and Laskey, R.A. (1991) Nuclear targeting sequences – a consensus? *TIBS*, **16**, 478–481.
- Doebley, J. and Stec, A. (1993) Inheritance of the morphological differences between maize and teosinte: comparison of results for two F₂ populations. *Genetics*, **134**, 559–570.
- Doebley, J., Stec, A. and Gustus, C. (1995) *Teosinte branched1* and the origin of maize: evidence for epistasis and the evolution of dominance. *Genetics*, **141**, 333–346.
- Doebley, J., Stec, A. and Hubbard, L. (1997) The evolution of apical dominance in maize. *Nature*, **386**, 485–488.
- Felsenstein, J. (1985) Confidence limits on phylogenies: an approach using the bootstrap. *Evolution*, **39**, 783–791.
- Hauge, B., Hanley, S., Cartinhour, S. *et al.* (1993) An integrated genetic/RFLP map of the *Arabidopsis thaliana* genome. *Plant J.* **3**, 745–754.
- Heery, D.M., Kalkhoven, E., Hoare, S. and Parker, M.G. (1997) A signature motif in transcriptional co-activators mediates binding to nuclear receptors. *Nature*, **387**, 733–736.
- Holland, P.W.H., Garcia-Fernandez, J., Williams, N.A. and Sidow, A. (1994). Gene duplications and the origins of vertebrate development. *Development*, **Suppl.**, 125–133.
- Hunter, T. and Karin, M. (1992) The regulation of transcription by phosphorylation. *Cell*, **70**, 375–387.
- Hurst, H.C. (1994) Transcription factors: 1 bZIP proteins. *Protein Profile*, **1**, 123–168.
- Kosugi, S. and Ohashi, Y. (1997) PCF1 and PCF2 specifically bind to *cis* elements in the rice proliferating cell nuclear antigen gene. *Plant Cell*, **9**, 1607–1619.
- LaCasse, E.C. and Lefebvre, A. (1995) Nuclear localisation signals overlap DNA- or RNA binding domains in nucleic acid-binding proteins. *Nucl Acids Res.* **23**, 1647–1656.
- Lander, E.S., Green, P., Abrahamson, J., Barlow, A., Lincoln, S. and Newburg, A. (1987). MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. *Genomics*, **1**, 174–181.
- Landschulz, W.H., Johnson, P.F. and McKnight, S.L. (1988) The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins. *Science*, **240**, 1759–1762.
- Lesczynski, J.W. and Rose, G.D. (1986) Loops in globular proteins: a novel category of secondary structure. *Science*, **234**, 849–855.
- Littlewood, J. and Evans, G. (1994) Transcription factors: 2 helix-loop-helix. *Protein Profile*, **1**, 639–669.
- Luo, D., Carpenter, R., Vincent, C., Copsey, L. and Coen, E. (1996) Origin of floral asymmetry in *Antirrhinum*. *Nature*, **383**, 794–799.
- Lupas, A. (1996) Coiled coils: new structures and new functions. *Trends Biochem. Sci.* **10**, 375–382.
- Lupas, A., Van Dyke, M. and Stock, J. (1991) Predicting coil coils from protein sequences. *Science*, **252**, 1162–1164.

- Murre, C., McCaw, P.S. and Baltimore, D.** (1989) A new DNA binding and dimerization motif in immunoglobulin enhancer binding, Daughterless, MyoD and myc proteins. *Cell*, **56**, 777–783.
- Ratcliffe, O.J., Maya, I., Vincent, C.A., Rothstein, S., Carpenter, R., Coen, E.S. and Bradley, D.J.** (1998) A common mechanism controls the life cycle and architecture of plants. *Development*, **125**, 1609–1615.
- Rounsley, S.D., Ditta, G.S. and Yanofsky, M.F.** (1995) Diverse roles for MADS-box genes in *Arabidopsis* development. *Plant Cell*, **7**, 1259–1269.
- Shore, P. and Sharrocks, A.D.** (1995) The MADS-box family of transcription factors. *Eur J. Biochem.* **229**, 1–13.
- Smyth, D.R., Bowman, E.M. and Meyerowitz, E.M.** (1990) Early flower development in *Arabidopsis*. *Plant Cell*, **2**, 755–767.
- Stebbins, G.L.** (1974) *Flowering Plants, Evolution Above the Species Level*. Massachusetts: Harvard University Press.
- Stultz, C.M., White, J.V. and Smith, T.F.** (1993) Structural analysis based on state-space modelling. *Protein Sci.* **2**, 305–314.
- Thompson, J.D., Higgins, D.G. and Gibson, T.J.** (1994) ClustalW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting position specific gap penalties and weight matrix choice. *Nucl Acids Res.* **22**, 4673–4680.
- White, J.V., Stultz, C.M. and Smith, T.F.** (1994) Protein classification by stochastic modelling and optimal filtering of amino-acid sequences. *Math. Biosci.* **119**, 35–75.
- Zophonias, O.J. and Hubscher, U.** (1997) Proliferating cell nuclear antigen: more than a clamp for DNA polymerases. *Bioessays*, **19**, 967–975.

GenBank accession numbers AF072691 and AF072134.