

# Heterogeneous Selection on *LEGCYC* Paralogs in Relation to Flower Morphology and the Phylogeny of *Lupinus* (Leguminosae)

Richard H. Ree,\*<sup>1</sup> H  l  ne L. Citerne,† Matt Lavin,‡ and Quentin C. B. Cronk\*

\*University of British Columbia, Botanical Garden and Centre for Plant Research, Vancouver, British Columbia, Canada;

†Royal Botanic Garden Edinburgh, Edinburgh, UK; ‡Montana State University

An analysis of the molecular evolution of two *LEGCYC* paralogs in *Lupinus* (Genisteeae: Leguminosae) reveals a varied history of site-specific and lineage-specific evolutionary rates and selection both within and between loci. *LEGCYC* genes are homologous to regulatory loci known to control floral symmetry and adaxial flower organ identity in *Antirrhinum* and its relatives. Within *Lupinus*, *L. densiflorus* is unusual in having flowers with a proportionally smaller standard (fused adaxial petals) and larger wings (lateral petals) than other lupin species. Phylogenetic estimates of the nonsynonymous/synonymous substitution rate ratio,  $\omega$ , suggest that along the *L. densiflorus* lineage, positive selection ( $\hat{\omega} > 1$ ) acted at some codon sites of one paralog, *LEGCYC1B*, and greater purifying selection ( $\hat{\omega} < 1$ ) acted at some sites of the other paralog, *LEGCYC1A*. Overall, *LEGCYC1A* appears to be evolving faster than *LEGCYC1B*, and both paralogs are evolving faster than the internal transcribed spacer (ITS) region of nr DNA. The predominant historical pattern inferred is a highly heterogeneous "selectional mosaic" which we suggest may be typical of the teosinte branched 1-cycloidea-PCF (TCP) class of transcriptional activators, and possibly other genes. Codon models that do not account for both site-specific and lineage-specific variation in  $\omega$  do not detect positive selection at these loci. We suggest a modification of existing branch-site models involving an additional  $\omega$  parameter along the foreground branch, to account for the effects of both greater positive selection and greater purifying selection at different codon sites along a particular branch. The higher rates of evolution and congruent phylogenetic signal of both *LEGCYC* paralogs show promise for the use of these genes as markers for phylogeny reconstruction at low taxonomic levels in Genisteeae.

## Introduction

From the rapidly advancing field of evolutionary developmental genetics, it is emerging that morphological phenotypes can be dramatically affected by genetic mutations at regulatory loci (e.g., Cronk 2001; Doebley and Lukens 1998). Studies of regulatory gene evolution may thus provide insights into the origins of morphological diversity. Thus far, however, evidence that evolutionary rates and directional selection at regulatory loci are correlated with rates of morphological evolution is mixed. In the morphologically diverse Hawaiian silverswords, for example, nonsynonymous substitution rates for the floral homeotic (MADS-box) genes *APETALA3* and *APETALAI* were found to be higher than for the (nonregulatory) photosynthetic structural gene *ASCAB9* (Barrier, Robichaux, and Purugganan 2001), but directional selection in the *DELLA* subfamily of putative transcription factors could not be detected in the silverswords compared to their North American relatives (Remington and Purugganan 2002). Neither could selection be detected in duplicated floral organ identity genes (*CYCLOIDEA/DICHOTOMA*) in *Antirrhinum* and its relatives (Hileman and Baum 2003) or in *tb1*-like genes in maize (Lukens and Doebley 2001). The paucity of direct evidence for adaptive evolution in regulatory genes could reflect a biological generality that phenotypic change is the outcome of changes in the location and timing of regulatory gene expression, rather than change at the amino acid sequence level of the

regulatory genes themselves (Doebley and Lukens 1998). Nevertheless, the prominence of transcription factors in our emerging understanding of morphological development is a compelling reason to investigate their molecular evolution in relation to comparative morphology.

Inferences of adaptive molecular evolution from nucleotide sequence data typically focus on the rate of nonsynonymous substitution ( $d_N$ ) relative to that of synonymous substitution ( $d_S$ ), with  $\omega = d_N/d_S$  (Goldman and Yang 1994; Muse and Gaut 1994). The most common method of estimating  $\omega$  from interspecific data is to optimize stochastic models of codon substitution on phylogenetic trees representing the genealogical history of the sequences under scrutiny. Codon models of molecular evolution in a phylogenetic context are becoming increasingly fine-grained in identifying substitution patterns that represent potentially adaptive evolution, to a point where it is now possible to test statistically for individual sites with elevated  $d_N/d_S$  ratios along specific lineages of a phylogenetic tree (Yang and Nielsen 2002). Such analyses of genes that regulate morphological development are thus a potentially powerful means of connecting directional or diversifying selection at the molecular level to adaptive phenotypic change at the morphological level.

The focus of this study is the angiosperm group *Lupinus* (Leguminosae: Papilionoideae: Genisteeae). Papilionoid legumes have flowers that are characteristically zygomorphic (i.e., differentiated along but symmetric about a dorsoventral axis), in which the dorsal, lateral, and ventral petals make up the standard, wings, and keel, respectively. Typically the standard is reflexed (i.e., held upright) and is large in proportion to the rest of the flower, presumably functioning to attract pollinators and effect outcrossing.

The morphology and anatomy of flowers in Leguminosae have been well-studied (e.g. see Ferguson and Tucker 1994); in contrast, the genetic control of legume flower development is only beginning to be studied and

<sup>1</sup> Present address: Botany Department, Field Museum of Natural History, Chicago, Illinois.

Key words: nonsynonymous/synonymous rate ratio, flower morphology, *LEGCYC*, *CYCLOIDEA*, Papilionoideae, *Lupinus*.

E-mail: quentin.cronk@ubc.ca.

*Mol. Biol. Evol.* 21(2):321–331. 2004

DOI: 10.1093/molbev/msh022

Advance Access publication December 5, 2003

*Molecular Biology and Evolution* vol. 21 no. 2

   Society for Molecular Biology and Evolution 2004; all rights reserved.

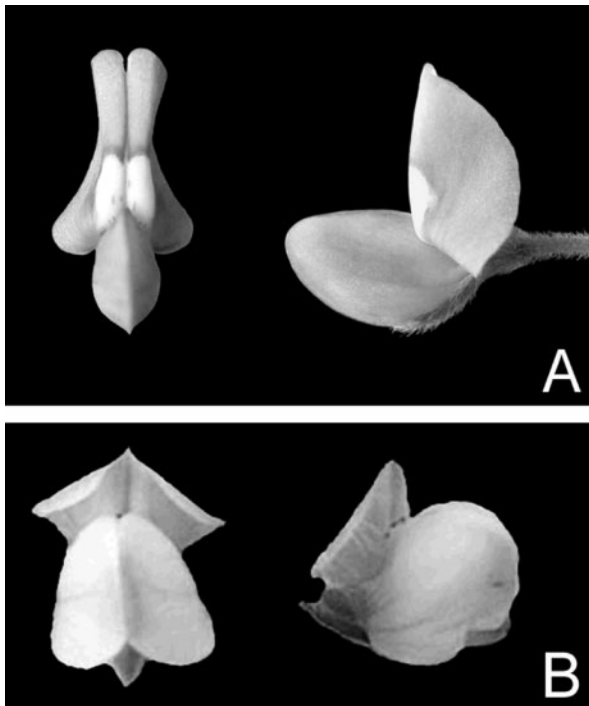


FIG. 1.—Flowers of (A) *Lupinus texensis* and (B) *L. densiflorus*, examples of standard-dominated and wing-dominated flowers, respectively.

understood. Molecular phylogenetic studies of Leguminosae (Citerne et al. 2003) discovered genes with homology to *CYCLOIDEA* (*CYC*) genes in *Antirrhinum majus*. *CYC* genes are transcription factors in the TCP gene family that are known to be involved in flower development, in particular conferring adaxial floral organ identity and hence zygomorphy (Luo et al. 1996). In legumes, two major groups (I and II) of *LEGCYC* genes have been identified, and within Genisteae two copies of group I genes, *LEGCYC1B* and *LEGCYC1A*, are known (Citerne et al. 2003). These latter genes are paralogs resulting from a gene duplication event that predates the origin of Genisteae (Citerne et al. 2003). Both are expressed in the dorsal region of the developing perianth (H. Citerne, unpublished data), which is consistent with expression patterns of *CYC* in *Antirrhinum* (Luo et al. 1999).

In this article we describe the molecular evolution of *LEGCYC* group I paralogs in *Lupinus* in relation to flower morphology, with emphasis on the western New World species. Despite considerable interspecific variation in foliage and seed coat characters (Ainouche and Bayer 1999; Heyn and Herrstadt 1977), all species of *Lupinus* have typical papilionoid flowers. Much of the floral variation in the group occurs in the size of the dorsal petal (standard) relative to the lateral petals (wings). In this respect most lupin species are fairly uniform, having large standards relative to the wings (i.e., the flowers are standard-dominated; fig. 1A). Flowers of *L. densiflorus* are an exception: they are wing-dominated and smaller than those of other lupin species, characteristics that are possibly associated with self-pollination. *Lupinus densiflorus* is thus a particularly interesting species because its

wing-dominated floral morphology is unusual within lupins and may represent a morphological adaptation corresponding to a shift in breeding system. In this article we use likelihood models of codon evolution and phylogenetic inference to investigate the hypothesis that in the *L. densiflorus* lineage, the change to wing-dominated flowers is associated with higher rates of nonsynonymous evolution in genes known to regulate floral development. As a point of comparison, we also test the lineage leading to the North American lupin species flock, along which we have no reason to believe *LEGCYC* paralogs should have experienced directional selection, on account of the plesiomorphic morphological uniformity of flowers in that group. By focusing on closely related species, the time scale of this study provides a “snapshot” of gene evolution compared to previous broader surveys within Leguminosae (Citerne et al. 2003). Inferences of *LEGCYC* evolution at the scale of all legumes were limited because *LEGCYC* homologs were too divergent to reliably align outside of the conserved TCP and R domains. By focusing more narrowly, we hope to obtain a more fine-grained view of the molecular evolution of genes that are likely to be important in flower development.

Our secondary focus here is on the utility of *LEGCYC* genes for phylogeny reconstruction in *Lupinus*, reasoning that rapidly evolving nuclear loci may provide phylogenetic signal beyond that of molecular markers more commonly used for phylogeny estimation. Analyses of the internal transcribed spacer (ITS) region of nuclear rDNA in *Lupinus* reveal five main clades in *Lupinus*, but support for relationships among and within these clades is poor overall (Ainouche and Bayer 1999). One of these, the clade of western New World species, appears to represent a recent rapid radiation; within it, relationships of the North American species cannot be resolved by ITS alone (Ainouche and Bayer 1999). We analyze the *LEGCYC* paralogs singly and in combination with ITS to determine whether an improvement in clade resolution and support is obtained.

## Methods

### Study Species

Twenty-three individuals were sampled from fifteen species of *Lupinus* and one outgroup species in Genisteae, *Genista tenera* (table 1). All but four of the lupin species are distributed primarily in western North America. Three species (*L. albus*, *L. cosentinii*, and *L. digitatus*) are distributed in the Old World, and *L. texensis* is from the southeastern United States and represents an eastern New World clade (Ainouche and Bayer 1999).

### DNA Isolation and Sequencing

Genomic DNA was isolated from fresh leaf tissue using DNEasy Plant System kits (Qiagen, Chatsworth, CA) from greenhouse-germinated seedlings of *Lupinus albus*, *L. cosentinii*, *L. densiflorus*, *L. digitatus*, *L. nanus*, and *L. texensis*, from botanical garden specimens of *L. polyphyllus* and *Genista tenera*, and from field-collected individuals of other North American species of *Lupinus* (table 1).

**Table 1**  
**Taxa Sampled**

Taxon sample	Voucher	GenBank Accession Numbers		
		<i>LEGCYC1A</i>	<i>LEGCYC1B</i>	ITS
<i>Lupinus albus</i>	USDA <sup>1</sup> PI 170528 [UBC]	AY338920	AY338871	AY338946
<i>L. andersonii</i>	Lavin 0815985 (Mt. Rose, NV) [MONT]	AY338908	AY338859	AY338934
<i>L. arboreus</i>	UBC Botanical Garden	AY338918	AY338869	AY338944
<i>L. argenteus</i> 1	Delgado & Lavin s.n. (Fairy Lake, MT) [MEXU]	AY338903	AY338854	AY338929
<i>L. argenteus</i> 2	Lavin 0815982 (Mt. Rose, NV) [MONT]	AY338905	AY338856	AY338931
<i>L. argenteus</i> 3	Lavin s.n. (Bozeman, MT) [MONT]	AY338910	AY338861	AY338936
<i>L. breweri</i> ssp. <i>breweri</i>	Lavin 7031 (Graeagle, CA) [MONT]	AY338912	AY338863	AY338938
<i>L. breweri</i> ssp. <i>bryoides</i>	Lavin 0815983 (Mt. Rose, NV) [MONT]	AY338906	AY338857	AY338932
<i>L. cosentinii</i>	USDA W6 17967 [UBC]	AY338921	AY338872	AY338947
<i>L. densiflorus</i> 1	UBC Botanical Garden (Holland Point, B.C.)	AY338914	AY338865	AY338940
<i>L. densiflorus</i> 2	UBC Botanical Garden (Macauley Point, B.C.)	AY338915	AY338866	AY338941
<i>L. densiflorus</i> 3	UBC Botanical Garden (Finlayson Point, B.C.)	AY338916	AY338867	AY338942
<i>L. densiflorus</i> 4	UBC Botanical Garden (Trial Island, B.C.)	AY338917	AY338868	AY338943
<i>L. digitatus</i>	USDA W6 11995 [UBC]	AY338922	AY338873	AY338948
<i>L. grayi</i>	Lavin 7033 (Graeagle, CA) [MONT]	AY338913	AY338864	AY338939
<i>L. lepidus</i> ssp. <i>utahensis</i>	Lavin 0815984 (Mt. Rose, NV) [MONT]	AY338907	AY338858	AY338933
<i>L. nanus</i>	Commercial seed (Sutton Seeds, Paignton, Devon, UK)	AY382155	AY382156	AY338928
<i>L. polyphyllus</i> 1	Lavin 0815981 (Mt. Rose, NV) [MONT]	AY338904	AY338855	AY338930
<i>L. polyphyllus</i> 2	UBC Botanical Garden	AY338919	AY338870	AY338945
<i>L. rivularis</i>	Lavin 7040 (Smith River, CA) [MONT]	AY338911	AY338862	AY338937
<i>L. sericeus</i>	Lavin s.n. (Bear Trap, MT) [MONT]	AY338909	AY338860	AY338935
<i>L. texensis</i>	USDA PI 577291 [UBC]	AY338923	AY338874	AY338949
<i>Genista tenera</i>	UBC Botanical Garden 22652-050-82	AY338924	AY338875	AY338950

<sup>1</sup> USDA source: US Department of Agriculture, Agricultural Research Service, Pullman, Washington.

*LEGCYC1B* and *LEGCYC1A* were amplified by polymerase chain reaction (PCR) using 34 cycles with denaturation at 94°C for 60 s, annealing at 50°C to 55°C for 60 s, and extension at 72°C for 90 s. We amplified each locus separately in two to three fragments each using a combination of primers designed for *LEGCYC* group 1, and primers designed specifically for *LEGCYC1A* and *LEGCYC1B*. Primer sequences are available from the authors on request. Amplified fragments were sequenced directly and assembled into contigs using Sequencher (Gene Codes, Ann Arbor, Mich.). In some cases locus-specific primers failed to amplify their target; in others, apparent allelic length differences in two regions of trinucleotide repeats in *LEGCYC1B* resulted in unreadable chromatograms downstream of the point of variation. For these samples, PCR was performed using the general primers *LEGCYC1\_F1* and *LEGCYC1\_R1* (Citerne, unpublished) and the products were cloned. Clones containing *LEGCYC1B* and *LEGCYC1A* were selected for sequencing based on diagnostic digests using *EcoR1*. The ITS region, including the 5.8S subunit, was amplified with forward primer ITS5 and reverse primer ITS4 (White et al. 1990) using 34 cycles with denaturation at 94°C for 60 s, annealing at 55°C to 58°C for 60 s, and extension at 72°C for 90 s. In some cases multiple bands were consistently amplified, so the appropriate fragment (ca. 800 bp in length) was isolated either by agarose gel extraction or by cloning.

The PCR products were purified using QiaQuick kits (Qiagen) and sequenced directly or cloned into the pCR 2.1-TOPO vector using TOPO TA cloning kits (Invitrogen, Carlsbad, Calif.). One to four clones were sequenced from each ligation reaction using vector primers M13F and

M13R. Automated sequencing was carried out on ABI PRISM 377 DNA sequencers.

#### Phylogenetic Analyses

One representative sequence of each gene (*LEGCYC1B*, *LEGCYC1A*, and ITS) was selected from each sample. Sequences were aligned “by eye.” Phylogenetic analyses were conducted using PAUP\* version 4b10 (Swofford 1998) and MrBayes (Huelsenbeck and Ronquist 2001). Gaps were treated as missing data. The most appropriate nucleotide substitution model for phylogenetic analysis of each gene was chosen using the Akaike Information Criterion (AIC) (Akaike 1974). Hierarchical model comparisons (Posada and Crandall 2001) were facilitated by the use of MrModeltest (Nylander 2002). Estimates of substitution parameters from the model selection process were used as starting values for subsequent phylogenetic analyses.

For individual genes, maximum-likelihood (ML) trees were found using PAUP\* by heuristic searches employing ten random-addition replicates with tree bisection-reconnection (TBR) branch swapping. Bayesian posterior probabilities of clades were found by Metropolis-coupled Markov chain Monte Carlo (MCMC) using MrBayes. For individual genes, we ran MCMC analyses with four chains (three “heated”) for 1,300,000 generations. We discarded the first 300,000 generations as burn-in after visually confirming probable chain stationarity from plots of likelihood against generation. We also combined the *LEGCYC1B*, *LEGCYC1A*, and ITS alignments into a gene-partitioned “total evidence” data set for a Bayesian analysis in which each partition was assigned its optimal substitution model with parameters that were

estimated independently for each gene. This analysis ran for 3,000,000 generations (including 1,000,000 for burn-in) with five chains (four heated). For all MCMC runs, chains were sampled every 100 generations.

### Phylogenetic Congruence Tests

We used the Shimodaira-Hasegawa (SH) test (Shimodaira and Hasegawa 1999) implementation in PAUP\* to assess the degree to which the different data sets are in agreement with respect to tree topology. We tested each data set against a set of 1,413 trees comprising the ML topologies from each data set and all topologies visited by the Markov chain in the total evidence analysis. This reasonably fulfills the requirement of the SH test that the set of trees  $\mathcal{M}$  contains all topologies that can reasonably be considered to be the true tree (Goldman, Anderson, and Rodrigo 2000).

### Analyses of *LEGCYC* Coding Region Evolution

Models of codon evolution and tests for selection on *LEGCYC* paralogs were evaluated on phylogenies obtained in the previous analyses using the program codeml from the PAML package, version 3.13 (Yang 1997). To represent our best possible estimate of *Lupinus* phylogeny, we selected the Bayesian total-evidence tree with maximum posterior probability that could not be rejected statistically by any individual data set. In addition, we evaluated all models on the optimal trees from single-gene analyses of each *LEGCYC* gene to assess the sensitivity of our results to tree topology.

We evaluated a range of codon models for measuring selection and compared nested models by the standard likelihood ratio test (LRT) statistic,  $2\Delta\ell$ , against the  $\chi^2$  distribution with degrees of freedom equal to the difference in number of parameters. In all cases, we ran the analyses multiple times with different starting values of  $\kappa$  and  $\omega$  to ensure that parameter estimates were stable. The codon models are described in more detail in the original articles, but we provide a brief overview of them here, following the model names used in Yang and Nielsen (2002). M0 is the simplest codon model with a single  $\omega$  parameter for all sites and branches of the phylogeny. M1, M2, and M3 are the site models in which  $\omega$  varies among sites but is constant across the phylogeny. M1 is the “neutral” model with two site classes that restricts  $\omega_0 = 0$  (implying strict purifying selection) and  $\omega_1 = 1$  (neutral evolution). M2, the “selection” model, is an extension of M1 and has an additional site class for which  $\omega_2$  can take any value. M3, the “discrete” model, is an extension of M0 and has  $k$  site classes, each with a separate  $\omega$  parameter. In this study we used  $k = 2$  because additional site classes did not improve the likelihood, and having two site classes facilitates comparison with branch-site model B (see below). M7 and M8 use a discrete beta distribution to approximate among-site  $\omega$  variation, differing only in the latter, allowing an additional site class with  $\omega > 1$  (Yang et al. 2000).

The branch models allow  $\omega$  to vary among branches on the phylogeny but not among sites. We used two variants: the “free-ratio” model, in which  $\hat{\omega}$  is estimated

separately for each branch on the phylogeny, and a “two-ratio” model, in which the phylogeny is partitioned into “background” and “foreground” branches and  $\hat{\omega}$  is estimated independently for each branch partition (Yang 1998). We targeted the *Lupinus densiflorus* branch and its sister lineage, the branch subtending the North American species, as foreground branches.

The branch-site models of Yang and Nielsen (2002) allow sites to have different nonsynonymous rates along foreground branches specified *a priori* and as such are similar to the two-ratio model described above. These models incorporate three  $\omega$  parameters: two “background” ratios,  $\omega_0$  and  $\omega_1$ , that are constrained to be less than or equal to one, and one ratio specific to the foreground lineage of interest,  $\omega_2$ , that may exceed one. Each codon site has either  $\omega_0$  or  $\omega_1$  across the entire tree (i.e., is not under selection), or just along background lineages, with  $\omega_2$  along foreground lineages (i.e., positively selected). Thus there are four site classes in these models. Model A of Yang and Nielsen (2002) restricts  $\omega_0 = 0$  and  $\omega_1 = 1$ , and is therefore a branch-specific extension of model M1. Model B places no restrictions on  $\omega$  parameters and is thus a branch-specific extension of model M3.

To assess variation along the length of the *LEGCYC* paralogs, estimates of substitution rate and  $\hat{\omega}$  were obtained along a 60-bp sliding window of the alignments. HYPHY (Muse and Pond 2002) was used to estimate substitution rate and codeml was used to estimate  $\hat{\omega}$ .

## Results

### Sequence Data

Within individuals, clones of the same locus showed minimal variation (< 0.01% raw pairwise divergence); this was interpreted as PCR error. Allelic length variation was observed in *LEGCYCIB* at two trinucleotide repeat regions. On the sequence of *Lupinus nanus*, the first region (GGT, glycine) is located at base position 274–288, and the second is located at position 1083–1098 (AAC, asparagine). Repeat number varied from two to seven repeats in the first region and from two to eight repeats in the second.

### Phylogenetic Analysis

The alignment of *LEGCYCIB* is 1,450 nucleotides in length, with coding sequence from sites 1–1320, followed by a short (approx. 95-bp) intron, and a short region of further coding sequence that was not used in tests of selection (see below). The alignment of *LEGCYCIA* is 1,292 nucleotides in length, with coding sequence from sites 1–1133. The aligned length of ITS, including the 5.8S region, was 681 nucleotides. For phylogenetic inference, AIC model comparisons indicated the use of the GTR + I +  $\Gamma$  model for *LEGCYCIB* and *LEGCYCIA*, and the GTR +  $\Gamma$  model for ITS.

Phylogenetic analyses of each gene (*LEGCYCIB*, *LEGCYCIA*, and ITS) each yielded single ML trees that are shown in figure 2. The relationships inferred by each gene are generally consistent with one another and with previous analyses of ITS by Ainouche and Bayer (1999).

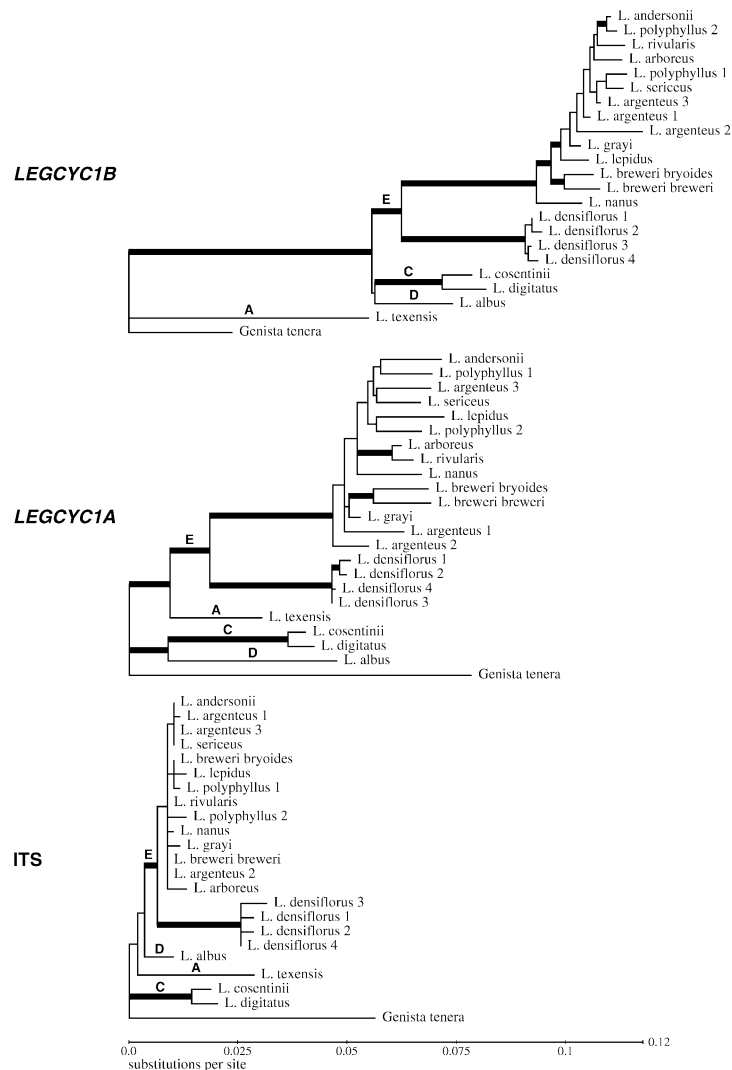


FIG. 2.—Maximum-likelihood phylogenies of *LEGCYC1B*, *LEGCYC1A*, and ITS. Clades labeled A, C, D, and E correspond to those similarly labeled in Ainouche and Bayer (1999). Branch lengths on each tree are drawn to a common scale (bottom). Bold branches are supported by posterior probabilities  $\geq 0.95$ .

In that study, five major groups within *Lupinus* were identified: two comprising eastern and western New World species, respectively (clades A and E) and three comprising Old World species (clades B, C, and D), but ITS alone did not provide sufficient signal to resolve relationships among these groups. In this study the eastern New World lupins (clade A) are represented by *L. texensis*, but representatives of clade B were not sampled.

*LEGCYC1B* and *LEGCYC1A* provide better resolved phylogenetic hypotheses than ITS. In fact, ITS alone strongly supports only three clades, whereas *LEGCYC1B* and *LEGCYC1A* each support nine. Within the western New World species (clade E of Ainouche and Bayer 1999), ITS only supports the monophyly of *L. densiflorus*, whereas each *LEGCYC* gene strongly supports a clade of strictly North American species exclusive of *L. densiflorus*. This clade is also supported in the total-evidence analysis (see below).

In only a single case, the placement of *L. texensis*, do different genes (*LEGCYC1B* and *LEGCYC1A*) each

support conflicting relationships with posterior probability  $\geq 0.95$ . If we consider the root of the phylogeny to be along the branch subtending the outgroup, *Genista tenera*, *LEGCYC1B* places *L. texensis* as the sister group to the rest of *Lupinus*, while *LEGCYC1A* places the clade of Old World species (*L. cosentinii* + *L. digitatus* + *L. albus*) as the sister group to the New World clades. It is possible that more comprehensive taxonomic sampling would resolve this conflict, which may be due to long-branch attraction of *L. texensis* to *Genista tenera*.

Combining the three genes for total-evidence phylogenetic analysis provides a more resolved and robust estimate of *Lupinus* phylogeny than analyses of single genes. The Markov chain sampled 1,410 distinct tree topologies from the posterior distribution. The tree with highest posterior probability (0.062) is shown in figure 3. All strongly supported clades inferred from single genes are also strongly supported by the total-evidence Bayesian analysis, with the exception of the position of *L. texensis*, yielding a total of 11 strongly supported nodes. This tree

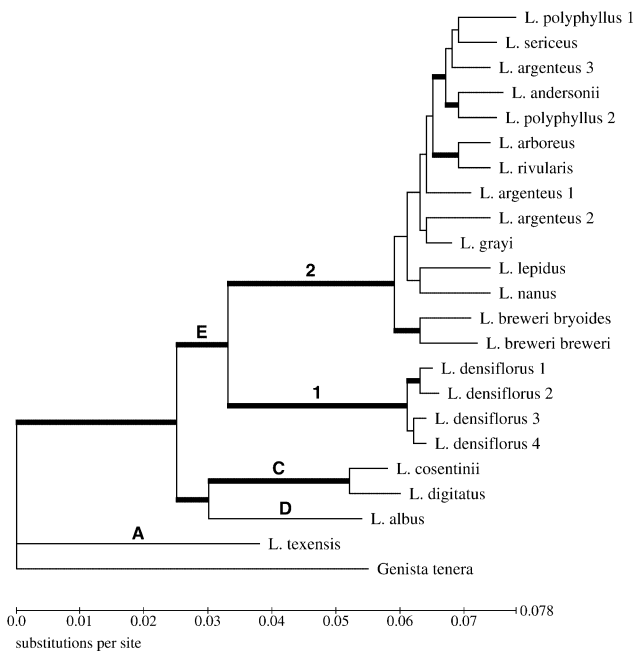


FIG. 3.—Total-evidence phylogeny inferred by Bayesian Markov chain Monte Carlo with substitution model parameters estimated separately for each gene. The topology shown has the highest posterior probability (0.062) of the 1,410 topologies visited along the chain. Branch lengths are the means of their Bayesian posterior distributions. Bold branches are supported by posterior probabilities  $\geq 0.95$ . Clade labels follow figure 2; clades numbered 1 and 2 correspond to the *Lupinus densiflorus* and North American foreground branches used in tests for positive selection on *LEGCYC1B* and *LEGCYC1A*.

also cannot be rejected statistically (SH test) by any single-gene data set, and we therefore interpret it to be the best overall estimate of relationships among the species sampled, and use it in subsequent analyses of codon evolution in *LEGCYC1B* and *LEGCYC1A*. Perhaps the most notable new relationship revealed by the *LEGCYC* and three-gene analyses is within the western New World lineage, the North American species form a clade distinct from *L. densiflorus*, whose North American distribution may represent a disjunction from its origins in a primarily South American lineage.

In both single-gene and total-evidence analyses, sequences from the same species do not always coalesce, e.g., for *L. argenteus* and *L. polyphyllus*. This is likely due to persistent ancestral polymorphisms at these loci or introgression between species. Ancestral polymorphism seems a more plausible explanation, as there does not appear to be any common pattern among the different genes with regard to species parphyly.

#### Patterns of Molecular Evolution

Mean treelengths (substitutions/site  $\pm$  SD) from the posterior distributions of trees visited by MCMC for each gene are as follows: *LEGCYC1B*,  $0.415 \pm 0.026$ ; *LEGCYC1A*,  $0.508 \pm 0.030$ ; ITS,  $0.269 \pm 0.029$ . Given that identical sets of taxa were sampled for each gene, and assuming independent gene trees, treelengths from the single-gene analyses reflect overall evolutionary rates for

each gene. By this measure both *LEGCYC* genes have higher rates than ITS, which is consistent with their apparently greater phylogenetic signal. *LEGCYC1A* also has a higher substitution rate overall than *LEGCYC1B*, despite the fact that, being more ultrametric, its gene tree appears shorter than that of *LEGCYC1B* in terms of root-to-tip distance (fig. 2). The *LEGCYC1A* tree has longer branches, particularly toward the tips, and it has a greater length overall than *LEGCYC1B*.

For both *LEGCYC1B* and *LEGCYC1A*, substitution rates and  $\hat{\omega}$  values vary substantially among codon sites and across the phylogeny. Likelihood scores and parameter estimates for the various codon models for each gene are given in table 2. We estimated ancestral sequences with model M0 and mapped the synonymous and nonsynonymous changes on the phylogeny (fig. 4). On many branches, *LEGCYC1B* and *LEGCYC1A* have undergone different amounts of synonymous and nonsynonymous evolution.

#### Site Models

For both *LEGCYC* paralogs, the neutral model (M1) fits the data poorly: M0, with only a single site class but unrestricted  $\omega$ , gives better likelihood scores than M1 with two constrained site classes. Model M2, which extends M1 by adding a third site class with unrestricted parameter  $\omega_2$ , provides a significant improvement in likelihood (*LEGCYC1B*:  $2\Delta\ell = 24.892$ ,  $df = 2$ ,  $P < 0.001$ ; *LEGCYC1A*:  $2\Delta\ell = 28.679$ ,  $df = 2$ ,  $P < 0.001$ ), and a large proportion of sites are inferred to be in the third class. This suggests that, across the whole tree, most sites are not evolving under absolute purifying selection or neutrally, but somewhere in between; the restrictions imposed by model M1 thus appear to be too stringent. Likelihood ratio tests of M0 versus M3 with  $k = 2$  are also significant in favor of the latter model (*LEGCYC1B*:  $2\Delta\ell = 20.391$ ,  $df = 2$ ,  $P < 0.001$ ; *LEGCYC1A*:  $2\Delta\ell = 27.811$ ,  $df = 2$ ,  $P < 0.001$ ). This indicates that sites are not homogeneous with respect to  $\omega$ : i.e., allowing at least two site classes with independent  $\omega$  parameters provides a better fit to the data.

In general, analyses of *LEGCYC* alignments using the site models with selection (M0, M3, and M8) provide no evidence for positive selection having acted on sites across the whole phylogeny. For *LEGCYC1A* with M8,  $\hat{\omega} = 1.510$ , but the proportion of sites in this class is exceedingly small ( $< 0.00001$ ) and the support for this model over M7 is nonsignificant. In contrast, the sliding-window analysis reveals two regions of each *LEGCYC* paralog for which  $\hat{\omega} > 1$  (fig. 5). Likelihood ratio tests show that for *LEGCYC1B*, these peaks are not significantly greater than one, but for *LEGCYC1A*, the windows at positions 702 ( $\hat{\omega} = 6.33$ ), 960 ( $\hat{\omega} = 4.25$ ), and 966 ( $\hat{\omega} = 7.29$ ) do support the hypothesis of positive selection ( $P = 0.049$ ,  $0.038$ , and  $0.005$ , respectively). Thus positive selection may have acted at some sites in these windows, but model M8 is unable to detect them. The difference in results may be due to the effect of conditioning on codon frequencies estimated from the whole alignment versus frequencies local to the sliding window.

**Table 2**  
**Codon Model Parameter Estimates for *LEGCYC* Paralogs**

Model	$p$	<i>LEGCYCIA</i>			<i>LEGCYCIB</i>	
		$\ell$	Parameter Estimates	Positively Selected Sites	$\ell$	Parameter Estimates
M0 (one ratio)	1	-3186.12	$\hat{\omega} = 0.259$	None	-2758.49	$\hat{\omega} = 0.202$
Site models						
M1: neutral	1	-3188.60	$\hat{p}_0 = 0.558$ ( $\hat{p}_1 = 0.442$ )	NA	-2758.80	$\hat{p}_0 = 0.592$ ( $\hat{p}_1 = 0.408$ )
M2: selection	3	-3176.16	$\hat{p}_0 = 0.260$ , $\hat{p}_1 = 0.105$ ( $\hat{p}_2 = 0.634$ ); $\hat{\omega}_2 = 0.257$	None	-2744.46	$\hat{p}_0 = 0.471$ , $\hat{p}_1 = 0.029$ ( $\hat{p}_2 = 0.49975$ ); $\hat{\omega}_2 = 0.381$
M3: discrete (k = 2)	3	-3175.92	$\hat{p}_0 = 0.753$ ( $\hat{p}_1 = 0.247$ ); $\hat{\omega}_0 = 0.115$ , $\hat{\omega}_1 = 0.717$	None	-2744.58	$\hat{p}_0 = 0.494$ ( $\hat{p}_1 = 0.506$ ); $\hat{\omega}_0 = 0.001$ , $\hat{\omega}_1 = 0.430$
M7: beta	2	-3176.12	$\hat{p} = 0.533$ , $\hat{q} = 1.475$	NA	-2744.91	$\hat{p} = 0.397$ , $\hat{q} = 1.375$
M8: beta& $\omega$	4	-3176.12	$\hat{p}_0 = 0.759$ , $\hat{p} = 13.354$ , $\hat{q} = 99.000$ ( $\hat{p}_1 = 0.241$ ); $\hat{\omega} = 0.723$	None	-2744.91	$\hat{p}_0 = 1.000$ , $\hat{p} = 0.397$ , $\hat{q} = 1.375$ ( $\hat{p}_1 = 0.000$ ); $\hat{\omega} = 1.51$
Branch models						
Free-ratio		-3160.73		NA	-2738.30	
Foreground 1: <i>Lupinus densiflorus</i> branch						
Two ratios	2	-3185.57	$\hat{\omega}_0 = 0.250$ , $\hat{\omega}_1 = 0.401$	NA	-2758.41	$\hat{\omega}_0 = 0.205$ , $\hat{\omega}_1 = 0.167$
Branch-site model A	3	-3185.90	$\hat{p}_0 = 0.293$ , $\hat{p}_1 = 0.205$ ( $\hat{p}_2 + \hat{p}_3 = 0.502$ ); $\hat{\omega}_2 = 14.956$	34A, 194S, 240K, 299K, 379H, 391A, 430F (0.5 < P < 0.8), 249Q (P = 0.94)	-2757.61	$\hat{p}_0 = 0.293$ , $\hat{p}_1 = 0.205$ ( $\hat{p}_2 + \hat{p}_3 = 0.502$ ); $\hat{\omega}_2 = 0.003$
Branch-site model B	5	-3173.31	$\hat{p}_0 = 0.698$ , $\hat{p}_1 = 0.276$ ( $\hat{p}_2 + \hat{p}_3 = 0.026$ ); $\hat{\omega}_0 = 0.097$ , $\hat{\omega}_1 = 0.641$ , $\hat{\omega}_2 = 12.126$	249Q (P = 0.96), 379H (P = 0.83)	-2744.50	$\hat{p}_0 = 0.407$ , $\hat{p}_1 = 0.419$ ( $\hat{p}_2 + \hat{p}_3 = 0.174$ ); $\hat{\omega}_0 = 0.001$ , $\hat{\omega}_1 = 0.436$ , $\hat{\omega}_2 < 0.001$
Foreground 2: North American branch						
Two ratios	2	-3179.89	$\hat{\omega}_0 = 0.299$ , $\hat{\omega}_1 = 0.069$	NA	-2757.52	$\hat{\omega}_0 = 0.213$ , $\hat{\omega}_1 = 0.113$
Branch-site model A	3	-3177.93	$\hat{p}_0 = 0.096$ , $\hat{p}_1 = 0.084$ ( $\hat{p}_2 + \hat{p}_3 = 0.820$ ); $\hat{\omega}_2 = 0.002$	None	-2754.34	$\hat{p}_0 = 0.074$ , $\hat{p}_1 = 0.050$ ( $\hat{p}_2 + \hat{p}_3 = 0.876$ ); $\hat{\omega}_2 = 0.089$
Branch-site model B	5	-3169.91	$\hat{p}_0 = 0.188$ , $\hat{p}_1 = 0.067$ ( $\hat{p}_2 + \hat{p}_3 = 0.745$ ); $\hat{\omega}_0 = 0.126$ , $\hat{\omega}_1 = 0.796$ , $\hat{\omega}_2 < 0.001$	None	-2743.57	$\hat{p}_0 = 0.188$ , $\hat{p}_1 = 0.187$ ( $\hat{p}_2 + \hat{p}_3 = 0.625$ ); $\hat{\omega}_0 = 0.000$ , $\hat{\omega}_1 = 0.459$ , $\hat{\omega}_2 = 0.060$

NOTE.—Following Yang and Nielsen (2002),  $p$  is the number of free parameters for the  $\omega$  ratios. Parameters in bold indicate positive selection; those in parentheses are not free parameters, e.g., under M1  $p_1 = 1 - p_0$ . Possible sites under positive selection for *LEGCYCIB* are given using *L. nanus* as the reference; no positively selected sites were identified for *LEGCYCIA*.

### Free-Ratio Model

Under the free-ratio model,  $\hat{\omega} < 1$  for nearly all branches for both genes (in each case,  $\hat{\omega}$  slightly exceeds one for only a single branch that is not the same for both genes). The free-ratio model and M0 are not significantly different (*LEGCYCIB*:  $2\Delta\ell = 50.777$ ,  $df = 44$ ,  $P = 0.224$ ; *LEGCYCIA*:  $2\Delta\ell = 40.387$ ,  $df = 44$ ,  $P = 0.627$ ), indicating that the data do not warrant fitting individual  $\omega$  parameters to branches if  $\omega$  is assumed constant across sites. The free-ratio model is parameter-rich relative to model M0, however, so to more rigorously examine branch-specific differences in  $\omega$  it is necessary to reduce the number of free parameters and target specific branches for hypothesis testing.

### Tests for Selection Along the *L. densiflorus* and North American Branches

#### Two-Ratio Model

For both *LEGCYC* paralogs the two-ratio model does not suggest positive selection along either the *L. densiflorus* branch or the North American branch. Likelihood

ratio tests of M0 versus the two-ratio model are only significant for *LEGCYCIB* and the North American branch (table 3), and in this case the foreground  $\hat{\omega}_1$  is close to zero and is less than the background  $\hat{\omega}_0$ , suggesting that, if anything, purifying selection was greater along the North American branch than elsewhere in the tree.

#### Branch-Site Models

The branch-site models provide evidence of positive selection having acted on *LEGCYCIB* along the *L. densiflorus* branch. Under both models (A and B) the foreground branch parameter  $\hat{\omega}_2$  is large, and eight codon sites are identified with a posterior probability greater than 0.5 of being positively selected; for one site, the posterior probability under model B is greater than 0.95. This suggests that *LEGCYCIB* has experienced positive selection along the branch where a putative shift from standard-dominated to wing-dominated flowers occurred. Likelihood ratio tests indicate that the branch-site models are better fitted to the data than the site models, but the improvement is not significant at the level of 0.05 (table 3). In contrast, for *LEGCYCIA* there is no evidence for positive selection

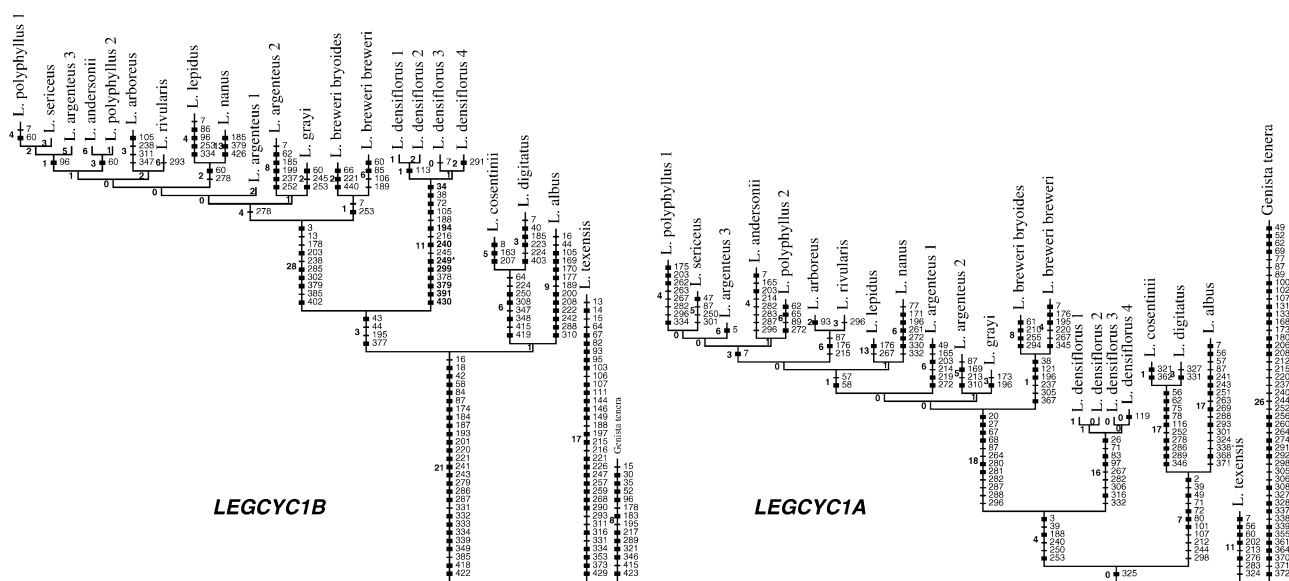


FIG. 4.—Substitution history of *LEGCYC1B* and *LEGCYC1A* based on maximum-likelihood ancestral sequence estimates. Hashmarks indicate nonsynonymous substitutions and are labeled by codon site position. Bold hashmarks indicate unique changes. Bold hashmark labels along the *L. densiflorus* branch (*LEGCYC1B*) indicate sites that have a posterior probability of positive selection  $\geq 0.5$ ; the single site with posterior probability  $>0.95$  is marked with an asterisk. Inferred numbers of synonymous codon changes are shown to the left of each branch.

along the *L. densiflorus* branch and likelihood ratio tests for the branch-site models are nonsignificant.

The positions of the positively selected sites in *LEGCYC1B* are scattered across the coding region, as are the seven other sites inferred to have undergone replacement substitution along the *L. densiflorus* branch (fig. 6). None fall within the TCP or R domains. Only one putative positively selected site is situated 5-prime of the TCP domain. In terms of secondary structure, most of the sites that change fall outside of predicted alpha helix regions (fig. 6).

Along the North American foreground branch,  $\hat{\omega}_2$  is close to zero for both *LEGCYC* genes and no sites are inferred to be positively selected. The branch-site models fit the *LEGCYC1B* data significantly better than the site models, whereas for *LEGCYC1A*, model A is significantly better than M1 but model B is not significantly better than M3 (table 3). One interpretation of these results is that purifying selection was especially strong for some sites of both genes along the North American branch.

#### Sensitivity of Results to Tree Topology

The results from the branch-site tests for selection appear to be somewhat sensitive to tree topology. Using the ML tree topology from the *LEGCYC1A* single-gene analysis,  $\hat{\omega}_2 < 1$  and models M1 and M3 cannot be rejected in favor of models A and B, for both the *L. densiflorus* and North American branches ( $P \gg 0.05$  in all cases). If the *LEGCYC1A* tree is a more accurate estimate of that gene's history than the total-evidence tree, our inferences of positive selection along the *L. densiflorus* branch may be an artifact of apparent homoplasy in nonsynonymous substitutions caused by enforcing an incorrect topology on the data. However, with the SH test, the *LEGCYC1A* topology is rejected by both the

*LEGCYC1B* data set and the total-evidence data set. Thus the preponderance of evidence would recommend the total-evidence tree as the preferred topology.

## Discussion

### Estimating Phylogenies of Recently Diverged Species

Our results provide an example of the use of rapidly evolving nuclear markers that improve estimates of phylogenetic relationships among recently diverged species. Both *LEGCYC* genes have greater phylogenetic signal than ITS, and the total-evidence analysis yields more strongly supported nodes than any single-gene analysis. However, because of limited taxonomic sampling, we caution against weighting too heavily the relationships found in this study. Even in the total-evidence analysis, we are unable to resolve many relationships, particularly among the North American lupins. This lack of signal is to be expected in cases where biological processes such as introgression or retained polymorphism prohibit such inference. For the North American lupins, our data suggest that population-level analyses with multiple markers may be more appropriate for phylogenetic inference.

### Inferring Selection on *LEGCYC* from Codon Models

Codon-based evolutionary models suggest that selection pressure on *LEGCYC* genes in *Lupinus* has been varied among sites and among lineages. Estimates of raw nucleotide and synonymous and nonsynonymous rates of evolution describe a heterogeneous pattern of episodic strong purifying selection, relaxed purifying selection, and positive selection, with little congruence in the molecular signatures of these processes across genes. We term this pattern a "selectional mosaic" and suggest that it may be

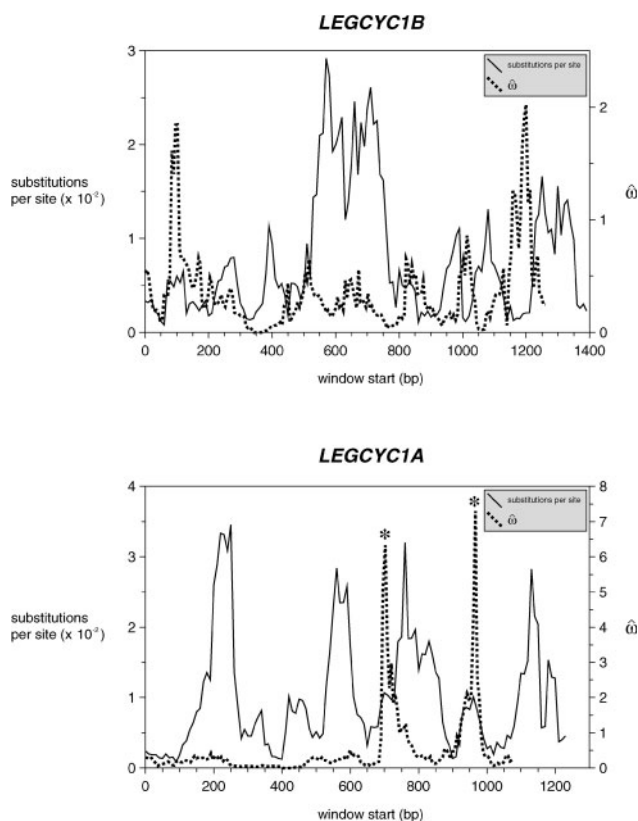


FIG. 5.—Sliding-window graphs of nucleotide substitution rate (60-bp window, 10 bp increments) and  $\hat{\omega}$  (60-bp window, 6 bp increments). Peaks where  $\hat{\omega}$  is significantly greater than one are marked with an asterisk.

typical of teosinte branched 1-cycloidea-PCF (TCP) genes, most of which show high rates of evolution.

Our results indicate that sites in a coding region may experience either greater positive selection or greater purifying selection along a particular branch of a phylogeny. The branch-site models of Yang and Nielsen (2002) appear to be more sensitive in detecting variation in  $\omega$  than models that accommodate among-site or among-branch variation in  $\omega$ , but not both. Positive selection is not supported at either *LEGCYC* locus using any site or branch model, but this may be due to a preponderance of neutral or purifying selection over most of the phylogeny, a pattern that is consistent with the sliding window analysis. This accords with what realistically might occur when natural selection acts on a protein, i.e., only a small proportion of amino acids (e.g. at active binding sites) change as its function evolves. In this light the branch-site models thus represent a substantial advance in inferential power for studies of molecular evolution, but a better understanding of the accuracy and power of LRTs (e.g., see Anisimova Bielawski, and Yang 2000) based on the branch-site models are needed. It is possible that previous studies that failed to detect positive selection with the site or branch models (e.g., Remington and Purugganan 2002; Hileman and Baum 2003) might upon reanalysis find evidence for positive selection with the branch-site models.

The idea of a selectional mosaic suggests one aspect of the current branch-site models that might be refined, namely the single foreground parameter  $\omega_2$  and its use as

**Table 3**  
Likelihood Ratio Tests for Sites Having Different  $\omega$  Ratios Along Foreground Branches

Comparison	Foreground Branch					
	<i>Lupinus densiflorus</i>			North American spp.		
	$2\Delta\ell$	$P$	$\hat{\omega}_{\text{fg}}$	$2\Delta\ell$	$P$	$\hat{\omega}_{\text{fg}}$
<i>LEGCYC1B</i>						
M0 vs. two-ratio	1.099	0.577	0.401	12.455	0.002	0.069
M1 vs. model A	5.409	0.067	14.956	21.339	<0.001	0.002
M3 vs. model B	5.222	0.073	12.126	12.031	0.002	0.000
<i>LEGCYC1A</i>						
M0 vs. two-ratio	0.168	0.919	0.167	1.932	0.381	0.113
M1 vs. model A	2.380	0.304	0.003	8.906	0.012	0.089
M3 vs. model B	0.165	0.921	0.000	2.029	0.363	0.060

an indicator of selection pressure along the foreground lineage. Consider model B and *LEGCYC1B*: the large value of  $\hat{\omega}_2$  along the *L. densiflorus* branch suggests positive selection, but statistical support for the model falls just on the side of nonsignificance, whereas for the North American branch, support for the model is highly significant but  $\hat{\omega}_2$  is close to zero. The difference in support between branches is likely related to the proportion of sites in each case that have  $\omega_2$  along the foreground branch ( $\hat{p}_2 + \hat{p}_3$ ): for the *L. densiflorus* branch  $\hat{p}_2 + \hat{p}_3$  is small (0.026), while for the North American branch  $\hat{p}_2 + \hat{p}_3$  is large (0.745). In other words, only 2.6% of sites are estimated to have higher nonsynonymous rates along the *L. densiflorus* branch, while 74.5% of sites are estimated to have lower nonsynonymous rates along the North American branch. These results do not preclude the possibility that positive selection did occur at some sites along the North American branch; after all, six synapomorphic amino acid substitutions are inferred along it without homoplasy. However, any signal of positive selection would be swamped out by the large proportion of sites that have lower nonsynonymous rates along the North American branch (or alternatively have higher nonsynonymous rates elsewhere in the tree). Likewise, detecting the small number of sites under positive selection along the *L. densiflorus* branch is contingent on the apparent absence of sites that have higher nonsynonymous rates elsewhere in the tree.

Given the objective of identifying adaptive coding region evolution along a particular branch of a gene phylogeny, one might consider alternative formulations of the branch-site model. For example, consider a model with two  $\omega$  parameters associated with the foreground branch rather than just one. This would account for sites with lower nonsynonymous rates along the foreground branch in addition to sites with higher nonsynonymous rates, as might be expected if adaptation of the protein required amino acid fixation at some sites and replacement substitutions at others. One drawback of this proposal would be an increase in model complexity, with six site classes required instead of four, and the resulting loss of statistical power (flattening of the likelihood surface). To compensate for this, other alternatives could be explored, e.g., a model with two foreground  $\omega$  parameters and only a single background  $\omega$  parameter, which is the converse of how those parameters are distributed in model B; this

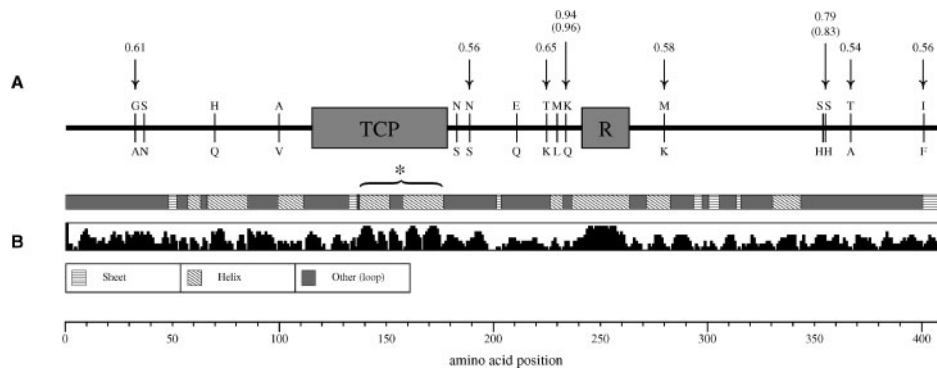


FIG. 6.—*A*, Map of *LEGCYC1B* coding region showing inferred nonsynonymous mutations along the *L. densiflorus* branch. Ancestral and derived amino acids are shown below and above the line, respectively. Numbers indicate posterior probabilities of positive selection under model A (probabilities under model B are shown in parentheses). No mutations are inferred within the TCP or R domains. *B*, predicted secondary structure (Ouali and King 2000) showing sheet, helix, and loop regions with prediction reliability scores shown below. The helix-loop-helix motif characteristic of the TCP domain is marked with an asterisk.

would yield three site classes. The main liability of having two  $\omega$  ratios along the foreground branch, however, would be the interpretation of results: it might be difficult to tell whether support for the model comes from those sites under stronger positive selection along the foreground branch, or those sites under stronger purifying selection, or both. To discern this, it would be necessary to compare likelihood scores and  $\omega$  parameter estimates with those of the branch-site models of Yang and Nielsen (2002).

#### Rates of *LEGCYC* Evolution

What are the main factors responsible for the overall faster evolutionary rates of *LEGCYC* genes compared to *ITS*? One hypothesis is that rapid, divergent evolution of *LEGCYC* somehow promotes speciation; this has been shown in gamete recognition proteins of animals (e.g., Hellberg et al. 2000; Metz and Palumbi 1996). Evolution of transcriptional regulators is commonly believed to be a chief agent in the generation of novel morphological phenotypes that could drive lineage divergence (Doebley and Lukens 1998). However, it is unlikely that *LEGCYC* evolution has consistently driven speciation in *Lupinus*, which apart from *L. densiflorus* shows relatively little variation in floral morphology; if it had, we would expect to see a stronger signal of positive (diversifying) selection on sites over the whole phylogeny.

An alternative explanation is that the fast rate of evolution is a consequence of the selectional mosaic. The idea is that most of the time, for most sites in the gene, purifying selection is relaxed, allowing neutral substitutions to accumulate rapidly. Strong purifying selection acts periodically and only on a small percentage of sites, and therefore has little effect on constraining gene evolution. This pattern of evolution comes from the fact that only certain sites (the DNA binding sites) are important, while the rest are relatively less constrained.

The faster substitution rate of *LEGCYC1A* over *LEGCYC1B* could be the result of functional redundancy following gene duplication (Ohno 1970), releasing *LEGCYC1A* from functional constraints and leading to relaxed purifying selection. Preliminary evidence suggests that like *CYC* genes in *Antirrhinum*, the role of *LEGCYC* lies in

establishing floral organ identity: both *LEGCYC* genes are expressed adaxially in the developing flower, but *LEGCYC1A* apparently has a smaller area of expression than *LEGCYC1B* (H. Citerne, unpublished data).

#### Role of *LEGCYC* in Papilionoid Flower Evolution

In this study we find a high rate of nonsynonymous evolution at a handful of sites in a floral regulatory gene, *LEGCYC1B*, associated with a phylogenetic branch along which a morphological shift occurred from standard-dominated flowers to the wing-dominated flowers of *L. densiflorus*. Support for the branch-site models falls just short of statistical significance, but given the function of *LEGCYC* as a transcriptional regulator, we expect the number of sites under selection to be small and therefore less likely to be detected. Furthermore, there is a striking contrast in pattern with *LEGCYC1A*, which appears to have undergone greater purifying selection along the *L. densiflorus* branch. For these reasons we are inclined to favor the hypothesis that positive selection acted on *LEGCYC1B* in conjunction with this morphological transition. Of the two *LEGCYC* paralogs, *LEGCYC1B* appears to be the more promising candidate for further studies of the genetics of legume flower development. Determining quantitative expression profiles of each gene in *L. densiflorus* floral organs, and gene silencing/knockout mutant studies might provide additional evidence of functional differentiation of these loci and their role in developmental regulation.

Studies of *LEGCYC* in other legume taxa representing more dramatic morphological shifts away from the “typical” papilionoid flower are also needed, particularly those in which selection on standard morphology is implicated. For example, shifts to bird pollination have occurred in several legume groups such as *Erythrina* (Bruneau 1997), presumably involving intense selection for floral traits characteristic of that syndrome, e.g., an elongated, tubelike perianth and reduced standard size. *Cadia*, an enigmatic legume with unusual radially symmetrical flowers likely derived from a papilionoid ancestor, is also a prime candidate for *LEGCYC* analysis as this may help determine whether the flower has evolved as a result of homeosis or heterochrony (Cronk 2002).

## Acknowledgments

We thank Sally Otto and Ziheng Yang for helpful statistical discussions, Jeannette Whitton for generously providing temporary laboratory space, and Toby Pennington for helpful encouragement throughout. We also wish to acknowledge Dorothy Cheung for laboratory assistance, and Ron Rollo and Elaine Le Marquand for greenhouse support in seedling germination.

## Literature Cited

- Akaike, H. 1974. A new look at the statistical model identification. *IEEE Trans. Automat. Contr.* **19**:716–723.
- Anisimova, M., J. P. Bielawski, and Z. Yang. 2000. Accuracy and power of the likelihood ratio test in detecting adaptive molecular evolution. *Mol. Biol. Evol.* **18**:1585–1592.
- Ainouche, A.-K., and R. J. Bayer. 1999. Phylogenetic relationships in *Lupinus* (Fabaceae: Papilionoideae) based on internal transcribed spacer sequences (ITS) of nuclear ribosomal DNA. *Am. J. Bot.* **86**:590–607.
- Barrier, M., R. H. Robichaux, and M. D. Purugganan. 2001. Accelerated regulatory gene evolution in an adaptive radiation. *Proc. Natl. Acad. Sci. USA* **98**:10208–10213.
- Bruneau, A. 1997. Evolution and homology of bird pollination syndromes in *Erythrina* (Leguminosae). *Am. J. Bot.* **84**:54–71.
- Citerne, H. L., D. Luo, R. T. Pennington, E. Coen, and Q. C. B. Cronk. 2003. A phylogenomic investigation of *CYCLOIDEA*-like TCP genes in the Leguminosae. *Plant Physiol.* **131**:1042–1053.
- Cronk, Q. C. B. 2001. Plant evolution and development in a post-genomic context. *Nat. Rev. Genet.* **2**:607–619.
- . 2002. Perspectives and paradigms in plant evo-devo. Pp. 1–14 in Q. C. B. Cronk, R. M. Bateman, and J. A. Hawkins, eds. *Developmental genetics and plant evolution*. Taylor and Francis, London.
- Doebley, J., and L. Lukens. 1998. Transcriptional regulators and the evolution of plant form. *Plant Cell* **10**:1075–1082.
- Ferguson, I. K., and S. C. Tucker (eds.). 1994. *Advances in Legume Systematics 6: Structural Botany*. Royal Botanic Gardens Kew, Richmond, Surrey, UK.
- Goldman, N., J. P. Anderson, and A. G. Rodrigo. 2000. Likelihood-based tests of topologies in phylogenetics. *Syst. Biol.* **49**:652–670.
- Goldman, N., and Z. Yang. 1994. A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Mol. Biol. Evol.* **11**:725–736.
- Hellberg, M. E., G. W. Moy, and V. D. Vacquier. 2000. Positive selection and propeptide repeats promote rapid interspecific divergence of a gastropod sperm protein. *Mol. Biol. Evol.* **17**:458–466.
- Heyn, C. C., and I. Herrnstadt. 1977. Seed coat of Old World *Lupinus* species. *Bot. Not.* **130**:427–435.
- Hileman, L. C., and D. A. Baum. 2003. Why do paralogs persist? Molecular evolution of *CYCLOIDEA* and related floral symmetry genes in Antirrhineae (Veronicaceae). *Mol. Biol. Evol.* **20**:591–600.
- Huelsenbeck, J. P., and F. R. Ronquist. 2001. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**:745–755.
- Lukens, L., and J. Doebley. 2001. Molecular evolution of the teosinte branched gene among maize and related grasses. *Mol. Biol. Evol.* **18**:627–638.
- Luo, D., R. Carpenter, L. Copsey, C. Vincent, J. Clark, and E. Coen. 1999. Control of organ asymmetry in flowers of *Antirrhinum*. *Cell* **99**:367–376.
- Luo, D., R. Carpenter, C. Vincent, L. Copsey, and E. S. Coen. 1996. Origin of floral asymmetry in *Antirrhinum*. *Nature* **383**:794–799.
- Metz, E. C., and S. R. Palumbi. 1996. Positive selection and sequence rearrangements generate extensive polymorphism in the gamete recognition protein bindin. *Mol. Biol. Evol.* **13**:397–406.
- Muse, S. V., and B. S. Gaut. 1994. A likelihood approach for comparing synonymous and nonsynonymous nucleotide substitution rates, with application to the chloroplast genome. *Mol. Biol. Evol.* **11**:715–724.
- Muse, S. V., and S. L. K. Pond. 2002. HYPHY: hypothesis testing using phylogenies. Distributed by the authors from <http://www.hyphy.org>.
- Nylander, J. A. A. 2002. MrModeltest, version 1.1b. Distributed by the author, Department of Systematic Zoology, EBC, Uppsala University, Sweden; [johan.nylander@ebc.uu.se](mailto:johan.nylander@ebc.uu.se).
- Ohno, S. 1970. *Evolution by gene duplication*. Springer-Verlag, New York.
- Ouali, M., and R. D. King. 2000. Cascaded multiple classifiers for secondary structure prediction. *Prot. Sci.* **9**:1162–1176.
- Posada, D., and K. A. Crandall. 2001. Selecting the best-fit model of nucleotide substitution. *Syst. Biol.* **50**:580–601.
- Remington, D. L., and M. D. Purugganan. 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.
- Shimodaira, H., and M. Hasegawa. 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Mol. Biol. Evol.* **16**:1114–1116.
- Swofford, D. L. 1998. PAUP\* 4.0. Sinauer Associates, Sunderland, Mass.
- White, T. J., T. Bruns, S. Lee, and J. Taylor. 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. Pp. 315–322 in M. Innis, D. Gelfand, J. Sninsky, and T. J. White, eds. *PCR Protocols: A guide to methods and applications*. Academic Press, San Diego, Calif.
- Yang, Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput. Appl. Biosci.* **13**:555–556.
- . 1998. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. *Mol. Biol. Evol.* **15**:568–573.
- Yang, Z., and R. Nielsen. 2002. Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. *Mol. Biol. Evol.* **19**:908–917.
- Yang, Z., R. Nielsen, N. Goldman, and A.-M. K. Pederson. 2000. Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* **155**:431–449.

Brandon Gaut, Associate Editor

Accepted September 22, 2003