

# Evidence for Positive Selection on the Floral Scent Gene Isoeugenol-*O*-methyltransferase

Todd J. Barkman

Department of Biological Sciences, Western Michigan University, Kalamazoo, Michigan

Isoeugenol-*O*-methyltransferase (IEMT) is an enzyme involved in the production of the floral volatile compounds methyl eugenol and methyl isoeugenol in *Clarkia breweri* (Onagraceae). IEMT likely evolved by gene duplication from caffeic acid-*O*-methyltransferase followed by amino acid divergence, leading to the acquisition of its novel function. To investigate the selective context under which IEMT evolved, maximum likelihood methods that estimate variable  $d_N/d_S$  ratios among lineages, among sites, and among a combination of both lineages and sites were utilized. Statistically significant support was obtained for a hypothesis of positive selection driving the evolution of IEMT since its origin. Subsequent Bayesian analyses identified several sites in IEMT that have experienced positive selection. Most of these positions are in the active site of IEMT and have been shown by site-directed mutagenesis to have large effects on substrate specificity. Although the selective agent is unknown, the adaptive evolution of this gene may have resulted in increased effectiveness of pollinator attraction or herbivore repellence.

## Introduction

Volatile compounds mediate many interactions between organisms, including interplant signaling in response to pathogen infection (Shulaev et al. 1997), plant-parasitoid signaling in response to herbivory (Turlings, Tumlinson, and Lewis 1990), and plant-pollinator communication during flowering. As pollinator attractants, volatiles are important cues that help insects locate flowers and signal the presence of food or mates (Knudsen, Tollsten, and Bergstrom 1993). The floral scent chemical compositions of hundreds of species have been enumerated; however, only recently has the molecular genetic basis of the biosynthesis of these compounds begun to be elucidated. Numerous studies in *Clarkia breweri* (Onagraceae) and *Antirrhinum majus* (Scrophulariaceae) have revealed the identity and numbers of genes involved in scent production and their expression patterns (Wang et al. 1997; Dudareva et al. 1998; Nam, Dudareva, and Pichersky 1999; Ross et al. 1999; Dudareva et al. 2000; Kolosova et al. 2001). Within *Clarkia*, *C. breweri* is the only species with strongly scented flowers, suggesting that floral fragrance production has recently evolved in this taxon (Raguso and Pichersky 1995). One possible function of floral scent in *C. breweri* may be the attraction of pollinating moths as suggested by electroantennogram studies showing that the floral volatiles elicit physiological responses in its primary hawkmoth pollinators (Raguso, Light, and Pichersky 1996; Raguso and Light 1998). An additional role of floral fragrance may be to repel herbivores, because many of the same volatile compounds produced in flowers are also released from leaves in response to herbivore damage (Kessler and Baldwin 2001). A novel scent-producing gene, Isoeugenol-*O*-methyltransferase (IEMT), has been characterized in *C. breweri* and catalyzes the production of methyleugenol and methylisoeugenol from eugenol and isoeugenol using the methyl-group donor *S*-adenosyl-L-methionine (SAM). This gene has high levels of sequence similarity to caffeic

acid-*O*-methyltransferase (COMT), which is a ubiquitous enzyme in plants that participates in the biosynthesis of lignin by converting caffeic acid to ferulic acid (Wang and Pichersky 1998). IEMT, likely duplicated from COMT, is expressed only in floral tissue of *C. breweri*, suggesting a potential role in pollinator attraction (Wang et al. 1997). Elegant site-directed mutagenesis studies of IEMT and COMT from *C. breweri* have shown that very few amino acids in a small region of the coding sequences are responsible for substrate specificity (Wang and Pichersky 1999). Although much is known about the molecular biology of IEMT, little attention has been paid to the selective context leading to its evolutionary divergence and acquisition of novel function.

The importance of positive Darwinian selection as a process shaping the evolution of protein coding genes has been suggested by numerous recent studies (Zhang, Rosenberg, and Nei 1998; Zanotto et al. 1999; Bishop, Dean, and Mitchell-Olds 2000; Yang and Bielawski 2000 and references cited therein; Swanson et al. 2001). The study of selection on protein coding genes relies on estimates of the nonsynonymous/synonymous ( $d_N/d_S$ ) rate ratio,  $\omega$ . Cases in which  $\omega = 1$  suggest the genes in question evolve neutrally, whereas purifying, or negative, selection is inferred when  $\omega < 1$ . Evidence for positive selection is obtained only in cases where  $\omega > 1$  (Yang and Bielawski 2000). Recently, maximum likelihood methods have been developed for the detection of positive selection among lineages (Yang 1998), among sites within a gene (Nielsen and Yang 1998; Yang et al. 2000), or among sites within specific lineages of a phylogeny (Yang and Nielsen 2002). In addition to simply inferring whether positive selection has been an important force shaping protein evolution, a recently developed Bayesian method has been applied to identify the most likely sites under positive selection in cases where estimates of  $\omega > 1$  (Nielsen and Yang 1998). In this paper, these recently developed methods are utilized to investigate the evolution of the floral scent gene, IEMT, in *Clarkia breweri*.

## Materials and Methods

A total of 13 sequences were analyzed from GenBank (fig. 1). These included IEMT from *C. breweri* and 12

Key words: *Clarkia breweri*, evolution of floral scent, gene duplication, maximum likelihood, positive selection.

E-mail: tbarkman@wmich.edu.

*Mol. Biol. Evol.* 20(2):168–172. 2003

DOI: 10.1093/molbev/msg030

© 2003 by the Society for Molecular Biology and Evolution. ISSN: 0737-4038

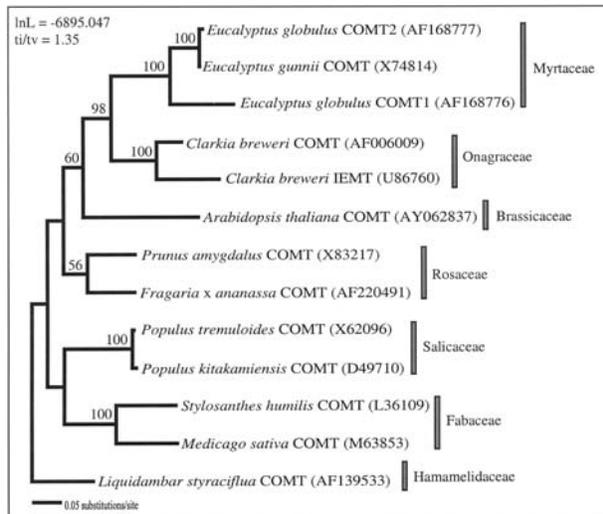


FIG. 1.—Phylogenetic relationships among COMT sequences from multiple representative rosoid families and IEMT from *Clarkia breweri*. GenBank accession number and family are listed next to each species. Tree was estimated using maximum likelihood (HKY85). Bootstrap proportions (>50) are listed above each branch.

closely related COMT sequences. Maximum likelihood estimates of the COMT/IEMT phylogeny (fig. 1) were obtained assuming the HKY85 model of nucleotide substitution using PAUP\*4.0 (Swofford 1998). The ti/tv ratio was estimated from the data and 100 random addition sequences were used during heuristic searches. Bootstrap proportions (Felsenstein 1985) were obtained from 500 pseudoreplicates with the ti/tv ratio estimated during each search. Neighbor-joining (assuming the HKY 85 model of nucleotide substitution) and unweighted parsimony were also used to estimate phylogenetic relationships among the genes. In all cases, the trees estimated using various models of nucleotide substitution or methods were similar in branching pattern and level of bootstrap support to the ML tree shown in figure 1. The strongly supported relationships implied by this tree estimate are congruent with traditional taxonomic treatments (Cronquist 1981) and recent phylogenetic studies based on independent data sets (Angiosperm Phylogeny Group 1998). PAML 3.1 (Yang 2000) was used for all likelihood calculations and estimates of  $\omega$  for all of the models used assuming the tree shown in figure 1.

## Results and Discussion

A series of likelihood ratio tests (LRTs) were performed to test hypotheses related to the molecular evolution of IEMT. The first test compared model M0 with the branch-specific model to detect whether selective pressures differ between the branch of interest (IEMT) and the other branches of the phylogeny. The M0 model of Goldman and Yang (1994) assumes a single  $\omega$  ratio for all nucleotide sites and lineages. The branch-specific model corresponds to that used by Yang (1998) and estimates two  $\omega$  ratios for both a “foreground” and “background” lineage. In the case examined here, the foreground lineage

**Table 1**  
Likelihood Scores and Parameter Estimates Assuming Various Models for the Methyltransferase Data

	lnL	Parameter Estimates
M0: one ratio	-7408.70	$\omega = 0.1061$
Branch-specific model: two ratios	-7391.20	$\omega_{\text{Background}} = 0.095$ $\omega_{\text{Foreground}} = 0.395$
M3: site-specific model; discrete ( $K = 3$ )	-7279.48	$\omega_0 = 0.02$ ( $P_0 = 0.60$ ) $\omega_1 = 0.20$ ( $P_1 = 0.35$ ) $\omega_2 = 0.68$ ( $P_2 = 0.05$ )
M3: site-specific model; discrete ( $K = 2$ )	-7287.28	$\omega_0 = 0.04$ ( $P_0 = 0.73$ ) $\omega_1 = 0.33$ ( $P_1 = 0.27$ )
Branch-site model	-7262.25	$\omega_0 = 0.03$ ( $P_0 = 0.59$ ) $\omega_1 = 0.31$ ( $P_1 = 0.22$ ) $\omega_2 = 1.93$ ( $P_2 = 0.19$ )

is *Clarkia breweri* IEMT, and all of the COMT sequences comprise the background lineage (fig. 1). This *a priori* assignment of ratios to different lineages is based on the previous studies of IEMT, suggesting its recent evolution from COMT (Wang and Pichersky 1998). The LRT statistic for this comparison was  $2(\text{diff. lnL}) = 35.00$  ( $P < 0.001$ , 1 *df*). This significantly better fit of the branch-specific model to the data suggests that the IEMT lineage has experienced different levels of selective constraint than the COMT lineages. The fact that  $\omega$  is 40 times greater in the foreground lineage suggests that there is either relaxed selective constraint or positive selection on IEMT.

The second test compared model M0 with M3 (discrete model), the latter of which was presented in Yang et al. (2000) to investigate whether selective pressure varies among sites. Model M3 (with three site classes) assumes that there is some proportion of sites belonging to each of three classes with different  $\omega$  ratios (table 1). In this case, both the proportions and  $\omega$  ratios for three site classes were estimated from the data. The LRT statistic for this comparison was  $2(\text{diff. lnL}) = 242.84$  ( $P < 0.001$ , 4 *df*). This significant result suggests that the  $\omega$  ratio is variable among sites in the COMT and IEMT genes, with  $\omega_0 = 0.02$  at 60% of the sites,  $\omega_1 = 0.20$  at 35% of the sites, and  $\omega_2 = 0.68$  at 5% of the sites. Although  $\omega$  appears to be heterogeneous for these sequences, this test provides no evidence for sites under positive selection in these genes, because in no case was  $\omega$  estimated to be greater than 1. However, because the same  $\omega$  ratios were estimated for all lineages, any difference in selective constraint in IEMT only could not be detected.

The third test involved a comparison of model M3 (two site classes) with the branch-sites model (two ratios and M3 with two site classes) to test whether sites are under positive selection along the IEMT branch in figure 1. M3 with two site classes was defined in Yang et al. (2000) and estimates the proportion of sites that belong to each of two site classes (table 1). The branch-sites model was recently developed in Yang and Nielsen (2002) and relies on the *a priori* specification of foreground and background lineages. Proportions of sites belonging to each of two site classes ( $\omega_0$  and  $\omega_1$ ) are estimated for all lineages, and a third site class ( $\omega_2$ ) is estimated for the foreground lineage alone (table 1). A significant LRT for this model comparison would indicate that sites in IEMT

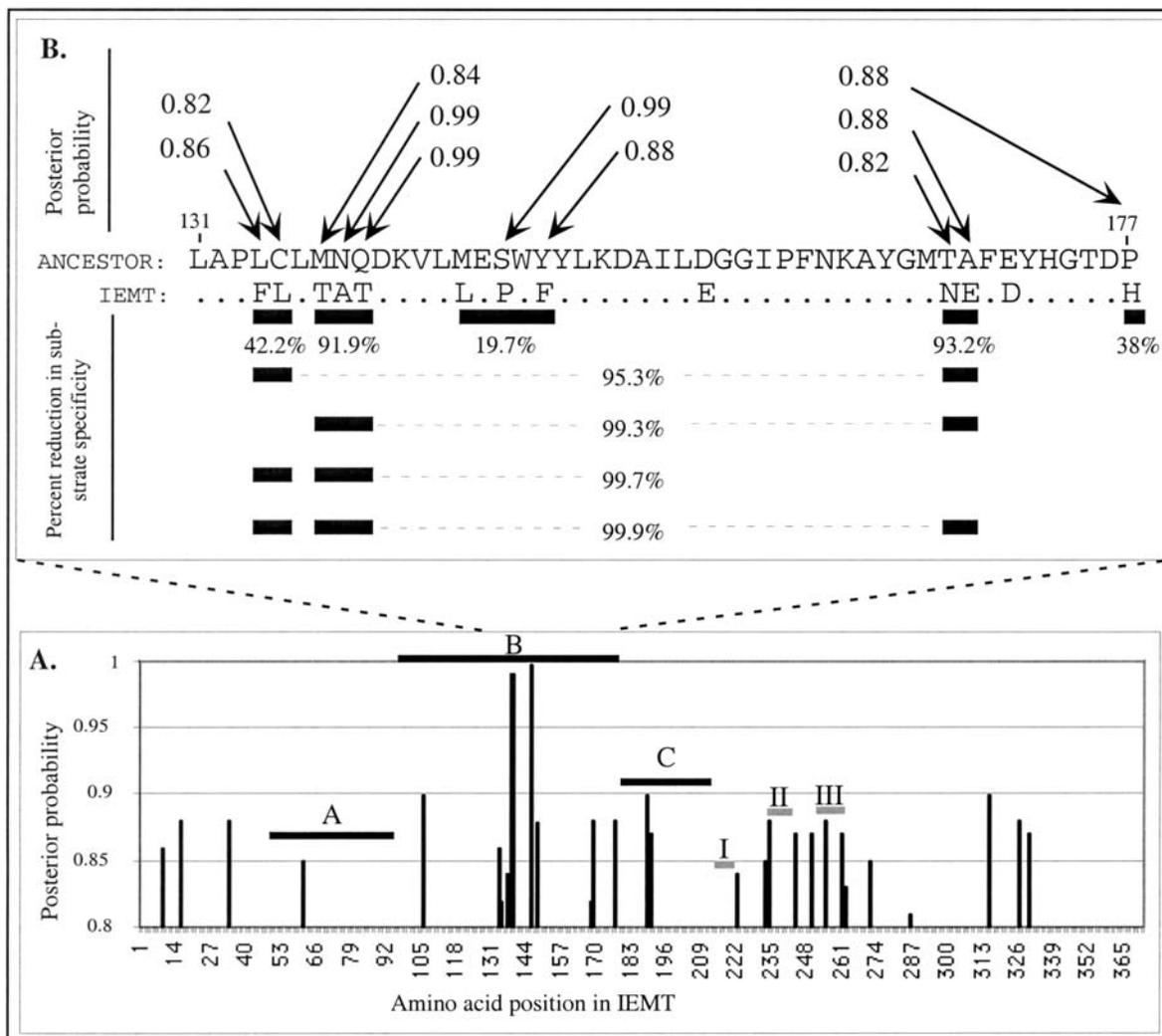


FIG. 2.—A, Posterior probabilities ( $>0.8$ ) that a particular site in IEMT is from the site class  $\omega = 1.93$  estimated with PAML. Regions A–C were used in domain swapping experiments between COMT and IEMT (Wang and Pichersky 1998). Only region B had an effect on substrate specificity of IEMT. Regions I–III are highly conserved domains among plant methyltransferases and are known to be involved in binding of the methyl group donor SAM (Gang et al. 2002). B, Amino acids 131 to 177 representing the active site of IEMT, with posterior probabilities listed for those sites inferred to be under positive selection. The underlined amino acids represent IEMT sites mutated by site-directed mutagenesis studies (Wang and Pichersky 1999). The sites were mutated in IEMT to match the amino acids present in the ancestral sequence shown. The percent reduction in the ratio of methyl isoeugenol and ferulic acid produced when using isoeugenol and caffeic acid as substrates for mutated IEMT sequences relative to wild-type IEMT is listed below the amino acids (Wang and Pichersky 1999).

are under differential selection and if  $\omega_2$  is greater than 1, this test provides evidence for positive selection (Yang and Nielsen 2002). The LRT statistic was  $2(\text{diff. lnL}) = 50.06$  ( $P < 0.001$ , 2 *df*). In the branch-sites model, 59% of sites were estimated with  $\omega_0 = 0.03$ , 22% with  $\omega_1 = 0.31$ , and 19% with  $\omega_2 = 1.93$ . These results strongly suggest that there has been positive selection on some sites in the IEMT lineage since its divergence from COMT.

Assuming the branch-sites model, PAML 3.1 was used to calculate the posterior probabilities that a particular site belongs to the site class that has experienced positive selection. As shown in figure 2A, 30 sites were assigned to the site class with  $\omega = 1.93$  (posterior probability  $> 0.8$ ). Of these substitutions, 11 were in the region of the coding sequence that has been shown to be important for the

different substrate specificities of COMT and IEMT based on chimeric COMT-IEMT protein constructs (region B in fig. 2A) (Wang and Pichersky 1998). Furthermore, within this region, all of the sites shown by site-directed mutagenesis to have major effects on substrate discrimination of IEMT had high posterior probabilities (fig. 2B) (Wang and Pichersky 1999). It should also be noted that two of the sites with the highest posterior probabilities (138 and 139) were shown to have some of the largest effects on substrate discrimination.

The branch-sites model that allowed for sites to be under positive selection within IEMT provided a statistically better fit to the data than models without such a parameter. The fact that  $\omega_2 > 1$  suggests that some substitutions in this enzyme are adaptive and should have

positive effects on fitness. Interestingly, the overall spatial pattern of positively selected amino acid replacements in IEMT is not homogeneous. There are relatively few positively selected sites in the first and last 100 amino acids, whereas there are two to three times more substitutions between positions 131 to 177 and 222 to 274. This heterogeneous pattern of amino acid replacement would be expected for a gene that has evolved by duplication and functional divergence because most of the substitutions occur in the catalytic regions of IEMT. The fact that most of the positively selected sites in IEMT have significant effects on substrate specificity (fig. 2B) suggests that adaptive evolution of this gene has been driven by a selective agent that is sensitive to the activities of this protein.

If it is assumed that efficient biosynthesis of methyl eugenol and methyl isoeugenol is important for the fitness of *Clarkia breweri*, then the inference that multiple IEMT sites have evolved by positive selection is supported by the experimental studies on enzyme kinetics (Wang and Pichersky 1999). Unlike proteins that have experienced Darwinian selection for increased proportions of charge changing residues (e.g., arginine in eosinophil cationic protein [Zhang, Rosenberg, and Nei 1998]), it appears that IEMT has been selected for increased substrate specificity. Although the order in which mutations accumulated is unknown, experimental studies do suggest that the many of the sites shown to be under positive selection could have had synergistic effects on the ancestral IEMT substrate specificity. For example, Wang and Pichersky (1999) demonstrated that mutations of IEMT at positions 134 and 135 led to an enzyme with a 42% lower ability to discriminate substrates than wild-type (fig. 2B). Combined mutations of sites 134 and 135 and 168 and 169 reduced the substrate specificity by 95% compared with wild-type (fig. 2B). Finally, it was shown that mutations of 134 and 135, 168 and 169, and 137 and 139 resulted in an enzyme with almost no ability to catalyze the formation of methyl isoeugenol. If the selective agent were sensitive to varying levels of emissions, then the pattern of mutations in the active site of this enzyme is consistent with positive Darwinian selection for increased substrate specificity because the rate of volatile production is probably directly related to substrate preference. Further studies are needed to determine the role of IEMT because the volatiles it produces may attract pollinators, repel herbivores from the flowers, or have some other as of yet unknown function.

In a general context, the results of this study have implications for the study of genes involved in secondary metabolism. Much like IEMT, only a small number of amino acid changes were responsible for novel substrate specificities in another methyltransferase, eugenol-*O*-methyltransferase (EOMT), relative to its sister gene, chavicol-*O*-methyltransferase (Gang et al. 2002). If positive selection is localized to a few codons in one or a few lineages, the recently developed branch-sites model used here will likely become increasingly important in the study of adaptive evolution of genes involved in secondary metabolism.

## Acknowledgments

Ziheng Yang, DeWayne Shoemaker, and two anonymous reviewers are thanked for comments and suggestions on earlier drafts of this manuscript.

## Literature Cited

- Angiosperm Phylogeny Group. 1998. An ordinal classification of the families of flowering plants. *Ann. Mo. Bot. Gard.* **85**:531–553.
- Bishop, J. G., A. M. Dean, and T. Mitchell-Olds. 2000. Rapid evolution in plant chitinases: molecular targets of selection in plant-pathogen coevolution. *Proc. Natl. Acad. Sci. USA* **97**:5322–5327.
- Cronquist, A. 1981. An integrated system of classification of flowering plants. Columbia University Press, New York.
- Dudareva, N., R. A. Raguso, J. Wang, J. R. Ross, and E. Pichersky. 1998. Floral scent production in *Clarkia breweri*. *Plant Physiol.* **116**:599–604.
- Dudareva, N., L. M. Murfitt, C. J. Mann, N. Gorenstein, N. Kolosova, C. M. Kish, C. Bonham, and K. Wood. 2000. Developmental regulation of methyl benzoate biosynthesis and emission in snapdragon flowers. *Plant Cell* **12**:949–961.
- Felsenstein, J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39**:783–791.
- Gang, D., N. Lavid, C. Zubieta, F. Chen, T. Beuerle, E. Lewinsohn, J. P. Noel, and E. Pichersky. 2002. Characterization of phenylpropene *O*-methyltransferases from sweet basil: facile change of substrate specificity and convergent evolution within a plant *O*-methyltransferase family. *Plant Cell* **14**:505–519.
- Goldman, N., and Z. Yang. 1994. A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Mol. Biol. Evol.* **11**:725–736.
- Kessler, A., and I. T. Baldwin. 2001. Defensive function of herbivore-induced plant volatile emissions in nature. *Science* **291**:2141–2144.
- Knudsen, J. T., L. Tollsten, and G. Bergstrom. 1993. Floral scents—a checklist of volatile compounds isolated by head-space techniques. *Phytochemistry* **33**:253–280.
- Kolosova, N., N. Gorenstein, C. M. Kish, and N. Dudareva. 2001. Regulation of circadian methyl benzoate emission in diurnally and nocturnally emitting plants. *Plant Cell* **13**:2333–2347.
- Nam, K. H., N. Dudareva, and E. Pichersky. 1999. Characterization of benzylalcohol acetyltransferases in scented and non-scented *Clarkia* species. *Plant Cell Physiol.* **40**:916–923.
- Nielsen, R., and Z. Yang. 1998. Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics* **148**:929–936.
- Raguso, R. A., and D. M. Light. 1998. Electroantennogram responses of male *Sphinx perelegans* hawkmoths to floral and ‘green-leaf volatiles.’ *Entomol. Exp. Appl.* **86**:287–293.
- Raguso, R. A., D. M. Light, and E. Pichersky. 1996. Electroantennogram responses of *Hyles lineata* (Sphingidae: Lepidoptera) to volatile compounds from *Clarkia breweri* (Onagraceae) and other moth-pollinated flowers. *J. Chem. Ecol.* **22**:1735–1766.
- Raguso, R. A., and E. Pichersky. 1995. Floral volatiles from *Clarkia breweri* and *C. concinna* (Onagraceae): recent evolution of floral scent and moth pollination. *Plant Syst. Evol.* **194**:55–67.
- Ross, J. R., K. H. Nam, J. C. D’Auria, and E. Pichersky. 1999. S-Adenosyl-L-Methionine: salicylic acid carboxyl methyltransferase, an enzyme involved in floral scent production and

- plant defense, represents a new class of plant methyltransferases. *Arch. Biochem. Biophys.* **367**:9–16.
- Shulaev, V., P. Silverman, and I. Raskin. 1997. Airborne signalling by methyl salicylate in plant pathogen resistance. *Nature* **385**:718–721.
- Swanson, W. J., Z. Yang, M. F. Wolfner, and C. F. Aquadro. 2001. Positive Darwinian selection drives the evolution of several female reproductive proteins in mammals. *Proc. Natl. Acad. Sci. USA* **98**:2509–2514.
- Swofford, D. L. 1998. PAUP\*: phylogenetic analysis using parsimony (\*and other methods). Version 4.0. Sinauer Associates, Sunderland, Mass.
- Turlings, T. C. J., J. H. Tumlinson, and W. J. Lewis. 1990. Exploitation of herbivore-induced plant odors by host-seeking parasitic wasps. *Science* **250**:1251–1253.
- Wang, J., and E. Pichersky. 1998. Characterization of S-adenosyl-L-methionine: (iso)eugenol *O*-methyltransferase involved in floral scent production in *Clarkia breweri*. *Arch. Biochem. Biophys.* **349**:153–160.
- . 1999. Identification of specific residues involved in substrate discrimination in two plant *O*-methyltransferases. *Arch. Biochem. Biophys.* **368**:172–180.
- Wang, J., N. Dudareva, S. Bhakta, R. A. Raguso, and E. Pichersky. 1997. Floral scent production in *Clarkia breweri* (Onagraceae) II. Localization and developmental modulation of the enzyme S-adenosyl-L-methionine: (iso)eugenol *O*-methyltransferase and phenylpropanoid emission. *Plant Physiol.* **114**:213–221.
- Yang, Z. 1998. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. *Mol. Biol. Evol.* **15**:568–573.
- . 2000. Phylogenetic analysis by maximum likelihood (PAML). Version 3.0. University College, London.
- Yang, Z., and J. P. Bielawski. 2000. Statistical methods for detecting molecular adaptation. *Trends Ecol. Evol.* **15**:496–503.
- 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. Pedersen. 2000. Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* **155**:431–449.
- Zanotto, P. M. de A., E. G. Kallas, R. F. de Souza, and E. C. Holmes. 1999. Genealogical evidence for positive selection in the *nef* gene of HIV-1. *Genetics* **153**:1077–1089.
- Zhang, J., H. F. Rosenberg, and M. Nei. 1998. Positive Darwinian selection after gene duplication in primate ribonuclease genes. *Proc. Natl. Acad. Sci. USA* **95**:3708–3713.

Geoffrey McFadden, Associate Editor

Accepted September 24, 2002