

LETTERS

Directed Mutagenesis Confirms the Functional Importance of Positively Selected Sites in Polygalacturonase Inhibitor Protein

John G. Bishop¹

Max Planck Institute For Chemical Ecology, Jena, Germany; and School of Biological Sciences, Washington State University

Polygalacturonase inhibitor proteins (PGIPs) protect plants against invasion by diverse microbial and invertebrate enemies that use polygalacturonase (PG) to breach the plant cell wall. Directed mutagenesis has identified specific natural mutations conferring novel defensive capability in green bean PGIP against a specific fungal PG. These same sites are identified as positively selected by phylogenetic codon-substitution models, demonstrating the utility of such models for connecting retrospective comparative analyses with contemporary, ecologically relevant variation.

Introduction

Phylogeny-based models of codon substitution have significantly advanced the goal of detecting selection on protein-encoding DNA segments (Nielsen and Yang 1998; Yang et al. 2000). By identifying individual codons that experience advantageous substitutions, the models offer a valuable tool for studying adaptation at the molecular level that can complement structural and biochemical tools. In several systems inferred, positively selected sites occur in regions known from structural or functional analyses to be involved in protein-protein or protein-ligand interactions (e.g., Bishop, Dean, and Mitchell-Olds 2000; Urwin et al. 2002; Yang and Swanson 2002; Bishop et al. 2005). Despite the surprising discovery of congruence between statistical analysis of sequence variation and other approaches, there is concern that the codon-based analyses may produce numerous false positives (Suzuki and Nei 2001, 2004). Although judicious application of the models (Anisimova, Bielawski, and Yang 2001, 2002; Wong et al. 2004) and improved methods (Yang, Wong, and Nielsen, 2005) appear to mitigate many of these concerns, further confirmation of their reliability is desirable. A useful follow-up to the codon-substitution analyses is to experimentally manipulate putative positively selected codons and perform subsequent protein- or organismal-level assays to determine the functional significance of variation at these sites. Such manipulations are not expected to affect function if codon models are prone to misclassification of neutrally evolving sites.

Even when sites with a history of adaptive substitutions are correctly identified, it is not obvious whether extant variants at these sites should confer contemporary functional variation, owing to the retrospective nature of the analyses. A common thread in most cases of positive selection is that they involve proteins in which an arms race or some other coevolutionary process may be responsible for rapid protein evolution (i.e., reproductive genes [Swanson and Vacquier 2002], gastropod toxins used to

subdue prey [Duda and Palumbi 1999, 2000], salamander courtship pheromones that increase female receptivity [Watts et al. 2004], host immune response genes [Urwin et al. 2002; Yang and Swanson 2002; Wang et al. 2003], and plant defense-related glycanhydrolases [Bishop, Dean, and Mitchell-Olds 2000; Bishop et al. 2005]). Identifying positively selected sites in these systems might only constitute resurrecting the ghost of ancient adaptations, left behind by rapid coevolution. On the other hand, a diverse spatial mosaic of coevolutionary outcomes might maintain numerous extant functional variants, especially among gene family members.

Polygalacturonase inhibitor proteins (PGIPs) are secreted by plants into the cell wall where they defend the wall against polygalacturonases (PGs) secreted by diverse plant enemies, including fungi and hemipteran insects (De Lorenzo and Ferrari 2002; D'Ovidio et al. 2004). Previous studies uncovered evidence of positive selection acting on the PGs of fungi and oomycetes as well as on the PGIPs of dicotyledonous plants, strongly suggestive of arms race dynamics (Stotz et al. 2000; Götesson et al. 2002). Green bean (*Phaseolus vulgaris* var. BAT93, hereafter "PvB") and pinto bean (*P. vulgaris* var. Pinto, hereafter "PvP") each harbor four PGIPs (PGIP1–PGIP4) that differ in their effectiveness against various PGs produced by plant enemies. PvP_PGIP2 is the only protein capable of inhibiting *Fusarium moniliforme* PG (FmPG), whereas Pv_PGIP3 and Pv_PGIP4 are the only ones capable of inhibiting hemipteran PGs (D'Ovidio et al. 2004). PvP_PGIP1 and PvP_PGIP2 also differ in inhibitory activity toward the PGs of other pathogens but are separated by only eight amino acid substitutions.

By using directed mutagenesis to recreate each of the eight substitutions differentiating PvP_PGIP1 and PvP_PGIP2, Leckie et al. (1999) demonstrated that a single substitution, K224Q, causes PvP_PGIP1 to gain the ability to bind FmPG. Conversely, Q224K caused PvP_PGIP2 to lose 70% of its inhibitory activity toward FmPG, and in combination with substitutions V152G or A297S all inhibition of FmPG was abolished. Mutations Q224K and A152G also decreased PGIP2 affinity for *Aspergillus niger* PG (AnPG), but to a much lesser extent.

To test whether any of the functionally important substitutions identified by Leckie et al. reside at sites predicted

¹ Present address: Max Planck Institute for Chemical Ecology, Hans-Knoll Strasse 8, Jena, Germany 07745.

Key words: polygalacturonase inhibitor protein, directed mutagenesis, bean, *Phaseolus vulgaris*, positive selection, codon-substitution models.

E-mail: bishop@vancouver.wsu.edu.

Mol. Biol. Evol. 22(7):1531–1534, 2005

doi:10.1093/molbev/msi146

Advance Access publication April 13, 2005

Table 1
Positively Selected and Manipulated Sites

Site ^a	Res ^b	CODEML		FEL		Beta (NEB) 1-Post	Beta (BEB) 1-Post	FEL <i>P</i> value
		Non	Syn	dN	dS			
39	3	3	0	1.6	0		0.21	0.09
(152 ^c)	4 ^d	6	1	4.2	1.5	0.10	0.24	0.47
175	4	5	2	4.2	2.6	0.02	0.06	0.74
178^e	5^d	6	0	4.7	0	0.007	0.03	0.02
200	4 ^d	5	0	2.8	0	0.03	0.11	0.05
224^f	6^{d,g}	6	0	3.1	0	0.03	0.10	0.12
232	4	7	0	4.3	0	0.009	0.04	0.36
271	6 ^d	6	1	4.7	2.9	0.015	0.06	0.43
272 ^c	4 ^d	5	0	2.0	0	0.11	0.24	0.10
296	3	3	0	1.5	0	0.35		0.08
(297 ^c)	4	3.5	2.5	1.5	4.6	0.34	0.50	

^a Selected sites were included if the posterior probability ≥ 0.9 in CODEML M8 (beta model) using BEB method or $P \leq 0.1$ in FEL. Sites with superscripts were manipulated by Leckie et al. (1999), and those in parentheses are not positively selected. Results for all models are given in Supplementary table 2 (Supplementary Material online).

^b Res: number of different residues per site among the 12 sequences; Non and Syn: nonsynonymous and synonymous substitutions (CODEML); dN and dS: nonsynonymous and synonymous rates for each site estimated by FEL. NEB: Naive Empirical Bayes.

^c Directed mutagenesis caused some effects on specificity.

^d Site is within leucine-rich repeat region.

^e Directed mutagenesis detected no effect on specificity.

^f Directed mutagenesis strongly affected specificity.

^g Includes a deletion in PvP_PGIP1.

to experience strong positive selection, codon-evolution models were applied to a set of 12 PGIPs comprising four paralogs each from PvP and PvB as well as four PGIPs from the closely related species *Glycine max* (soybean) (see *Methods*). All CODEML models found significant evidence of positive selection (e.g., M8 vs. M7 LRT = 20.4, 2 df, $P < 0.0002$; Supplementary table 1, Supplementary Material online). Nine out of 318 total sites were identified as positively selected by fixed effects likelihood (FEL) or CODEML (using the conservative Bayes empirical Bayes [BEB] method), including three sites manipulated

by Leckie et al. (table 1). Inspection of the PvP_PGIP2 crystal structure reveals that seven of the nine positively selected sites are arrayed around the negatively charged pocket implicated in binding PG (fig. 1) (Di Matteo et al. 2003). Five of these seven are within the leucine-rich repeat region implicated in protein-protein interactions. Docking models of PvP_PGIP2 with FmPG (1HG8) indicate that selected residues to the right of the binding pocket may interact with the loops 113–124 and 177–185 on the margin of the active site cleft of FmPG (Supplementary Material online).

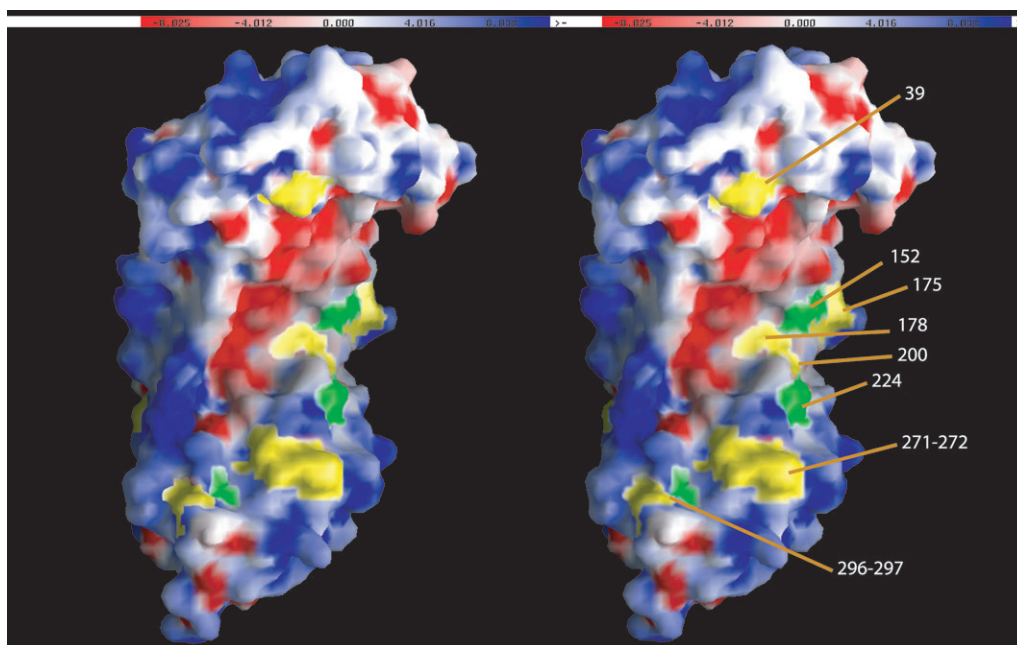


FIG. 1.—Electrostatic potential surface of green bean PGIP2 (1OGQ), with areas of strong negative charge in red and positive charge in blue. A negative pocket (arrow) is thought to bind the PG ligand. Positively selected sites (yellow) and sites shown to affect binding of FmPG (green) are labeled.

Site 224, responsible for conferring inhibition of FmPG, was estimated to have sustained six nonsynonymous substitutions and was identified as positively selected by the CODEMLs beta (M8) and discrete (M3) models but not by FEL. Site 178 also sustained six coding substitutions and was the only site clearly identified by both FEL and CODEML models. Mutation S178A was found by Leckie et al. (their S207A) to have no effect on inhibition of FmPG or AnPG.

Of the other two sites known to affect FmPG inhibition, site 152 was positively selected only in the least conservative models (Supplementary table 2, Supplementary Material online) but was also identified by an analysis of dicot PGIPs (Stotz et al. 2000). Although site 297 was never significant, it was included in the positively selected class by most CODEML models, and neighboring sites 296 and 295 were significant in the FEL model and in dicots (Stotz et al. 2000), respectively. The three residues surround a prominent negative bulge formed by Asp294 on the inhibitor's flank (fig. 1), providing strong evidence of selection on this region if not precisely on site 297.

A common concern regarding codon-evolution models is that they misclassify variable, but neutrally evolving, sites as positively selected. In PGIP, the selection analysis identified nine sites that are most likely involved in adaptive evolution. The adaptive importance of one, site 224, is proven, and the location of the others suggests that experiments against a more diverse set of ligands (plant pathogenic bacteria, oomycetes, and nematodes also produce PGs) will reveal important functional variants. Analogous results were obtained in recent studies of vertebrate glutathione *S*-transferases (GST) and primate TRIM5 α (Ivarsson et al. 2003; Sawyer et al. 2005). Ivarsson et al. (2003) identified three positively selected sites, two of which differed between human GST paralogs possessing different substrate affinities. A biochemically "conservative" Thr \rightarrow Ser substitution at one of these sites in one paralog conferred a 10^3 increase in the ability to process the substrate of the second paralog. In TRIM5 α , the codon-evolution models identified a 13 residue "patch" of positively selected sites that, when swapped between human and rhesus monkey orthologs, accounted for species-specific antiretroviral activity. Taken together, these studies suggest that codon-substitution models, when judiciously applied, will frequently identify sites sustaining adaptive mutations.

Codon-based selection models provide a novel method for sorting through and prioritizing substitutions for experimental analysis of protein adaptation. However, even when a very small proportion of sites is identified, the large number of substitutions occurring at these sites makes for an array of possibilities requiring DNA shuffling and high throughput assays. With four to six different residues per site, the nine selected sites in PGIP yield $\sim 2 \times 10^6$ substitutional combinations, but applying the results to paralogs or haplotypes with known functional differences, as in PGIP and GST, can greatly circumscribe the relevant substitutions. For example, with differences at ~ 70 of 318 sites, experimentally investigating the basis of fungal PG versus insect PG inhibition that contrast PGIP2 and PGIP4 is difficult to contemplate, but the 7 positively

selected sites at which the two paralogs differ provides a reasonable entry point. Thus, in the absence of additional structural or stereochemical considerations, treating the codon model results as testable hypotheses will advance the study of molecular adaptation and enhance further model development.

Methods

Positively selected sites were identified in a set of 12 green bean (*P. vulgaris*) and soybean (*G. max*) PGIPs using the modified maximum likelihood codon-evolution models implemented in CODEML (PAML 3.14) with settings to account for codon bias. With a total sequence divergence of 2.43 substitutions per codon, this represents a moderately diverged data set for which these models should have relatively high power and accuracy (Anisimova, Bielawski, and Yang 2002). Models M2 and M8 apply a more conservative method for assigning sites to rate categories that accounts for uncertainty in parameter estimates when calculating posterior probabilities by using the BEB method (Z. Yang, W. S. W. Wong, and R. Nielsen, unpublished data). A phylogeny for use with these models was estimated using MrBayes 3.0, which was identical to the tree produced by Neighbor-Joining. Accession numbers, alignments, and phylogeny are in the Supplementary Material online. In addition, the FEL model implemented in HyPhy was applied (Pond, Frost, and Muse 2004). The FEL model is a generalization of the models implemented in CODEML but differs in that FEL explicitly models site-to-site synonymous rate variation and estimates a *P* value for $dN > dS$ for each site (Pond and Frost 2005). For FEL, we used HKY85 as the substitution model, which was identified as the best matrix by HyPhy's model selection algorithm. To allow the models to count indels as coding changes, the corresponding gaps were recoded by replacement with codons chosen so as to count each deleted or inserted gap as a single nonsynonymous substitution based on the codon present in the most closely related sequence. Because of the slight distances between neighboring sequences, replacing the three codons was judged unlikely to lead to undercounting of synonymous substitutions. Three-dimensional structure was represented based on the crystal structure of PvP_PGIP2 (1OGQ, Di Matteo et al. 2003), using Swiss PDB Viewer 3.7. Site E1 corresponds to the first residue of the mature PvB_PGIP1, as in D'Ovidio et al. (2004).

Supplementary Material

Supplementary tables 1 and 2 are available at *Molecular Biology and Evolution* online (www.mbe.oupjournals.org).

Acknowledgments

I thank C. Palmer and J. Kroymann for discussion, H. Stotz and two anonymous reviewers for comments, and T. Mitchell-Olds and the Max Planck Institute for Chemical Ecology for support. Daniel Ripoll of the Cornell Theory Center kindly provided figure 1.

Literature Cited

- Anisimova, M., J. P. Bielawski, and Z. Yang. 2001. Accuracy and power of the likelihood ratio test in detecting adaptive molecular evolution. *Mol. Biol. Evol.* **18**:1585–1592.
- . 2002. Accuracy and power of Bayes prediction of amino acid sites under positive selection. *Mol. Biol. Evol.* **19**:950–958.
- 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.
- Bishop, J. G., D. R. Ripoll, S. Bashir, C. M. B. Damasceno, J. D. Seeds, and J. K. C. Rose. 2005. Selection on glycine beta-1,3-endoglucanase genes differentially inhibited by a *Phytophthora* glucanase inhibitor protein. *Genetics* **169**:1009–1019.
- De Lorenzo, G., and S. Ferrari. 2002. Polygalacturonase-inhibiting proteins in defense against phytopathogenic fungi. *Curr. Opin. Plant Biol.* **5**:295–299.
- Di Matteo, A., L. Federici, B. Mattei, G. Salvi, K. A. Johnson, C. Savino, G. De Lorenzo, D. Tsernoglou, and F. Cervone. 2003. The crystal structure of polygalacturonase-inhibiting protein (PGIP), a leucine-rich repeat protein involved in plant defense. *Proc. Natl. Acad. Sci. USA* **100**:10124–10128.
- D'Ovidio, R., A. Raiola, C. Capodicasa, A. Devoto, D. Pontiggia, S. Roberti, R. Galletti, E. Conti, D. O'Sullivan, and G. De Lorenzo. 2004. Characterization of the complex locus of bean encoding polygalacturonase-inhibiting proteins reveals sub-functionalization for defense against fungi and insects. *Plant Physiol.* **135**:2424–2435.
- Duda, T. F., and S. R. Palumbi. 1999. Molecular genetics of ecological diversification: duplication and rapid evolution of toxin genes of the venomous gastropod *Conus*. *Proc. Natl. Acad. Sci. USA* **96**:6820–68237.
- . 2000. Evolutionary diversification of multigene families: allelic selection of toxins in predatory cone snails. *Mol. Biol. Evol.* **17**:1286–1293.
- Götesson, A., J. S. Marshall, D. A. Jones, and A. R. Hardham. 2002. Characterization and evolutionary analysis of a large polygalacturonase gene family in the oomycete pathogen *Phytophthora cinnamomi*. *Mol. Plant Microbe Interact.* **15**:907–921.
- Ivarsson, Y., A. J. Mackey, M. Edalat, W. R. Pearson, and B. Mannervik. 2003. Identification of residues in glutathione transferase capable of driving functional diversification in evolution—a novel approach to protein redesign. *J. Biol. Chem.* **278**:8733–8738.
- Leckie, F., B. Mattei, C. Capodicasa, A. Hemmings, L. Nuss, B. Aracri, G. De Lorenzo, and F. Cervone. 1999. The specificity of polygalacturonase-inhibiting protein (PGIP): a single amino acid substitution in the solvent-exposed β -strand/ β -turn region of the leucine-rich repeats (LRRs) confers a new recognition capability. *EMBO J.* **18**:2352–2363.
- 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.
- Pond, S. L. K., and S. D. W. Frost. 2005. Not so different after all: comparison of various methods for detecting amino-acid sites under selection. *Mol. Biol. Evol.* **22**:1208–1222.
- Pond, S. L. K., S. D. W. Frost, and S. V. Muse. 2004. HyPhy: hypothesis testing using phylogenies. *Bioinformatics* **21**:676–679.
- Sawyer, S. L., L. I. Wu, M. Emerman, and H. S. Malik. 2005. Positive selection of primate TRIM5 α identifies a critical species-specific retroviral restriction domain. *Proc. Natl. Acad. Sci. USA* **102**:2832–2837.
- Stotz, H. U., J. G. Bishop, C. W. Bergmann, M. Koch, P. Albersheim, A. G. Darvill, and J. M. Labavitch. 2000. Identification of target amino acids that affect interactions of fungal polygalacturonases and their plant inhibitors. *Mol. Plant Pathol.* **56**:117–130.
- Suzuki, Y., and M. Nei. 2001. Reliabilities of parsimony-based and likelihood-based methods for detecting positive selection at single amino acid sites. *Mol. Biol. Evol.* **18**:2179–2185.
- . 2004. False-positive selection identified by ML-based examples from the *Sigl* gene of the diatom *Thalassiosira weissflogii* and the *tax* gene of a human T-cell lymphotropic virus. *Mol. Biol. Evol.* **21**:914–921.
- Swanson, W. J., and V. D. Vacquier. 2002. Reproductive protein evolution. *Annu. Rev. Ecol. Syst.* **33**:161–179.
- Urwin, R., E. C. Holmes, A. J. Fox, J. P. Derrick, and M. C. J. Maiden. 2002. Phylogenetic evidence for frequent positive selection and recombination in the meningococcal surface antigen PorB. *Mol. Biol. Evol.* **19**:1686–1694.
- Wang, H.-Y., H. Tang, C.-K. J. Shen, and C.-I. Wu. 2003. Rapidly evolving genes in humans. I. The glycoporphins and their possible role in evading malaria parasites. *Mol. Biol. Evol.* **20**:1795–1804.
- Watts, R. I., C. A. Palmer, R. C. Feldhoff, P. W. Feldhoff, L. D. Houck, A. G. Jones, M. E. Pfrender, S. M. Rollman, and S. J. Arnold. 2004. Stabilizing selection on behavior and morphology masks positive selection on the signal in a salamander pheromone signaling complex. *Mol. Biol. Evol.* **21**:1032–1041.
- Wong, W. S. W., Z. Yang, N. Goldman, and R. Nielsen. 2004. Accuracy and power of statistical methods for detecting adaptive evolution in protein coding sequences and for identifying positively selected sites. *Genetics* **168**:1041–1051.
- Yang, Z., R. Nielsen, N. Goldman, and A.-M. Krabbe Pedersen. 2000. Codon-substitution models for heterogenous selection pressure at amino acid sites. *Genetics* **155**:431–449.
- Yang, Z., W. S. W. Wong, and R. Nielsen. 2005. Bayes empirical Bayes inference of amino acid sites under positive selection. *Mol. Biol. Evol.* **22**:1107–1118.
- Yang, Z., and W. J. Swanson. 2002. Codon-substitution models to detect adaptive evolution that account for heterogenous selection pressures among site classes. *Mol. Biol. Evol.* **19**:49–57.

Robin Bush, Associate Editor

Accepted March 16, 2005