Positive Selection Acting on a Surface Membrane Protein of the Plant-Pathogenic Phytoplasmas†

Shigeyuki Kakizawa,¹ Kenro Oshima,¹ Hee-Young Jung,¹ Shiho Suzuki,¹ Hisashi Nishigawa,² Ryo Arashida,¹ Shin-ichi Miyata,² Masashi Ugaki,² Hirohisa Kishino,¹ and Shigetou Namba^{1*}

Division of Agricultural and Life Sciences, The University of Tokyo, 1-1-1 Yayoi, Bunkyo-ku, Tokyo 113-8657, Japan,¹
and Division of Integrated Biosciences, The University of Tokyo, 202 Bioscience Bldg., 5-1-5 Kashiwanoha,
Chiba 277-8562, Japan²

Received 18 November 2005/Accepted 17 February 2006

Phytoplasmas are plant-pathogenic bacteria that cause numerous diseases. This study shows a strong positive selection on the phytoplasma antigenic membrane protein (Amp). The ratio of nonsynonymous to synonymous substitutions was >1 with all the methods we tested. The clear positive selections imply an important biological role for Amp in host-bacterium interactions.

Phytoplasmas are an economically important group of plant-pathogenic bacteria that belong to the class *Mollicutes* and cause numerous diseases (11, 13). These organisms are characterized by their small genome size, lack of cell walls, and habitat of the cytoplasm of eukaryotic cells. Phytoplasmas infect the phloem sieve elements of plants and are transmitted between plants by phloem-feeding insect hosts (11, 13). Recently, we sequenced the 860-kb genome of the “*Candidatus* Phytoplasma asteris” OY strain, line OY-M (22). The phytoplasma genome lacks several important metabolic genes, such as the ATP synthase gene, which is possibly the result of reductive evolution during adaptation to a nutrient-rich environment. Since phytoplasmas are endocellular and lack cell walls, their membrane proteins must interact directly with the cytoplasm of the host plant and insect cells and may play important roles in host-phytoplasma interactions. However, phytoplasma surface proteins have not yet been analyzed in detail because of difficulties in culturing phytoplasmas in vitro. In most phytoplasmas, a single membrane protein, termed the immunodominant membrane protein, accounts for most of the protein in cell membrane preparations (26). Recently, genes that encode immunodominant membrane proteins have been isolated from several phytoplasmas (2, 10, 16). Since the mollicute membrane proteins probably play important roles in the attachment of the bacteria to their host cell surface, the immunodominant membrane protein is a candidate for involvement in host-phytoplasma interactions.

A comparison of synonymous and nonsynonymous nucleotide substitution rates is an effective approach for studying the mechanisms of DNA sequence evolution (21). The numbers of nonsynonymous and synonymous substitutions per site (d_N and d_S , respectively) (18) and their ratio ($\omega = d_N/d_S$) are important indicators of the selective pressure at the protein level, such that ω values of <1, 1, and >1 imply stabilizing selection,

neutral mutations, and diversifying positive selection, respectively. Many examples of positive selection at the molecular sequence level have been reported previously (1, 3, 5–7, 19, 27, 29). Most of these proteins play very important roles in their respective organisms and probably influence their fitness directly. Positive selection on a protein may indicate that it has an important role in the evolution of an organism (4). Analyzing the positively selected proteins from an organism should contribute to our understanding of the evolution of these organisms and the proteins.

To examine the positive selection acting on the phytoplasma antigenic membrane protein (Amp), we used 14 phytoplasmas: three strains of onion yellows phytoplasmas (OY-W, OY-M, and OY-NIM) (15, 23) and, porcelain vine witches’ broom phytoplasma (PvWB) (8) were previously described; plant tissues infected with mulberry dwarf phytoplasma (MD), eggplant dwarf phytoplasma (ED), tomato yellows phytoplasma (TY), lettuce yellows phytoplasma (LeY), Iceland poppy yellows phytoplasma (IPY), marguerite yellows phytoplasma (MarY), sumac witches’ broom phytoplasma (SWB), and paulownia witches’ broom phytoplasma (PaWB) were kindly provided by T. Shiomi, T. Usugi, N. Nishimura, and T. Tsuchizaki (20); DNA samples extracted from Bermuda grass white leaf phytoplasma (AYBG) and potato purple top phytoplasma (PPT) were kindly provided by A. Bertaccini (24). Total DNA was extracted as described previously (17) and used for PCR templates. The primers ES-1 (5′-TTG AGC TCG CGG CCG CAT GAA ACA AAA AAC AAT TAT C-3′) and Nad-2 (5′-TTT GCC TGA TTG AAA CCA TCT-3′) were used to amplify a 3.6-kb DNA fragment harboring the *amp* gene. The PCR conditions were as described previously (9), except for

FIG. 1. Physical map of the 3.6-kb DNA fragment from OY-W. It contains three complete genes encoding molecular chaperonin GroES (*groES*), GroEL (*groEL*), and antigenic membrane protein (*amp*) and an incomplete gene encoding NAD synthase (*nadE*).

* Corresponding author. Mailing address: Laboratory of Plant Pathology, Division of Agricultural and Life Sciences, The University of Tokyo, 1-1-1 Yayoi, Bunkyo-ku, Tokyo 113-8657, Japan. Phone: 81-3-5841-5053. Fax: 81-3-5841-5090. E-mail: snamba@ims.u-tokyo.ac.jp.

† Supplemental material for this article may be found at <http://jb.asm.org/>.

FIG. 2. A maximum-parsimony tree for the 3.6-kb nucleotide sequence containing *groES*, *groEL*, *amp*, and partial *nadE* genes. Numbers on the branches are bootstrap values obtained for 100 replicates (only values above 70% are shown).

the use of LA-*Taq* (Takara). The PCR-amplified fragments were purified using ExoSAP-IT (Amersham Bioscience) and then sequenced directly by using 11 primers: ES-1, ES-2 (5'-CAA TAG CAG GTG AAT CAT AGG C-3'), EL-1

(5'-AGA ATT CCA TAT GAG TAA AAA AAT ACT TTA TGG CAA-3'), EL-2 (5'-AAA TTA GTA CTG TGC AAG AAA-3'), EL-3 (5'-TAT CAA GAA TTA AAA GAT ACT-3'), Amp-N1 (5'-AAG AAT TCC ATA TGC AAA ATC AAA AAA CTC A-3'), Amp-3 (5'-CTG TTA AAG CTG TAG ATG GTA-3'), Amp-C1 (5'-AAG AGC TCG AGT TAT TTA TTG TTT TTG TTT TTT TTA AC-3'), Nad-1 (5'-TAC TCG AGG CGG CCG CCA TTT TGC ATA CTA TAA GCA T-3'), Nad-3 (5'-AAT AAT GAA TAA CTA AGG AAG C-3'), and Nad-2. Each 3.6-kb DNA fragment of each of the 14 phytoplasmas contained four genes in the order *groES*, *groEL*, *amp*, and *nadE* (Fig. 1). A previously reported 1.2-kb DNA fragment containing the *amp* gene of clover phyllody phytoplasma (GenBank accession number AF244541) (2) was completely identical to that from PPT.

For tree constructions, the 3.6-kb DNA fragments containing the *amp* genes from 14 phytoplasmas plus a 2.8-kb fragment containing the *amp* gene from aster yellows phytoplasma (AY) (GenBank accession number AF244540) (2) were aligned using ClustalW (28) and then adjusted manually. To construct trees, the neighbor-joining and maximum-parsimony (MP) methods in MEGA version 2.1 (12) were used. The MP tree is shown in Fig. 2.

To examine the positive selection, the nucleotide sequences

TABLE 1. Number of synonymous differences, number of nonsynonymous differences, and d_N/d_S (ω) of *amp* estimated by the Nei-Gojobori method^a

Parameter	Value for indicated phytoplasma							
	OY-W	OY-M	ED	LeY	AY	PaWB	PvWB	MD
OY-M								
d_N/d_S	2.124							
No. of <i>N</i>	7.00							
No. of <i>S</i>	1.00							
ED								
d_N/d_S	—	2.426						
No. of <i>N</i>	3.00	8.00						
No. of <i>S</i>	0.00	1.00						
LeY								
d_N/d_S	—	3.029	—					
No. of <i>N</i>	3.00	10.00	6.00					
No. of <i>S</i>	0.00	1.00	0.00					
AY								
d_N/d_S	2.732	—	3.336	3.636*				
No. of <i>N</i>	9.00	6.00	11.00	12.00				
No. of <i>S</i>	1.00	0.00	1.00	1.00				
PaWB								
d_N/d_S	3.068*	5.839**	3.216*	3.524**	6.140**			
No. of <i>N</i>	20.00	19.00	21.00	23.00	20.00			
No. of <i>S</i>	2.00	1.00	2.00	2.00	1.00			
PvWB								
d_N/d_S	3.107***	4.037***	3.169***	3.161***	3.937***	3.276***		
No. of <i>N</i>	47.25	48.25	48.25	48.25	47.25	54.75		
No. of <i>S</i>	4.75	3.75	4.75	4.75	3.75	5.25		
MD								
d_N/d_S	2.897***	3.446***	2.947***	2.880***	3.366***	3.168***	7.607***	
No. of <i>N</i>	53.25	52.25	54.25	53.25	51.25	62.75	25.00	
No. of <i>S</i>	5.75	4.75	5.75	5.75	4.75	6.25	1.00	
SWB								
d_N/d_S	2.510**	2.921***	2.550**	2.590***	2.546**	2.447**	2.498**	2.913***
No. of <i>N</i>	57.75	57.75	58.75	59.75	58.75	60.25	44.50	56.00
No. of <i>S</i>	7.25	6.25	7.25	7.25	7.25	7.75	5.50	6.00

^a The difference between d_S and d_N is significant at the 5% level (*), at the 1% level (**), and at the 0.1% level (***). Dashes indicate uncalculated pairs of d_N/d_S (ω) because of a d_S of 0. *N*, nonsynonymous differences; *S*, synonymous differences.

TABLE 2. Log-likelihood scores and parameters estimated for the *amp* gene from phytoplasmas

Model	l^a	Estimate(s) of parameters	Likelihood ratio test ($P < 0.001$)
M0 (one-ratio)	-1,618.885	$\omega = 2.675$	M0 vs M2
M1 (neutral)	-1,606.634	$p_0 = 0.498, p_1 = 0.502$	M1 vs M2
M2 (selection)	-1,567.415	$p_0 = 0.109, p_1 = 0.720, p_2 = 0.171, \omega_2 = 19.873$	M0 vs M3
M3 (discrete)	-1,563.095	$p_0 = 0.788, p_1 = 0.191, p_2 = 0.021, \omega_0 = 0.979, \omega_1 = 18.209, \omega_2 = 120.222$	M1 vs M3
M5 (gamma)	-1,566.361	$\alpha = 0.200, \beta = 0.030$	M0 vs M5
M7 (beta)	-1,606.635	$p = 0.003, q = 0.004$	M1 vs M5
M8 (beta and ω)	-1,567.508	$p_0 = 0.827, p_1 = 0.173, p = 1.733, q = 0.001, \omega = 22.656$	M7 vs M8

^a l , log-likelihood score.

of the *groES*, *groEL*, *amp*, and *nadE* genes from 14 phytoplasmas, as well as the *groEL*, *amp*, and *nadE* genes from AY, were each aligned (supplemental material). The *amp* sequence from PPT was excluded because this sequence contains many deletion sites, and it was not possible to determine the substitutions at each codon. ω was calculated as an average over all of the codon sites in each gene using the Nei-Gojobori method by MEGA version 2.1 (12) with Jukes-Cantor model or a maximum likelihood (ML) pairwise comparison (model M-2 in CODEML of PAML) (30). The significant difference test of ω was done according to a previously described procedure (14). These results are partly shown in Table 1. For 70 of the 91 pairs, ω was >1 , ranging from 1.819 (OY-M versus MerY) to 7.607 (PvWB versus MD). The values for the other 21 pairs could not be calculated because the values for d_s were 0. The ω values for 61 pairs were significantly greater than 1. Very similar results were obtained with the ML pairwise comparison (data not shown). For the other three genes (*groES*, *groEL*, and *nadE*), ω was <1 for all pairs (data not shown).

To confirm the results obtained by the Nei-Gojobori method and to identify the positively selected sites, the ML analysis was performed (30). In the ML approach, ω is assumed to follow a certain probability distribution among codon sites in the sequence. All ML analyses were performed by using the MP tree,

using the CODEML program in the PAML package version 3.13 (30). Table 2 shows the likelihood scores, the parameter estimates, and the results of likelihood ratio test calculated with the MP tree. Models M2, M3, M5, and M8 estimated that ω was $\gg 1$ and were significantly favored over model M0, M1, or M7 (P was <0.001 in all pairs), which is indicative of strong positive selection. The posterior probabilities of site classes along the *amp* gene are presented in Fig. 3 for the best-supported M3 model with the MP tree. One site, 83A, appears to be subjected to extraordinarily strong positive selection, indicated by ω values of 120.028. Similar ML analyses of the selection pressures were also performed using other tree topologies (e.g., the neighbor-joining tree, trees constructed only by *groEL* or *amp* sequences, or a "star" tree), and almost the same or more extreme results were obtained (data not shown). These analyses were repeated for the *groES*, *groEL*, and *nadE* genes. In all three genes and in all tree topologies, no models incorporating positive selection were significantly supported over models that did not (data not shown).

Combined, ω values estimated using both the Nei-Gojobori method (average ω over all codon sites) and the ML analysis (heterogeneous ω) provided strong evidence for positive selection acting on Amp, whereas there was no significant evidence of positive selection on the genes upstream or downstream of

FIG. 3. Posterior probabilities of site classes along the *amp* gene. Model M3 (discrete) with the maximum-parsimony tree topology was used; this assumes three classes of sites in the gene. The estimated frequencies and ω ratios for the three classes were as follows: p_0 , 0.788; p_1 , 0.191; p_2 , 0.021; ω_0 , 0.979; ω_1 , 18.209; ω_2 , 120.222 (Table 2). In each site, the total of posterior probabilities of the three classes is 1.000; e.g., the first codon site has posterior probabilities of 0.965, 0.035, and 0.000 with class 1 (white), 2 (gray), and 3 (black), respectively. The horizontal dashed line indicates a posterior probability of 0.95. Segments corresponding to two transmembrane regions (TM) and a coiled-coil structure (CC) are shown. The positions corresponding to sites with posterior probabilities of >0.95 are indicated by arrowheads.

the *amp* gene. The criterion that ω is >1 in the pairwise method is a very stringent one for detecting positive selection (25), so few genes were reported as having an ω of >1 by the Nei-Gojobori method (6, 27). Here, Amp shows high ω values in both methods, and Amp appears to have one of the highest ω ratios of previously reported proteins (1, 3, 7, 19, 29). The high ω values probably result from determining the ω ratio over a short evolutionary time, as closely related and biologically distinct phytoplasmas were compared. As shown in Table 1, it was observed that the actual number of nonsynonymous substitutions was greater than that of synonymous substitutions in most pairs (e.g., 62.75 and 6.25 in PaWB and MD, respectively). These data indicate that many nonsynonymous substitutions were accumulated within the time of accumulating few synonymous substitutions.

The Amp protein possesses a central hydrophilic region (amino acid positions 33 to 203) that is probably located on the phytoplasma exterior and two hydrophobic transmembrane regions (2, 10). In this study, the best-supported M3 model estimated 19 positively selected sites, and 18 of these 19 sites were located within the central hydrophilic region (Fig. 2). These results were consistent with the previous reports that the central region of *amp* is variable (2, 10), suggesting that there is strong selective pressure on the central region of Amp. It is likely that the positive selection on Amp is caused by the interaction between the phytoplasma and its extracellular environment, the host cytoplasm. However, it remains unclear whether the positive pressure was derived from either insects or plants. The detailed analysis of the source of positive pressure would be a further problem.

This study strongly suggests that the Amp protein plays a biologically important role in the phytoplasma-host interaction. Molecular evolutionary analysis might enable the identification of biologically important genes through the comparison of nucleotide sequences. Identification of positively selected proteins offers a good approach for understanding organisms, such as phytoplasmas, in which transformation or production of mutants is difficult. A search for other positively selected proteins of phytoplasmas would also be useful. Moreover, a search for the host proteins that interact with Amp should help to elucidate its role in host-phytoplasma interactions and the source of the positive pressure. These analyses will further our knowledge of the adaptation of phytoplasmas, contributing to our understanding of their biology.

Nucleotide sequence accession numbers. The sequences of the phytoplasmas have been deposited in GenBank with accession numbers AB124806 through AB124811, AB242231 through AB242237, and AB167357.

We thank T. Shiomi, T. Usugi, N. Nishimura, T. Tsuchizaki, and A. Bertaccini for the gifts of plant tissues and DNA samples.

This work was supported by grants-in-aid from the Ministry of Education, Culture, Sports, Science, and Technology of Japan (09460155).

REFERENCES

- Andrews, T. D., and T. Gojobori. 2004. Strong positive selection and recombination drive the antigenic variation of the PilE protein of the human pathogen *Neisseria meningitidis*. *Genetics* **166**:25–32.
- Barbara, D. J., A. Morton, M. F. Clark, and D. L. Davies. 2002. Immunodominant membrane proteins from two phytoplasmas in the aster yellows clade (chlorate aster yellows and clover phyllody) are highly divergent in the major hydrophilic region. *Microbiology* **148**:157–167.
- 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.
- Graur, D., and W.-H. Li. 2000. Rates and patterns of nucleotide substitution, p. 119–123. In A. D. Sinauer (ed.), *Fundamentals of molecular evolution*. Sinauer Associates, Sunderland, Mass.
- Hughes, A. L. 1991. Circumsporozoite protein genes of malaria parasites (*Plasmodium* spp.): evidence for positive selection on immunogenic regions. *Genetics* **127**:345–353.
- Hughes, A. L., and M. Nei. 1988. Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. *Nature* **335**:167–170.
- Jiggins, F. M., G. D. Hurst, and Z. Yang. 2002. Host-symbiont conflicts: positive selection on an outer membrane protein of parasitic but not mutualistic *Rickettsiaceae*. *Mol. Biol. Evol.* **19**:1341–1349.
- Jung, H.-Y., M.-C. Yae, J.-T. Lee, T. Hibi, and S. Namba. 2003. Aster yellows subgroup (*Candidatus* Phytoplasma sp. AY 16S-group, AY-sg) phytoplasma associated with porcelain vine showing witches' broom symptoms in South Korea. *J. Gen. Plant Pathol.* **69**:208–209.
- Jung, H. Y., T. Sawayanagi, S. Kakizawa, H. Nishigawa, S. Miyata, K. Oshima, M. Ugaki, J. T. Lee, T. Hibi, and S. Namba. 2002. '*Candidatus* Phytoplasma castaneae', a novel phytoplasma taxon associated with chestnut witches' broom disease. *Int. J. Syst. Evol. Microbiol.* **52**:1543–1549.
- Kakizawa, S., K. Oshima, H. Nishigawa, H. Y. Jung, W. Wei, S. Suzuki, M. Tanaka, S. Miyata, M. Ugaki, and S. Namba. 2004. Secretion of immunodominant membrane protein from onion yellows phytoplasma through the Sec protein-translocation system in *Escherichia coli*. *Microbiology* **150**:135–142.
- Kirkpatrick, B. C. 1992. Mycoplasma-like organisms: plant and invertebrate pathogens, p. 4050–4067. In A. Balows, H. G. Truper, M. Dworkin, W. Harder, and K. H. Schleifer (ed.), *The prokaryotes*. Springer, New York, N.Y.
- Kumar, S., K. Tamura, and M. Nei. 1994. MEGA: Molecular Evolutionary Genetics Analysis software for microcomputers. *Comput. Appl. Biosci.* **10**:189–191.
- McCoy, R. E., A. Caudwell, C. J. Chang, T. A. Chen, L. N. Chikowski, M. T. Cousin, J. L. Dale, G. T. De Leeuw, D. A. Golino, K. J. Hackett, B. C. Kirkpatrick, R. Marwitz, H. Petzold, R. C. Sinha, M. Suguira, R. F. Whitcomb, I. L. Yang, B. M. Zhu, and E. Seemüller. 1989. *The mycoplasmas*, vol. V. Academic Press, San Diego, Calif.
- Messier, W., and C. B. Stewart. 1997. Episodic adaptive evolution of primate lysozymes. *Nature* **385**:151–154.
- Miyahara, K., M. Matsuzaki, K. Tanaka, and N. Sako. 1982. A new disease of onion caused by mycoplasma-like organism in Japan. *Ann. Phytopathol. Soc. Jpn.* **48**:551–554.
- Morton, A., D. L. Davies, C. L. Blomquist, and D. J. Barbara. 2003. Characterization of homologues of the apple proliferation immunodominant membrane protein gene from three related phytoplasmas. *Mol. Plant Pathol.* **4**:109–114.
- Namba, S., S. Kato, S. Iwanami, H. Oyaizu, H. Shiozawa, and T. Tsuchizaki. 1993. Detection and differentiation of plant-pathogenic mycoplasma-like organisms using polymerase chain reaction. *Phytopathology* **83**:786–791.
- Nei, M. 1987. *Molecular evolutionary genetics*, p. 73–76. Columbia University Press, New York, N.Y.
- 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.
- Nishimura, N., S. Nakajima, T. Sawayanagi, S. Namba, T. Shiomi, I. Matsuda, and T. Tsuchizaki. 1998. Transmission of *Cryptotaenia japonica* witches' broom and onion yellows phytoplasma by *Hishimonus sellatus* Uhler. *Ann. Phytopathol. Soc. Jpn.* **64**:474–477.
- Ohta, T. 1992. The nearly neutral theory of molecular evolution. *Annu. Rev. Ecol. Syst.* **23**:263–286.
- Oshima, K., S. Kakizawa, H. Nishigawa, H. Y. Jung, W. Wei, S. Suzuki, R. Arashida, D. Nakata, S. Miyata, M. Ugaki, and S. Namba. 2004. Reductive evolution suggested from the complete genome sequence of a plant-pathogenic phytoplasma. *Nat. Genet.* **36**:27–29.
- Oshima, K., T. Shiomi, T. Kuboyama, T. Sawayanagi, H. Nishigawa, S. Kakizawa, S. Miyata, M. Ugaki, and S. Namba. 2001. Isolation and characterization of derivative lines of the onion yellows phytoplasma that do not cause stunting or phloem hyperplasia. *Phytopathology* **91**:1024–1029.
- Schneider, B., U. Ahrens, B. C. Kirkpatrick, and E. Seemüller. 1993. Classification of plant-pathogenic mycoplasma-like organisms using restriction-site analysis of PCR-amplified 16S rDNA. *J. Gen. Microbiol.* **139**:519–527.
- Sharp, P. M. 1997. In search of molecular darwinism. *Nature* **385**:111–112.
- Shen, W. C., and C. P. Lin. 1993. Production of monoclonal-antibodies against a mycoplasma-like organism associated with sweet-potato witches' broom. *Phytopathology* **83**:671–675.

27. **Tanaka, T., and M. Nei.** 1989. Positive darwinian selection observed at the variable-region genes of immunoglobulins. *Mol. Biol. Evol.* **6**:447–459.
28. **Thompson, J. D., D. G. Higgins, and T. J. Gibson.** 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**:4673–4680.
29. **Urwin, R., E. C. Holmes, A. J. Fox, J. P. Derrick, and M. C. Maiden.** 2002. Phylogenetic evidence for frequent positive selection and recombination in the meningococcal surface antigen PorB. *Mol. Biol. Evol.* **19**:1686–1694.
30. **Yang, Z., R. Nielsen, N. Goldman, and A. M. Pedersen.** 2000. Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* **155**:431–449.