J Mol Evol (2002) 54:437–457
DOI: 10.1007/s00239-001-0032-y

JOURNAL OF MOLECULAR EVOLUTION

A Phylogenomic Study of the OCTase Genes in *Pseudomonas syringae* Pathovars: The Horizontal Transfer of the *argK–tox* Cluster and the Evolutionary History of OCTase Genes on Their Genomes

Hiroyuki Sawada,[1] Shigehiko Kanaya,[2] Masataka Tsuda,[3] Fumihiko Suzuki,[1] Kozi Azegami,[1] Naruya Saitou[4]

[1] National Institute for Agro-Environmental Sciences, Tsukuba, Ibaraki 305-8604, Japan
[2] Nara Institute of Science and Technology, Ikoma, Nara 630-0101, Japan
[3] Graduate School of Life Sciences, Tohoku University, Sendai, Miyagi 980-8577, Japan
[4] National Institute of Genetics, Mishima, Shizuoka 411-8540, Japan

Received: 22 May 2001 / Accepted: 26 September 2001

Abstract. Phytopathogenic *Pseudomonas syringae* is subdivided into about 50 pathovars due to their conspicuous differentiation with regard to pathogenicity. Based on the results of a phylogenetic analysis of four genes (*gyrB, rpoD, hrpL,* and *hrpS*), Sawada et al. (1999) showed that the ancestor of *P. syringae* had diverged into at least three monophyletic groups during its evolution. Physical maps of the genomes of representative strains of these three groups were constructed, which revealed that each strain had five *rrn* operons which existed on one circular genome. The fact that the structure and size of genomes vary greatly depending on the pathovar shows that *P. syringae* genomes are quite rich in plasticity and that they have undergone large-scale genomic rearrangements. Analyses of the codon usage and the GC content at the codon third position, in conjunction with phylogenomic analyses, showed that the gene cluster involved in phaseolotoxin synthesis (*argK–tox* cluster) expanded its distribution by conducting horizontal transfer onto the genomes of two *P. syringae* pathovars (pv. *actinidiae* and pv. *phaseolicola*) from bacterial species distantly related to *P. syringae* and that its acquisition was quite recent (i.e., after the ancestor of *P. syringae* diverged into the respective pathovars). Furthermore, the results of a detailed analysis of *argK* [an anabolic ornithine carbamoyltransferase (anabolic OCTase) gene], which is present within the *argK–tox* cluster, revealed the plausible process of generation of an unusual composition of the OCTase genes on the genomes of these two phaseolotoxin-producing pathovars: a catabolic OCTase gene (equivalent to the orthologue of *arcB* of *P. aeruginosa*) and an anabolic OCTase gene (*argF*), which must have been formed by gene duplication, have first been present on the genome of the ancestor of *P. syringae*; the catabolic OCTase gene has been deleted; the ancestor has diverged into the respective pathovars; the foreign-originated *argK–tox* cluster has horizontally transferred onto the genomes of pv. *actinidiae* and pv. *phaseolicola*; and hence two copies of only the anabolic OCTase genes (*argK* and *argF*) came to exist on the genomes of these two pathovars. Thus, the horizontal gene transfer and the genomic rearrangement were proven to have played an important role in the pathogenic differentiation and diversification of *P. syringae.*

Key words: *Pseudomonas syringae* — *argK–tox* cluster — Ornithine carbamoyltransferase — Genome plasticity — Physical map — Codon usage — Molecular phylogeny — Gene duplication — Paralogue — Horizontal gene transfer

Introduction

Pseudomonas syringae, which is a phytopathogenic bacterium, is known to be rich in diversity and has hitherto

Correspondence to: H. Sawada; *email:* sawada@affrc.go.jp

been subdivided into about 50 pathovars, mainly with regard to phytopathogenicity and host ranges (Bradbury 1986; Dye et al. 1980; Rudolph 1995; Young et al. 1992). It is essential to identify the composition of pathogenicity-related genes of *P. syringae,* as well as the process in which they have undergone variations and functional differentiation, to understand the nature of this conspicuous pathogenic differentiation. In addition, there are cases in which different types of pathovars (pv.) affect the same plant species and/or cause infected plants to exhibit similar disease symptoms, so it is suspected that the same pathogenicity-related genes exist in various pathovars as a result of interpathovar horizontal transfer (Hatziloukas and Panopoulos 1992; Sawada et al. 1997, 1999). Genomes of phytopathogenic bacteria can be considered "vehicles of genes" for the passage of pathogenicity-related genes by horizontal transfer. Accordingly, if the course of genome evolution is identified, the course of the divergence and horizontal transfer of pathogenicity-related genes may be clarified by comparing genomes and pathogenicity-related genes in view of the process of evolution.

Thus, we tried to clarify the nature of the pathogenic differentiation of *P. syringae,* mainly through phylogenetic analysis. A comprehensive phylogenetic analysis of four genes (*gyrB, rpoD, hrpL,* and *hrpS*) of 56 strains belonging to 19 pathovars showed that the ancestor of *P. syringae* (the common ancestor prior to the divergence into pathovars) had diverged into at least three monophyletic groups (Groups 1 to 3) (Sawada et al. 1999). In addition, the results suggested that the gene cluster involved in phaseolotoxin synthesis (*argK–tox* cluster) may have expanded its distribution onto the genomes of two pathovars (*P. syringae* pv. *actinidiae* and *P. syringae* pv. *phaseolicola*) by conducting horizontal transfer (Sawada et al. 1997, 1999). Pathovar *actinidiae* is known as a causative agent of bacterial canker of kiwifruits, and pv. *phaseolicola* as a causative agent of halo blight disease of beans, both of which cause a disease symptom (chlorosis) on the leaves of the host plants. Because phaseolotoxin is a virulence factor that causes chlorosis by inhibiting the enzymatic activity in host plants (Bender et al. 1999), we suspected that the *argK–tox* cluster may have been directly involved in the pathogenic differentiation by conducting horizontal transfer among pathovars.

After the whole-genome sequence of *Haemophilus influenzae* was determined in 1995, numerous genome sequencing projects were implemented for prokaryotes, and genome sequences of more than 50 prokaryotes have been reported so far (e.g., see http://www.tigr.org/tdb/). The progress of such genome sequencing projects gave rise to genomics, study to elucidate the general outlook of life by thoroughly clarifying the whole (genomes) in addition to the pieces (genes) (Galperin et al. 1999; Watanabe et al. 1997).

Has the *argK–tox* cluster existing on genomes actually experienced horizontal transfer? How has this affected the evolution and diversification of *P. syringae?* To answer these questions, it is essential to conduct the above-mentioned genomic analysis to supplement phylogenetic analysis. Combination of these two approaches can be called "phylogenomic" analysis (proposed by Eisen 1998). In this study, we analyzed the *argK–tox* cluster from a phylogenomic approach, which was analyzed from a phylogenetic approach in our previous study (Sawada et al. 1999). The procedure of the genomic analysis conducted in this study was as follows: the genomic structure of representative strains of the respective monophyletic groups (Groups 1 to 3) of *P. syringae* was identified by I-*Ceu*I fragment analysis; physical maps of these strains were constructed, onto which the *argK–tox* cluster was mapped; and then an interpathovar comparison was made.

If the *argK–tox* cluster has in fact experienced horizontal transfer, what is its initial origin, pv. *actinidiae,* pv. *phaseolicola,* or a bacterial species other than *P. syringae* (i.e., of foreign origin)? Rigorous analyses were thereby conducted using *argK* [a gene that encodes anabolic ornithine carbamoyltransferase (anabolic OCTase; EC 2.1.3.3)], which is present within the *argK–tox* cluster, as a clue: (1) a phylogenetic tree was constructed using the information on all the already-reported OCTase genes to clarify the evolutionary history of OCTase genes, the "course of *argK* evolution" within this evolutionary history was identified, and then whether or not this course agrees with the "course of genomic evolution of *P. syringae*" was examined; and (2) as the codon usage pattern and GC content at the codon third position are known as highly effective criteria for determining whether or not genes are of foreign origin (Kanaya et al. 1996a, b, 1997), they were used as indices to compare *argK* with other genes on the *P. syringae* genomes.

This particular anabolic OCTase, called "ArgK," is known to play an extremely important role in the survival and pathogenicity of *P. syringae* (Jahn et al. 1985; Rudolph 1990; Staskawicz et al. 1980; Templeton et al. 1986). The final analysis conducted in this study focused on the clarification of the relationship between the course of OCTase evolution and the evolution and diversification of *P. syringae* by reproducing the evolutionary history of OCTase genes unfolded on the *P. syringae* genomes.

The results of these analyses showed that genomic rearrangements occurred on a large scale in the process of the divergence of the ancestor of *P. syringae* into the respective pathovars. In addition, our hypothesis that the "*argK–tox* cluster is of foreign origin and has expanded its distribution by conducting horizontal transfer" was confirmed to be consistent and unequivocal throughout the phylogenomic analyses. Furthermore, the plausible process of generating an unusual OCTase gene compo-

Table 1. Strains of *Pseudomonas* species used in this study

Species	Group[a]	Pathovar	Strain designation[b]	Geographic origin[c]	Source[b]
P. aeruginosa			PAO 1	?	J. Kato
P. fluorescens			MAFF 302208^{T}	UK	MAFF
P. putida			IAM 1236	USA	IAM
P. syringae	1	*actinidiae*	KW-11^{PT}	Shizuoka, Japan	Y. Takikawa
		maculicola	MAFF 302264^{PT}	New Zealand	MAFF
		theae	PT1^{PT}	Shizuoka, Japan	Y. Takikawa
		morsprunorum	MAFF 302280^{PT}	?	MAFF
		lachrymans	MAFF 302278^{PT}	USA	MAFF
	2	*syringae*	MAFF 302155^{T}	UK	MAFF
		pisi	MAFF 302269^{PT}	New Zealand	MAFF
		aptata	MAFF 302253^{PT}	USA	MAFF
		japonica	MAFF 301072^{PT}	Tochigi, Japan	MAFF
		aceris	MAFF 302273^{PT}	USA	MAFF
	3	*phaseolicola*	MAFF 302282^{PT}	Canada	MAFF
		phaseolicola	KUZ 1	Ibaraki, Japan	M. Sato
		glycinea	MAFF 302260^{PT}	New Zealand	MAFF
		tabaci	MAFF 302270^{PT}	Hungary	MAFF
		mori	MAFF 302279^{PT}	Hungary	MAFF

[a] The ancestor of *P. syringae* (the common ancestor prior to the divergence into pathovars) had so far diverged into at least three monophyletic groups (Groups 1 to 3) in the process of evolution (Sawada et al. 1999).
[b] IAM—IAM Culture Collection, Institute of Molecular and Cellular Biosciences, University of Tokyo, Tokyo 113-0032, Japan; MAFF—Microorganisms Section of the MAFF Gene Bank, National Institute of Agrobiological Resources, Tsukuba, Ibaraki 305-8602, Japan; J. Kato—Junichi Kato, Faculty of Engineering, Hiroshima University, Hiroshima, Japan; M. Sato—Mamoru Sato, National Institute of Agrobiological Resources, Japan; Y. Takikawa—Yuichi Takikawa, Faculty of Agriculture, Shizuoka University, Shizuoka, Japan. Superscripts: PT, pathotype strain; T, type strain.
[c] ?: Information was not available.

sition was revealed: in the initial stage, both the catabolic OCTase gene and the anabolic OCTase gene (*argF*) existed on the genome of the ancestor of *P. syringae*; the catabolic OCTase gene was deleted; the ancestor diverged into the respective pathovars; the *argK–tox* cluster horizontally transferred onto the genomes of pv. *actinidiae* and pv. *phaseolicola*; and hence two copies of only the anabolic OCTase genes (*argK* and *argF*) came to exist on the genomes of these two pathovars.

Incidentally, the research results reported in this paper have been presented orally (Sawada et al. 1998, 2000a,b, 2001).

Materials and Methods

Bacterial Strains

Fifteen strains belonging to 14 *P. syringae* pathovars, including pathotype strains, were used (Table 1). The sources of these strains are shown in Table 1, and detailed descriptions of their characteristics are given in our previous papers (Sawada et al. 1997, 1999). Reference strains of closely related species, which all belong to *Pseudomonas* (*sensu stricto*), were also used for comparison with *P. syringae* pathovars.

Preparation of High Molecular Weight Genomic DNA

Bacterial cells in the logarithmic phase of growth were suspended in a 10 m*M* Tris–HCl (pH 7.6) buffer containing 1 *M* NaCl, adjusted to 10^9 CFU/mL. Agarose plugs were prepared by mixing equal volumes of cell suspension and molten 2% low-melting point agarose. Bacterial cells embedded in agarose were lysed in the presence of RNase, treated with proteinase K, and stored in 0.5 *M* EDTA (pH 9.0) at 4°C. The agarose plugs containing intact genomic DNA were treated with phenylmethylsulfonyl fluoride (final concentration, 1 m*M*) to inactivate proteinase K and washed four times with at least 10 vol of TE [10 m*M* Tris–HCl, 1 m*M* EDTA (pH 8.0)] to remove the phenylmethylsulfonyl fluoride.

Restriction Digestion and Pulsed-Field Gel Electrophoresis (PFGE)

It is known that I-*Ceu*I is an endonuclease encoded by a group I intron in the LSU rRNA gene of *Chlamydomonas eugametos,* whose cleavage site exists only within the *rrl* gene on the genomes of many bacterial species (Liu et al. 1993). In addition, the number and position of *rrl* are strongly conserved among closely related species. Thus, the homologous I-*Ceu*I fragments on different genomes can easily be identified by combining I-*Ceu*I fragment size comparison using PFGE with Southern blot analysis. This I-*Ceu*I fragment analysis is also advantageous in that only genomic DNA can be analyzed without being affected by resident plasmids, because no I-*Ceu*I site exists on a plasmid and because, when conducting PFGE, closed circular DNA does not permeate the gels but is trapped in the well (Römling et al. 1996). Thus, I-*Ceu*I fragment analysis has been employed for analyzing genomes of various bacterial species (Liu et al. 1993, 1999).

For I-*Ceu*I digestion, agarose plugs were preincubated in a 2× concentrated I-*Ceu*I buffer supplied by the manufacturer for at least 1 h at 4°C and then incubated with 100 μL of a fresh 1× buffer containing 10 μg of bovine serum albumin (BSA) and 4 U of I-*Ceu*I (New England Biolabs) at 37°C for 4 h to digest DNA in plugs. For partial digestion with I-*Ceu*I, DNAs in agarose plugs were digested with 1 U of enzyme in 100 μL of a reaction mixture at 37°C, with some variations in the time of digestion. For digestion with all other enzymes,

Table 2. Genetic markers (probe[a]) used for Southern blot analysis and construction of physical maps, and their primers[b]

Designation of probe[a]	Target gene	Designation of primer[b]	Sequence of primer (5′ to 3′)	Position in submitted sequence	Accession No.	Amplicon size (bp)
16S[c]	*rrs* (16S rRNA)	16BF	GCTAACTCTGTGCCAGCAGCCGCGG	501–525	AB001439	594
		16BR-B	CGGGACTTAACCCAACATCTCACG	1094–1071	AB001439	
23S-1[c]	*rrl* (23S rRNA)	23S-F	AAGCTTGCTGGAGGTATCAGAAGTG	1216–1240	Y00432	354
		23S-R	AAGCTTTTCTTGGAAGCATGGCATC	1569–1545	Y00432	
23S-2[d]	*rrl* (23S rRNA)	23S-F2	CGGTGTTTGGCACCTCGATGTCGGC	2476–2500	Y00432	412
		23S-R2	AAGCCTCACGGGCAATTAGTACTGG	2887–2863	Y00432	
rpoD	*rpoD*	rpo F	AAGGCGAGATCGAAATCGCCAAGCG	347–371	U93292	582
		rpo R-B	GGAACATGCGCAGGAAGTCGGCTCG	928–904	U93292	
gyrB	*gyrB*	gyr AF3	CCGGCGGTAAGTTCGATGATAACTC	422–446	X54631	669
		gyr AR-B3	TGATCGCGGTCAGGCCTTCACGGGC	1090–1066	X54631	
hrpL	*hrpL*	hrpLF3	TTTTGGCTGGCACGGTTATCGCTATA	1025–1050	U16817	653
		hrpLR-B	TGTGGTTTTGCGTGCGAGTTGGTTCC	1677–1652	U16817	
hrpS	*hrpS*	hrpS-F2	CTCCAGGCCAAGCTGCTGAGGGTGC	363–387	U03852	341
		hrpS-RB	TTGAGCTCGCGGATATTGCCGGGCC	703–679	U03852	
argK	*argK*	OCTF-03	GACCGTCAAGGAAGAATTCGGGCGC	835–859	M94049	632
		OCTR	CGACCTTGTTGACCTCCCG	1466–1448	M94049	
orf3	ORF3[e]	orf3-F	TGTTGAAGATGTCGGGGTTGTCCGC	2298–2322	U27310	547
		orf3-R	TCCACAGGGCGTATCCCGCCCTCGC	2844–2820	U27310	
argF	*argF*	argF1F	CTGCGGTCTGTTGCAAAAAAGGTAG	270–294	M99382	534
		argF1R-B	ATATAGCTATTGCACATGTTGTTGC	803–779	M99382	
arcA	*arcA*	arcAF	TCCACTCCGAAGCCGGCAAACTGCG	56–80	X14694	838
		arcAR	TCGCGGTCGCAGAAGCTGAACACGG	893–869	X14694	
arcB	*arcB*	arc 1 F 2	TTCCCACAAGGAGATCAATCATGGC	44–68	X05637	591
		arc 1 R-B	CCTTCGGTGCGGCGATGCGTACGTC	634–610	X05637	
arcD	*arcD*	arcDF	TCCAGCCAAAAACTCCGACTAGGAG	36–60	M33223	530
		arcDR	ATGTCCAGCTTGAAGGCGAACAGGC	565–541	M33223	

[a] The products amplified by PCR were cloned, the sequence of the inserts was confirmed, and then the inserts separated by electrophoresis were excised from the gels and DIG-labeled for use as probes. The conditions of hybridization and detection using the heterologous gene probes in the Southern blot analyses (Figs. 6 and 7) are described under Materials and Methods and in the legend to Fig. 6.
[b] All the primers listed in this table were designed in this study.
[c] A probe that hybridizes with the upstream region of the 1-*Ceu*I site.
[d] A probe that hybridizes with the downstream region of the 1-*Ceu*I site.
[e] ORF3 is located in the *phtE* locus of the *argK–tox* cluster and is similar to the acetylornithine aminotransferase gene from *E. coli* (Zhang and Patil 1997).

agarose plugs were immersed in a 1× buffer supplied by the manufacturer and incubated at 4°C for at least 1 h. The buffer was then replaced with 100 μL of a fresh 1× buffer containing 10 μg of BSA and 10 U of enzyme and incubated for 4–16 h. The enzyme-digested DNA fragments were separated by a CHEF-DR II apparatus (Bio-Rad) according to the manufacturer's instructions. After PFGE, the gels were stained with SYBR Green I (FMC BioProducts). Phage λ multimeric DNA and yeast chromosomal DNA were used as molecular mass markers. For redigestion, restriction fragments separated by PFGE were excised from agarose gels under a long-wavelength UV light, and agarose plugs containing DNA fragments were trimmed to slices about 1 mm thick and digested with a second enzyme.

Preparation of DNA Probes

For Southern hybridization analysis and construction of physical maps, 13 genes were selected as genetic markers (Table 2). PCR was conducted using the primer sets listed in Table 2, and the genes selected as markers were amplified. Preparation of templates and amplification by PCR were performed based on the method described by Sawada et al. (1995, 1997). After the amplified products were cloned to pUC18, the sequences of cloned inserts were determined, which showed that the target amplified products were successfully cloned. The inserts separated by electophoresis were then excised from the gels, labeled with DIG-High Prime (Roche Diagnostics), and used as probes.

Southern Hybridization Analysis

DNA fragments separated by electrophoresis were transferred to a nylon membrane (Hybond-N+; Amersham Pharmacia Biotech). Southern hybridization and detection of the hybridized DNA were carried out using the DIG DNA Labeling and Detection Kit (Roche Diagnostics) following the manufacturer's instructions. In the Southern hybridization analysis conducted for detecting OCTase genes, the following conditions were established, particularly when heterologous gene probes were used: hybridization buffer, DIG Easy Hyb (Roche Diagnostics); concentration of the DIG-labeled probe in the hybridization solution, 25 ng/mL; temperature for hybridization, 40°C; temperature for washing in 0.5× wash solution, 60°C; and chemiluminescent substrate, CDP-Star (Tropix).

Sequence Alignments

Sequence data were compiled using DNASIS-Mac Version 3.6 (Hitachi Software Engineering) and MacClade Version 3.06 (Maddison and Maddison 1996), and sequence alignments were determined using CLUSTAL W version 1.7 (Thompson et al. 1994). DNA sequences for OCTase-encoding genes were converted into amino acid sequences, and multiple alignments among them were inferred. Some previously published alignment data (Hatziloukas and Panopoulos 1992; Hough-

ton et al. 1984; Itoh et al. 1988; Ruepp et al. 1995; Schurr et al. 1995; Ventura et al. 1997) were used as templates to correct the inferred alignments manually. Information on the functions of gene products was also used to aid in the multiple alignments. The final products of the multiple alignments are available in a machine-readable format. (For details, contact Hiroyuki Sawada by E-mail at sawada@affrc.go.jp.) Positions where gaps are present in any of the aligned sequences were excluded from the following analysis.

Phylogenetic Analysis

Phylogenetic trees were constructed using the maximum-likelihood (ML) method (Felsenstein 1981), according to the conditions described by Sawada et al. (1999). To obtain the topology of ML trees, local rearrangement searches were conducted using NucML and ProtML of MOLPHY version 2.3 (Adachi and Hasegawa 1996) with a neighbor-joining (NJ) tree topology and a topology obtained by the quick add OTUs search option as initial trees. The HKY85 model (Hasegawa et al. 1985) was used for the base substitution process, and the "F" option of the JTT model (Cao et al. 1994) for the amino acid substitution process. To evaluate the reliability of the inferred ML tree topology, the local bootstrap probability (LBP) was calculated using the RELL method (Hasegawa and Kishino 1994) with 1000 replications.

Analysis of Codon Usage and Adaptiveness of the argK–tox *Cluster to the* P. syringae *Genomes*

We characterized the latent structure of the species-specific diversity of genes in codon usage for the *P. syringae* genomes as follows. As we have described the methodology in one of our previous papers (Kanaya et al. 1996a), only an outline is given here.

The codon usage pattern for each gene of a species is described by a vector consisting of each codon frequency, i.e., a gene is represented as a point in the codon frequency space. To describe codon-usage diversity which was derived from factors such as cellular tRNA amounts and/or mutational pressure for evolutional processes, a scale with the largest variance in the codon frequency space was estimated on the basis of principal-component analysis.

Genes which are present on the *P. syringae* genomes were extracted from the DDBJ (DNA Data Bank of Japan). We further selected from the genes the 33 gene sequences longer than 500 bases in each coding region that satisfied the following two conditions: (1) genes in which the sum of codon occurrences for every amino acid except methionine and tryptophan was not zero were selected; (2) to avoid multiple entries of one gene, the longest sequence was selected when there were sequences with the same gene name but different coding lengths. In the case of genes with the same name and length but with variations in their sequences, one of them was selected arbitrarily. The selected genes were as follows: *argF, argK, avrA, avrB, dapB, hrmA, hrpC, hrpH, hrpJ, hrpL, hrpR, hrpS, hrpU2, hrpW, hrpX, hrpY, infC, lemA, mrsA, pgm, pilM, pilP, syrB, syrC, syrD, syrP, tabA, tabB, tblA, tpiA, ttr,* dihydropteroate synthase gene, and fatty acid desaturase gene (DES-ORF).

To eliminate the influence of gene size, amino acid composition, and codon box number, the codon-usage pattern of the *i*th gene was described as a vector $\mathbf{x_i}$ consisting of $x_{ij(m)}$ as expressed by Eq. (1).

$$x_{ij(m)} = f_{ij(m)} \Big/ \left[\sum_{j=1}^{M(m)} f_{ij(m)} \Big/ M(m) \right] \tag{1}$$

where $f_{ij(m)}$ denotes the *j*th synonymous codon number of the *m*th amino acid, and *M(m)* denotes its codon box number. For simplicity, variables for individual codon frequencies are denoted X_j. Principal-component analysis was conducted to estimate the direction of the gene distribution with the largest width in the codon-usage space. The *k*th principal component (PC) is represented by Eq. (2).

$$Z'_k = \sum_{j=1}^{M} b_{kj} X_j \tag{2}$$

$$\text{where } \sum_{j=1}^{M} b_{kj}^2 = 1.0.$$

The vector $\mathbf{b_k} = [b_{k1}, b_{k2}, \ldots, b_{kM}]$ is referred to as the b_k vector, which represents the direction of the species-specific diversity of codon usage in genes. The *k*th PC (i.e., Z'_k) was standardized by Eq. (3) and denoted Z_k.

$$Z_k = (Z'_k - \mathrm{Av}[Z'_k])/\mathrm{SD}[Z'_k] \tag{3}$$

where $\mathrm{Av}[Z'_k]$ and $\mathrm{SD}[Z'_k]$ are the average and standard deviation of the variable Z'_k for the data set. With this scaling, some statistical data of genes in the PC axes were obtained. For example, if the *i*th gene has a z_{ik} larger than 2, the codon usage becomes very biased among genes of the respective species because, in statistical theory, only 2.3% of genes are included in this interval.

Two parameters, i.e., the proportion in Eq. (4) and the factor loadings in Eq. (5), were used to interpret the PC axes. The former quantifies the width of distribution for the *k*th PC, and the latter represents the contribution of the *j*th variable to the *k*th PC.

$$\mathrm{Pr}[Z'_k] = \mathrm{Var}[Z'_k] \Big/ \sum_{j=1}^{M} \mathrm{Var}[X_j] \tag{4}$$

$$\mathrm{r}(Z'_k, X_j) = \mathrm{Cov}[Z'_k, X_j]/(\mathrm{Var}[Z'_k]\mathrm{Var}[X_j])^{1/2} \tag{5}$$

where Cov[A, B] and Var[A] denote the covariance between variable A and variable B and the variance of A, respectively.

Results and Discussion

Plasticity of the Genomes of P. syringae *Pathovars: Validation of the Horizontal Transfer of the* argK–tox *Cluster*

An attempt was made to clarify how the genomic structure of the three groups of *P. syringae* varied while they evolved separately after diverging from their ancestor. In addition, "whether or not the *argK–tox* cluster has actually experienced horizontal transfer" was determined by locating the *argK–tox* cluster on the genomes of pv. *actinidiae* and pv. *phaseolicola* and by comparing its position between these two pathovars.

*Construction of Physical Maps of Genomes by I-*Ceu*I Fragment Analysis.* I-*Ceu*I fragment analysis using PFGE and Southern blot analysis using the various probes listed in Table 2 were combined for constructing physical maps of the genomes of representative strains selected from the three monophyletic groups which had been identified by Sawada et al. (1999), i.e., pv. *actin-*

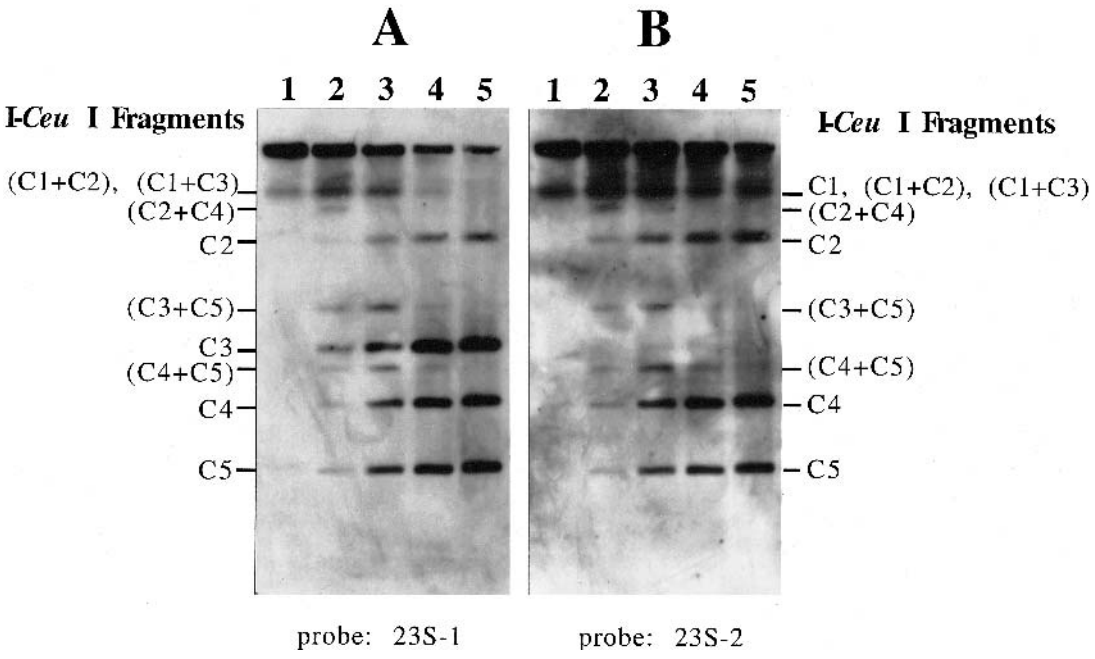

Fig. 1. Results of an I-*Ceu*I partial digestion combined with a Southern blot analysis using a probe that hybridizes with the upstream region of the I-*Ceu*I site (23S-1; **A**) and a probe that hybridizes with the downstream region of the I-*Ceu*I site (23S-2; **B**). First, the genomic DNA of *P. syringae* pv. *actinidiae* strain KW-11^{PT} (see Table 1) was partially digested using I-*Ceu*I, then the digests were separated by PFGE, Southern blotted, and probed (A, 23S-1; B, 23S-2; see Table 2). In both A and B, the degree of digestion by I-*Ceu*I of the electrophoresed samples gradually advances from Lane 1 to Lane 5, and the samples in Lane 5 are completely digested. Five I-*Ceu*I fragments generated by complete digestion are numbered in descending order of size (C1 to C5; see Fig. 2). The possible composition of the neighboring I-*Ceu*I fragments, which constitute each partially digested fragment, is indicated in parentheses. Fragments of ca. 1800 kb or smaller could be resolved satisfactorily by this particular PFGE run (CHEF DRII, 60 s ramping to 120 s, 200 V, 24 h, and 1.2% gel in 0.5 × TBE at 14°C), while fragments larger than 2000 kb were resolved by modifying these conditions.

idiae from Group 1, pv. *syringae* from Group 2, and pv. *phaseolicola* from Group 3 (Table 1).

PFGE was conducted for the above three strains on a preliminary basis, and the results showed that nothing permeated the gels when intact genomic DNA samples were used, but several fragments were separated when samples were used after being digested separately with either I-*Ceu*I, *Pac*I, or *Pme*I (data not shown). This shows that the genome of *P. syringae* is not linear and that *P. syringae* does not contain any linear plasmid. Subsequently, Southern blot analysis was conducted: each genomic DNA of the strains tested was first digested separately with either *Bam*HI, *Eco*RI, *Eco*RV, *Kpn*I, *Pst*I, or *Sal*I, which cuts the genomic DNA outside the *rrn* genes; then the digests were separated by PFGE, Southern blotted, and probed with either the 16S rRNA gene (16S, Table 2) or the 23S rRNA gene (23S-2). As a result, five equally intense bands were detected in all cases (data not shown), showing that five *rrn* operons exist on the *P. syringae* genome.

PFGE was then conducted for the I-*Ceu*I complete cleavage products, by which five fragments were separated. It was also revealed that the probe that hybridizes with the upstream region of the I-*Ceu*I site (16S or 23S-1; Table 2) and/or the probe that hybridizes with the downstream region of the I-*Ceu*I site (23S-2; Table 2) were infallibly hybridized to each of these five fragments (Fig. 1). Furthermore, based on I-*Ceu*I partial digestion analysis in conjunction with Southern hybridization using the respective probes shown in Table 2 (Fig. 1), these five I-*Ceu*I fragments were revealed to be linked to each other on the genome.

These data showed that I-*Ceu*I cleavage sites exist only within the *rrl* genes with respect to all three strains tested and that I-*Ceu*I mapping is an effective method for use in the genomic analysis of *P. syringae.* It is also known from these data that the genome of *P. syringae* is circular, and five *rrn* operons exist on it. Hereafter, the five I-*Ceu*I fragments of the genomes of the respective strains are referred to as C-1, C-2, C-3, C-4, and C-5, in descending order of size (Figs. 1 and 2).

The I-*Ceu*I partial digestion analysis in conjunction with the Southern hybridization was conducted in more detail (Fig. 1), and the order of the five I-*Ceu*I fragments was obtained as shown in Fig. 2. Incidentally, we had identified in our previous studies that the order of the respective rRNA genes in the *rrn* operons of *P. syringae* was 5′-*rrs*–*rrl*–*rrf*-3′ as in various bacterial species (Sawada et al. 1997, 1998, unpublished data). Therefore, in this study, we succeeded in determining the direction of transcription of each *rrn* operon (Fig. 2) by using a probe (16S or 23S-1) that hybridizes with the upstream

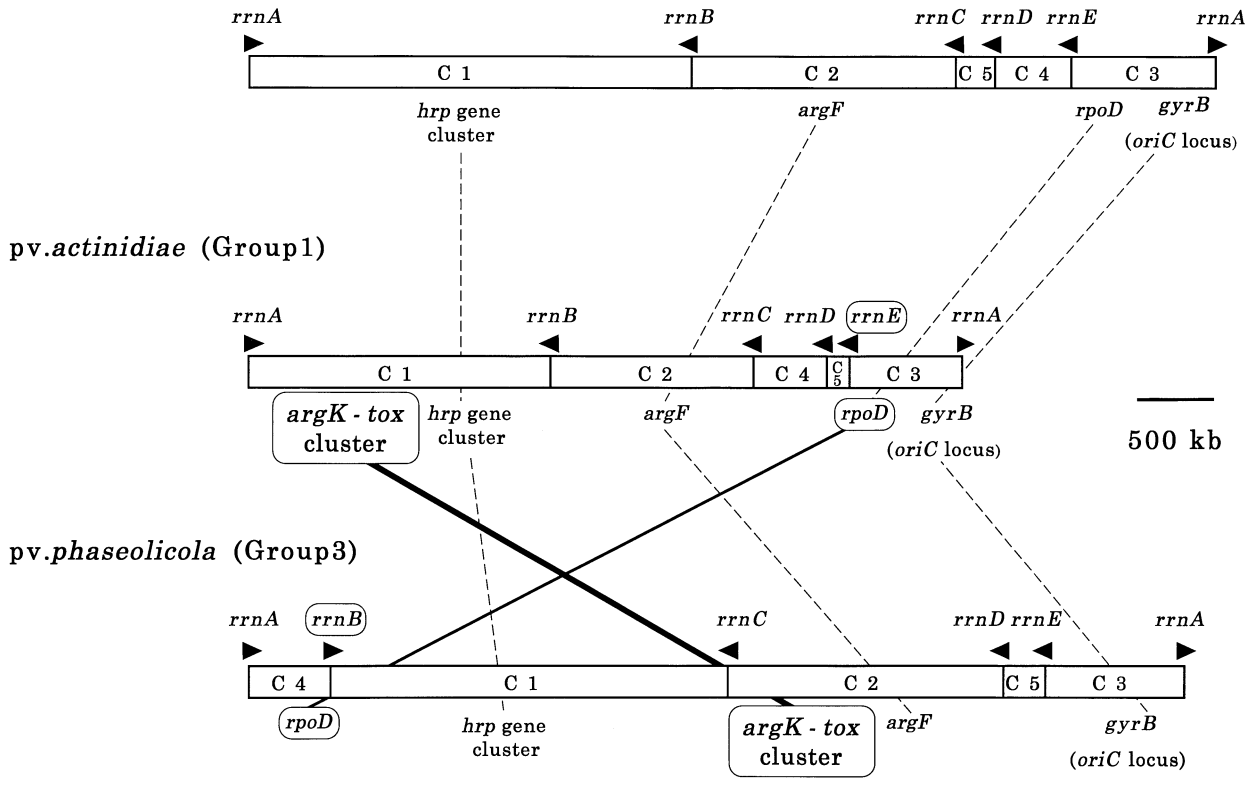

Fig. 2. Physical maps of the genomes of *P. syringae* pv. *syringae* strain MAFF 302155^{T} (Group 2), pv. *actinidiae* strain KW-11^{PT} (Group 1), and pv. *phaseolicola* strain MAFF 302282^{PT} (Group 3). These maps were constructed and the genetic markers mapped by combining the partial-digestion analyses using I-*Ceu*I and *Pme*I with the Southern blot analyses (see Fig. 1 and the text). These genomes are actually circular but are presented linearly here to facilitate comparisons. I-*Ceu*I fragments are numbered in descending order of size (C1 to C5), which are shown in the fragment bars. Each junction between the fragments is the endonuclease cleavage site of I-*Ceu*I. *Arrowheads* show the positions and directions of transcription of ribosomal RNA operons (*rrn* operons), and the labels above the arrowheads indicate the names of *rrn* operons designated by us. The zero point of the genome map was established in the *oriC* region, and the five *rrn* operons were named *rrnA* to *rrnE* with this zero point as the origin. The genetic markers are listed in Table 2. Each pair of orthologues is connected by either *dashed* or *solid lines:* the orthologues connected by dashed lines have positions that are conserved among the genomes, while those connected by solid lines have positions that differ greatly among the genomes. In the C1 fragment of pv. *actinidiae* and in the C2 fragment of pv. *phaseolicola,* the three genetic markers (*argK–tox* cluster, *hrp* gene cluster, and *argF*) are placed arbitrarily, so their positions in these two fragments do not necessarily agree with the actual order of their positions on the genomes.

region of the I-*Ceu*I site and a probe (23S-2) that hybridizes with the downstream region of the I-*Ceu*I site (Fig. 1).

An attempt was also made to identify more precisely the genomic structure by conducting PFGE analysis using I-*Ceu*I and *Pme*I. Specifically, genomic DNA samples were treated in four ways: (1) complete digestion using I-*Ceu*I and complete digestion using *Pme*I, (2) complete digestion using I-*Ceu*I and partial digestion using *Pme*I, (3) partial digestion using *Pme*I of the respective I-*Ceu*I fragments excised from the PFGE gels, and (4) partial digestion using *Pme*I. The digests of all these samples were then separated by PFGE, Southern blotted, and probed with the respective markers (Table 2) (data not shown). This helped us to map the marker genes roughly as shown in Fig. 2. Unfortunately, the arrangement of the *argK–tox* cluster and the *hrp* gene cluster in the C-1 fragment of the pv. *actinidiae* genome, as well as that of the *argK–tox* cluster and *argF* in the C-2 fragment of the pv. *phaseolicola* genome, could not be clearly determined. For the time being, these two pairs are arranged tentatively in their respective orders as shown in Fig. 2. More rigorous analysis will be conducted to determine the precise arrangement of the respective genes.

Structure of the P. syringae *Genome.* The results of the above analyses showed that all three pathovars tested had five *rrn* operons, all of which were present on one circular genome (Fig. 2). With regard to the genomes of bacteria, *oriC,* the origin of replication, is defined loosely as a locus harboring either the *dnaA* or the *gyrB* gene, which is linked to an *rrn* operon (Cole and Saint Girons 1994). With respect to various bacteria, it has been confirmed that *rrn* operons are generally arranged in two

groups transcribed divergently away from *oriC* (Cole and Saint Girons 1994; Honeycutt et al. 1993; Le Bourgeois et al. 2000). The genomic structure of *P. syringae* obtained in this study showed good agreement with this general tendency of bacteria. Regarding the genome of pv. *actinidiae,* for example, the C-3 fragment is assumed to contain *oriC,* and the five *rrn* operons are arranged transcribed toward two opposite directions, going far off from the C-3 fragment (Fig. 2).

Given that *oriC* is used as the reference point of bacterial genome maps in general (Römling et al. 1992), the zero point was established in the C-3 fragment of the pv. *actinidiae* genome in this study (Fig. 2). With this C-3 fragment as the origin, the five *rrn* operons were named *rrnA* to *rrnE* in the order shown in Fig. 2. Regarding the genomes of pv. *phaseolicola* and pv. *syringae,* the zero point was also established in the C-3 fragment, which is assumed to contain *oriC,* and the respective *rrn* operons were named as shown in Fig. 2 with the zero point as the origin.

Comparison of the Genomic Structure Among the Groups. A comparison of the three strains showed their genomes to be very different, not only in basic structure but also in size. Specifically, the genomes of pv. *syringae* and pv. *phaseolicola* were roughly the same size (about 6 Mb), while the genome of pv. *actinidiae,* at about 4.7 Mb, was markedly smaller (Fig. 2).

In addition, it was observed that *rrn* operons and *rpoD,* which are essential for survival, had undergone translocation, resulting in a variation of their positions on the genomes depending on the pathovar (Fig. 2). Specifically, in the case of the pv. *syringae* and pv. *actinidiae* genomes, the positions of *rrnE* and *rpoD* are relatively close (they coexist in the I-*Ceu*I–*Pme*I fragment, which is about 600 kb in size for pv. *syringae* and about 100 kb for pv. *actinidiae*) and are arranged in the same order in both of these pathovars. On the other hand, *rpoD* of pv. *phaseolicola* is present in the C-4 fragment, meaning that its position is opposite the position of *rpoD* in the other two pathovars, with the *oriC* region in the center.

Furthermore, the following results were obtained: (1) Southern hybridization analysis showed that only one copy of *rpoD* is present on the pv. *phaseolicola* genome; (2) a phylogenetic tree of *rpoD* showed that *rpoD* of pv. *phaseolicola* belongs to the lineage of Group 3 (Sawada et al. 1999); (3) the results of a *t* test conducted based on the distance of synonymous substitution showed that the base substitution pattern of *rpoD* of pv. *phaseolicola* is not significantly different from that of other essential genes on the same genome, such as *gyrB, hrpL,* and *hrpS* (Sawada et al. 1999, unpublished data); and (4) the results of PC analysis conducted based on the codon-usage pattern provided a Z_1 value of 0.388 for *rpoD* of pv. *phaseolicola,* indicating that the codon usage of *rpoD* of pv. *phaseolicola* agrees with that of other native genes existing on the *P. syringae* genome.

The results of the above four analyses indicate that *rpoD* of pv. *phaseolicola* is not of foreign origin but is a native gene which has resided on the pv. *phaseolicola* genome from the beginning. Thus, all these results led us to speculate that the reason the position of *rpoD* on the pv. *phaseolicola* genome is markedly different from that of the other two pathovars is because *rpoD* of pv. *phaseolicola* translocated from its initial position near *rrnE* up to the present position.

Meanwhile, *rrnB* in pv. *phaseolicola* was revealed to be present on the opposite side of that in the other two pathovars, with the *hrp* gene cluster in the center, and its direction of transcription was also opposite (Fig. 2). Another difference revealed was that *rrnB* and *rpoD* are close to each other in the case of pv. *phaseolicola* (150 kb or less), while in the other two pathovars, the closely located genes are *rrnE* and *rpoD.* Based on these results, *rrnB* of pv. *phaseolicola* appears to be an operon orthologous to *rrnE* of the other two pathovars and to have transferred to its present position by conducting translocation together with *rpoD.*

The above results indicate that the structure and size of the *P. syringae* genomes vary greatly depending on the pathovar, because the *P. syringae* genomes are quite rich in plasticity and underwent active and large-scale rearrangement, including insertion, deletion, and translocation. Similar plasticity has been reported for the genomes of various other bacteria, including *Escherichia coli* (Bergthorsson and Ochman 1998; Rode et al. 1999; Hayashi et al. 2001), *Bordetella pertussis* (Stibitz and Yang 1999), *Vibrio cholerae* (Khetawat et al. 1999), *Lactococcus lactis* (Le Bourgeois et al. 2000), and *Azospirillum* species (Martin-Didonet et al. 2000). This means in turn that there are diversified genome sizes and structures even within the same species. The genomes of *E. coli* and *Salmonella enterica* have a size variation of 1 Mb within the same species, respectively (Bergthorsson and Ochman 1998; Hayashi et al. 2001; Rode et al. 1999; Thong et al. 1997). As the difference of about 1.3 Mb in size observed between the genome of pv. *actinidiae* and the genomes of the other two pathovars is equivalent to the respective size variations of *E. coli* and *S. enterica,* the plasticity of the *P. syringae* genomes is concluded to be quite high.

Given that the genomic structure of pv. *phaseolicola* identified in this study is roughly the same as that reported by Esther De Ita et al. (1998), the genomic structure of pv. *phaseolicola* may be almost invariable. Nonetheless, the structural and size variations within one single pathovar of *P. syringae* need to be examined further using more strains with diverse origins, because there has been a report that the genome size of *S. enterica* serovar Typhimurium varies greatly even within this one serovar (Thong et al. 1997).

Phylogenomic Analysis of the argK–tox *Cluster: The* argK–tox *Cluster Has in Fact Experienced Horizontal Transfer.* The *argK–tox* cluster was mapped onto the physical maps constructed as described above, and then traces which indicate the horizontal transfer of the *argK–tox* cluster were searched for by phylogenomically reviewing the phylogenetic data obtained in our previous studies (Sawada et al. 1997, 1999), as well as the information on the genomic structure obtained in this study.

Except for pv. *actinidiae* and pv. *phaseolicola,* no pathovars of *P. syringae* have the *argK–tox* cluster on their genomes (Figs. 2 and 6) (Sawada et al. 1997, 1998, 1999). Using phylogenetic analysis, we showed in our previous paper that these two pathovars belong to two separate monophyletic groups (pv. *actinidiae* belongs to Group 1; pv. *phaseolicola,* Group 3) and that they are very distant from each other in a phylogenetic tree (Sawada et al. 1999). We also proved that the *argK* sequences of these two pathovars are completely identical, with not a single synonymous substitution observed (Sawada et al. 1997, 1999). We thus clarified that completely identical *argK* genes are present on the genomes of these two pathovars, although their genomic structures have become markedly different from each other (Fig. 2) as they evolved in two separate courses (Sawada et al. 1999).

In addition, the results of mapping the *argK–tox* cluster onto the physical maps of genomes showed that the position of the *argK–tox* cluster differed greatly between these two pathovars (Fig. 2); in the case of the pv. *actinidiae* genome, the *argK–tox* cluster is in a position close to the *hrp* gene cluster, while in the case of the pv. *phaseolicola* genome, the *argK–tox* cluster is present near *argF*. Thus, all the above results obtained by the phylogenetic and genomic analyses consistently and strongly support the hypothesis that the "*argK–tox* cluster has expanded its distribution onto the genomes of these two pathovars by conducting horizontal transfer after the ancestor of *P. syringae* diverged into the respective pathovars."

Comparison Between the OCTase Tree and the 16S rDNA Tree: Validation of the Foreign Origination of the argK–tox *Cluster (1)*

If the *argK–tox* cluster has in fact experienced horizontal transfer, where did it originally come from: pv. *actinidiae,* pv. *phaseolicola,* or a bacterial species other than *P. syringae* (i.e., of foreign origin)? To resolve this question, various analyses were conducted as discussed here and in the following sections. In this section, the results of a phylogenetic analysis of "*argK,*" which is a pathogenicity-related gene existing within the *argK–tox* cluster, are described. This *argK* is a gene that encodes anabolic ornithine carbamoyltransferase (anabolic OCTase; EC 2.1.3.3) (Hatziloukas and Panopoulos 1992; Mosqueda et al. 1990; Sawada et al. 1997) and is known to have a pathogenicity-related function (Rudolph 1990; Staskawicz et al. 1980; Templeton et al. 1986).

First, a phylogenetic tree was constructed using the currently available published information on OCTase gene sequences of a variety of organisms to understand the evolutionary history of OCTase genes, and then the course of evolution of *argK* within this evolutionary history was clarified. Subsequently, another phylogenetic tree was constructed using 16S rDNA as an index, which was used to confirm the course of the evolution of the *P. syringae* genome. Then these two courses of evolution were compared to find differences in properties which could suggest the foreign origination of *argK*.

Construction of an OCTase Tree: Elucidation of the Evolutionary History of OCTase Genes. OCTase is known to consist of two types of enzymes, i.e., anabolic and catabolic, which are involved in the biosynthesis and degradation of arginine, respectively (Cunin et al. 1986; Ramos et al. 1967). There are diverse homologues that encode these two types of OCTase, which are distributed widely in three domains, namely, *Bacteria, Archaea, and Eucarya.* In this study, such homologues were found to contain not only orthologues that have originated from speciation but also paralogues that have originated from gene duplication (Figs. 3 and 8, Table 6), which is explained later. Thus, to identify the entire aspect of the evolutionary history of OCTase genes, it is essential to conduct phylogenetic analyses precisely and discretely by sampling homologues (including both orthologues and paralogues) from a variety of organisms of the three domains.

The problem, however, is what to use as the outgroup when constructing a phylogenetic tree for clarifying the evolutionary history of OCTase genes of all three domains. In this study, we decided to use aspartate carbamoyltransferase (ACTase; EC 2.1.3.2), which is involved in pyrimidine biosynthesis, as the outgroup, and constructed a composite phylogenetic tree. As for *E. coli* genomes, it has been reported that *pyrB,* which encodes ACTase, is present near 96.4 min, being adjacent to *argI,* which is an anabolic OCTase gene (Jacoby 1971). In addition, it has been pointed out that OCTase and the catalytic subunit of ACTase are highly homologous with regard to the primary structure and the three-dimensional structure (Houghton et al. 1984; Van Vliet et al. 1984). These findings indicate that gene duplication, which occurred prior to the divergence into the three domains, generated a pair of paralogues, one of which differentiated into the ACTase gene and the other into the OCTase gene (Cunin et al. 1986; Houghton et al. 1984; Van Vliet et al. 1984). Thus, a composite phylogenetic tree can well be constructed using ACTase genes as the outgroup, and the overview of the evolutionary history of OCTase genes of the three domains can be clarified eventually.

Thus, a maximum-likelihood (ML) tree was con-

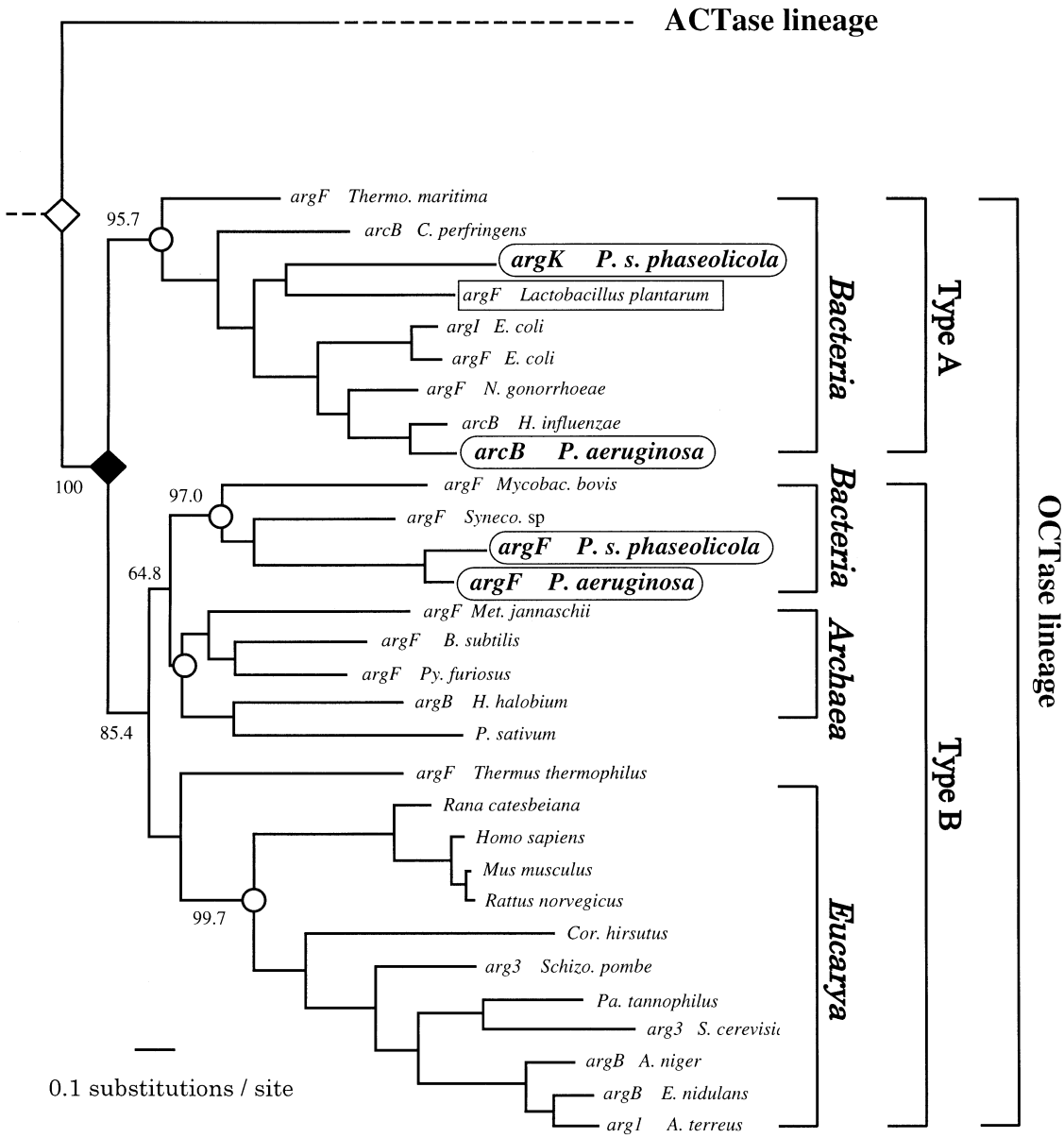

Fig. 3. OCTase tree with ACTase as the outgroup inferred by the maximum-likelihood (ML) method based on the JTT-F model. To clarify the phylogenetic position of *argK* in the evolutionary history of OCTase genes, this ML tree was constructed by processing all presently available amino acid sequences of OCTase using the ProtML program of MOLPHY version 2.3 (Adachi and Hasegawa 1996). The deepest root was placed arbitrarily at a point between the two lineages corresponding to the different proteins (OCTase and ACTase). The horizontal length of each branch is proportional to the estimated number of amino acid substitutions, and the figures above or below the internal branches show the local bootstrap probability (percentage) obtained for 1000 resamplings. ◇ and ◆, at the nodes of the tree, show gene duplication, and ○ shows the commencement of speciation in each domain (lineage). This tree revealed that the ACTase gene and OCTase gene were generated by gene duplication which took place in the very early stage of evolution (the node marked by ◇), the OCTase gene then underwent a second gene duplication (at the node marked by ◆), and hence two paralogues (categorized into Types A and B here) now exist in the OCTase gene.

structed as shown in Fig. 3. Gene duplication, which made genes diverge into ACTase and OCTase genes, appears at the deepest node of this ML tree (shown by the ◇). In addition, it was revealed that gene duplication occurred one more time in the OCTase lineage (at the node shown by the ◆), by which two duplication-originated paralogues were generated in the OCTase lineage. It has already been reported that two groups of prokaryotic OCTase genes exist (Bringel et al. 1997; Falmagne et al. 1985; O'Reilly and Devine 1994; Ruepp et al. 1995), and our present phylogenetic analysis clearly proved that the existence of these two groups is attributable to gene duplication. In this study, these two duplication-originated lineages (paralogues) of OCTase genes are called Type A and Type B (Fig. 3, Table 6).

The Type A lineage contains OCTase genes of only *Bacteria,* while the Type B lineage contains OCTase genes of a variety of organisms belonging to the three

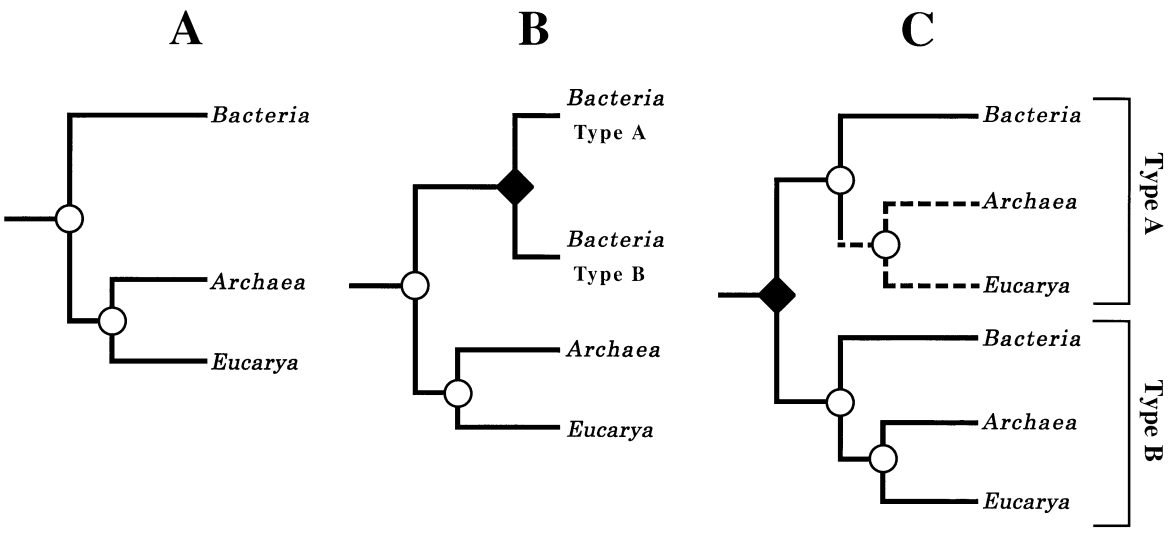

Fig. 4. Three hypotheses of the evolutionary history of OCTase genes of all organisms. ○ and ◆, at the nodes of the tree, show speciation and gene duplication, respectively. **A** The OCTase tree topology in the case when no OCTase gene duplication occurred in the process of evolution. **B** In the case when OCTase gene duplication occurred independently only on the genome of *Bacteria* after the LCA (last common ancestor) diverged into three domains (*Bacteria, Archaea,* and *Eucarya*), as a result of which a pair of paralogues, i.e., Type A and Type B, was generated. **C** In the case when a pair of paralogues (Type A and Type B) was generated due to OCTase gene duplication that occurred on the genome of LCA, followed by divergence into the three domains. The OCTase tree shown in Fig. 3 completely refutes the possibility of the topology in A, and it strongly supports the topology in C, without totally denying the topology in B. The branches shown by *dashed lines* in C denote that OCTase genes which are assumed to have followed these courses of evolution were not detected in the samples used in this analysis (Fig. 3).

domains, i.e., *Bacteria, Archaea,* and *Eucarya.* However, the bootstrap probability, which determines the branching order of the lineages corresponding to the respective domains, turned out to be comparatively low (64.8 and 85.4%). As the last common ancestor [LCA; the universal ancestor from which all extant life ultimately diverged (Glansdorff 2000)] is suspected to have diverged into the three domains during a short period of time in view of the time scale of evolution (Miyata et al. 1991), it may be unlikely to obtain highly reliable results with regard to their branching order using only limited data on OCTase genes. Thus, no definitive conclusion is given in this paper with regard to the branching order of the lineages corresponding to the respective domains in the OCTase tree.

Based on the results obtained above, the evolutionary history of the OCTase genes of all three domains was examined. If OCTase genes did not undergo duplication, their evolutionary history would have been expressed by a simple-shaped topology as shown in Fig. 4A; however, our analytical results (Fig. 3) refuted this possibility.

Figure 4C shows the evolutionary history of OCTase genes prepared based on the phylogenetic tree shown in Fig. 3. Specifically, Fig. 4C shows the following course of evolution: the OCTase gene underwent duplication prior to the divergence of LCA; a pair of paralogues, Types A and B, was generated; and then the divergence of LCA into the three domains (*Bacteria, Archaea,* and *Eucarya*) occurred. However, as no definitive conclusions regarding the branching order of the lineages corresponding to the respective domains can be drawn based on the phylogenetic tree shown in Fig. 3, the topology shown in Fig. 4B also remains as another possibility, meaning that the OCTase gene might have duplicated independently only in the lineage of *Bacteria* after LCA diverged into the three domains. Nonetheless, regardless of the correctness of either Fig. 4B or Fig. 4C, the basic course of evolution was clarified as follows: "After the ACTase gene and OCTase gene were generated by gene duplication, the OCTase gene was duplicated for a second time; as a result, a pair of duplication-originated paralogues (OCTase genes, one each belonging to Type A and Type B, respectively) now exists at least on the genomes of *Bacteria.*" These findings are used to examine the foreign origination of *argK* in the following subsection.

Construction of a 16S rDNA Tree (Criterion for Judging Foreign Origination) and Comparison with the OCTase Tree. An attempt was made to clarify the "course of evolution of the *P. syringae* genome" for use as the criterion for determining the foreign origination of the *argK–tox* cluster. Accordingly, another ML tree (Fig. 5), which expresses the phylogenetic relationship of representative *Bacteria* including *P. syringae,* was constructed using *Archaea* as the outgroup and 16S rDNA as the index. As the branching pattern reproduced in the lineage of *Bacteria* in this 16S rDNA tree (Fig. 5) agrees with the widely acknowledged pattern, i.e., "following the divergence of *Thermotoga,* the Gram-positive bacteria diverged, then *Proteobacteria* became further diverged" (Hugenholtz et al. 1998; Pace 1997), this tree is

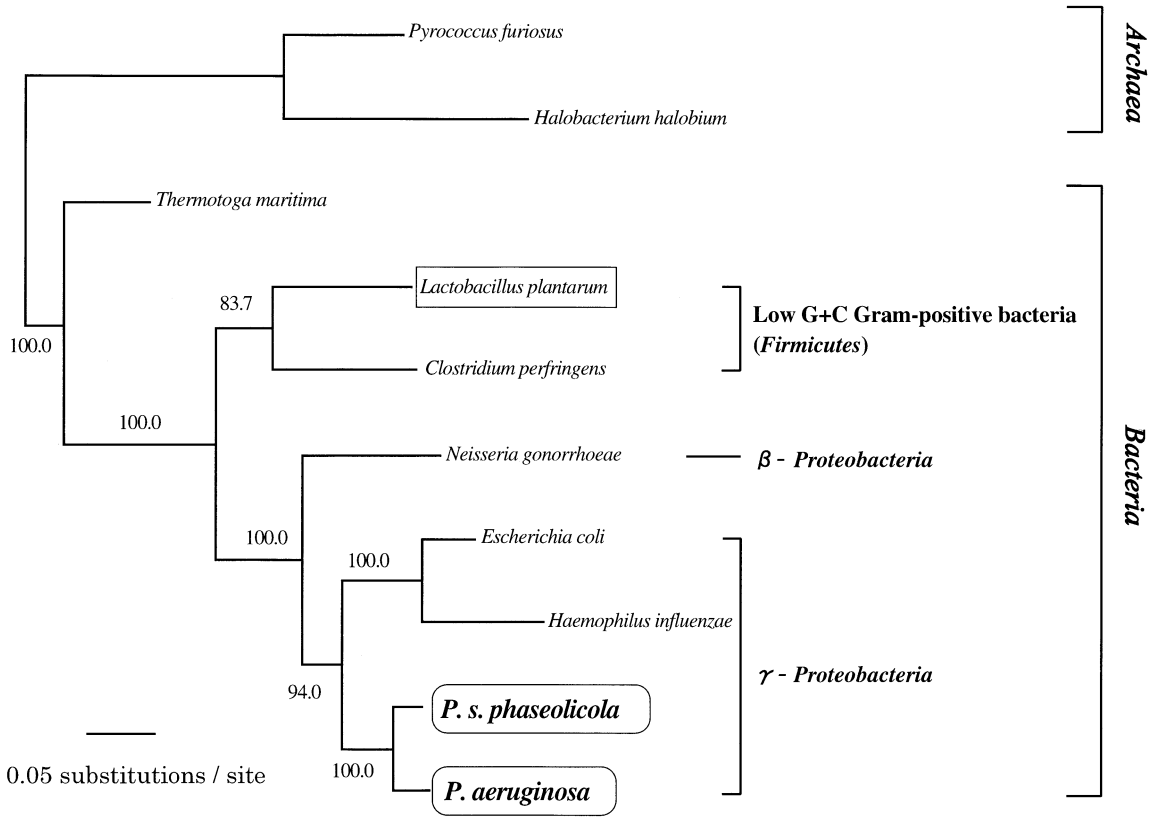

Fig. 5. 16S rDNA tree of *Bacteria,* with *Archaea* as the outgroup, obtained by the ML method using NucML (HKY85 model, $\alpha/\beta = 1.78$). The horizontal length of each branch is proportional to the estimated number of nucleotide substitutions, and the figures above or below the internal branches show the local bootstrap probability (percentage) obtained for 1000 resamplings. This tree shows the phylogenetic relationship among bacterial species which constitute the Type A lineage of the OCTase tree (Fig. 3).

considered to be an adequate criterion for judging whether or not *argK* is of foreign origin.

In the 16S rDNA tree, *P. syringae* belongs to the lineage of the γ-*Proteobacteria* and is located very close to *P. aeruginosa* (Fig. 5). Incidentally, it has already been reported that both *P. syringae* and *P. aeruginosa* belong to the genus *Pseudomonas* (*sensu stricto*) in the γ-*Proteobacteria* and are very closely related to each other (Anzai et al. 2000; Moore et al. 1996). This is evidently proven by the phylogenetic position of *P. syringae* in our 16S rDNA tree shown in Fig. 5.

On the other hand, the Type A lineage in our OCTase tree shows that *argK* is in close relation with the OCTase gene (*argF*) of *Lactobacillus plantarum* (a bacterial species belonging to the low-G+C Gram-positive bacteria or *Firmicutes*) (Fig. 3). These low-G+C Gram-positive bacteria are known to be extremely distantly related to *P. syringae* in terms of evolution (Hugenholtz et al. 1998; Pace 1997) and are located extremely far from *P. syringae* in the 16S rDNA tree (Fig. 5). Showing that the course of evolution of the *P. syringae* genome (Fig. 5) is greatly different from the course of evolution of *argK* (Fig. 3), these results strongly suggest the foreign origination of *argK.*

Species-Specific Diversity of Genes in Codon Usage: Validation of the Foreign Origination of the argK–tox *Cluster (2)*

To confirm whether or not the *argK–tox* cluster is adaptive to the genome of *P. syringae,* analysis was further conducted based on the species-specific diversity of genes in codon usage using a method developed by Kanaya et al. (1996a, b). This method, which has already been successfully applied to several bacterial species (Kanaya et al. 1996b, 1997, 1999; Nakayama et al. 1999), constructs a measure based on the widest scale of the gene distribution in codon frequency space. This axis is referred to as the first principal component (PC1) and its scale is denoted Z_1 (see Materials and Methods). Analysis of two genes that are present within the *argK–tox* cluster, namely, *argK* and the "fatty acid desaturase gene" (DES-ORF) (Hatziloukas et al. 1995), provided the following results.

In *P. syringae* genomic (native) genes, PC1 accounts for 30.4% of the total variance, which is much larger than those for other PCs (e.g., 10.7% for the second PC). This indicates that the *P. syringae*-specific diversity is roughly reflected in PC1. Table 3 shows factor loadings

Table 3. Factor loadings (FL) for the first principal component

Amino acid	NNA codon		NNU codon		NNG codon		NNC codon	
	Codon	FL	Codon	FL	Codon	FL	Codon	FL
Arg	CGA	−0.583	CGU	0.125	CGG	0.317	CGC	0.743
	AGA	−0.744	—	—	AGG	−0.462	—	—
Leu	UUA	−0.785	—	—	UUG	−0.431	—	—
	CUA	−0.448	CUU	−0.651	CUG	0.861	CUC	−0.183
Ser	UCA	−0.520	UCU	−0.740	UCG	0.285	UCC	−0.049
	—	—	AGU	−0.109	—	—	AGC	0.557
Ala	GCA	−0.784	GCU	−0.516	GCG	0.629	GCC	0.602
Gly	GGA	−0.815	GGU	−0.121	GGG	−0.124	GGC	0.671
Pro	CCA	−0.415	CCU	−0.611	CCG	0.641	CCC	0.095
Thr	ACA	−0.508	ACU	−0.516	ACG	−0.012	ACC	0.597
Val	GUA	−0.580	GUU	−0.718	GUG	0.667	GUC	0.312
Ile	AUA	−0.717	AUU	−0.359	—	—	AUC	0.849
Asn	—	—	AAU	−0.554	—	—	AAC	0.554
Asp	—	—	GAU	−0.760	—	—	GAC	0.760
Cys	—	—	UGU	−0.313	—	—	UGC	−0.313
Gln	CAA	−0.457	—	—	CAG	0.457	—	—
Glu	GAA	0.229	—	—	GAG	−0.229	—	—
His	—	—	CAU	−0.364	—	—	CAC	0.364
Lys	AAA	−0.235	—	—	AAG	0.235	—	—
Phe	—	—	UUU	−0.599	—	—	UUC	0.599
Tyr	—	—	UAU	−0.172	—	—	UAC	0.172

for PC1 which represent contributions of codon frequencies to PC1. The properties of the codon-usage diversity revealed in our analysis were as follows: most of the codon frequencies with A and U at the third codon position contributed negatively to PC1, while most of those with C and G at the third codon position contributed positively to PC1; however, the frequencies for two codons (GAA and CGU) contributed positively to PC1, while those for eight codons (AGG, UUG, GGG, ACG, GAG, CUC, UCC, and UGC) contributed negatively to PC1.

A *t* test was conducted to assess the effectiveness of Z_1 as a means of discriminating genomic (native) genes from foreign genes. Table 4 shows the *t* values for Z_1 and the GC content at the third codon position (3rd GC), which were obtained by comparing the average values and standard deviations of the two groups [the genomic (native) gene group (G-group) and the foreign gene group (F-group)]. These *t* values prove that Z_1 was better able to discriminate native genes from foreign genes than the 3rd GC, because the significance level of Z_1 ($p < 0.1\%$) was much lower than that of the 3rd GC ($0.1 < p < 1.0\%$). So Z_1 was used to assess the adaptiveness of *argK* and DES-ORF to the *P. syringae* genome, the results of which revealed that the Z_1 values of *argK* (−0.95) and DES-ORF (−1.36) were very close to the average of the foreign gene group (Table 5). This finding could roughly be confirmed by their 3rd GC values (0.552 and 0.554, respectively) as well, which were much lower than that (0.68) of the G-group (Table 5). Accordingly, the analytical results obtained by both the codon usage and the 3rd GC proved that the *argK–tox* cluster is of foreign origin.

Table 4. *t* values between the genomic (native) gene group (G-group) and the foreign gene group (F-group)[a]

	G-group		F-group		
	Av	(SD)	Av	(SD)	*t* value
Z_1	0.00	(1.00)	−1.15	(1.16)	4.20
3rd GC	0.68	(0.10)	0.57	(0.14)	3.36

[a] Av and SD represent the group average and standard deviation, respectively. Genes larger than 500 bases in plasmids and in transposable elements were included in the foreign gene group: *avrC*, *avrD* (race 0, race 1, race 2, race 3, race 6), *avrPmaA1*, *avrPphC*, *avrPpiB1.R3*, *cfa3*, *cfl*, *cmaA*, *cmaT*, *cmaU*, *copA*, *copB*, *copR*, *copS*, *corS*, *iaaH*, *iaaL*, *iaaM*, *ptz*, *rulB*, *stbA*, *stbD*, *tnpA*, *tnpR*, and ethylene-forming enzyme gene. Genomic (native) genes are listed under Materials and Methods.

Thus, all the results obtained by the analyses discussed in this section, the phylogenomic analyses discussed in the previous sections, and the phylogenetic analyses conducted in our previous studies (Sawada et al. 1997, 1999) strongly and consistently support the hypothesis that the "*argK–tox* cluster is of foreign origin and has expanded its distribution by horizontally transferring onto the genome of *P. syringae* from bacterial species distantly related to *P. syringae*."

Evolution of OCTase Homologues on the P. syringae *Genome*

The previous sections have described mainly the results of analysis of "*argK*" existing within the *argK–tox* clus-

Table 5. Comparison between the *P. syringae* genome and the *argK–tox* cluster based on the 3rd GC and codon usage

	P. syringae genome			*argK–tox* cluster	
	Whole genome[a]	G-group[b]	*argF*	*argK*	DES-ORF
GC content (%)	56	56–58	57.3	49.4	49.6
3rd GC (%)	—	68	67.8	55.2	55.4
Codon adaptation[c]	—	—	0.00	−0.95	−1.36

[a] Data from Palleroni (1984).
[b] Genomic (native) gene group (see Table 4 and Materials and Methods).
[c] Z_1 values (see Materials and Methods).

ter. Incidentally, "ArgK," which is one type of anabolic OCTase, has been revealed to have a salient pathogenicity-related function as follows: phaseolotoxin, in whose biosynthesis the *argK–tox* cluster is involved, is a virulence factor that makes host plants exhibit a disease symptom (chlorosis) by inhibiting the catalysis of their anabolic OCTase (Bender et al. 1999); however, phaseolotoxin-producing pathovars (pv. *actinidiae* and pv. *phaseolicola*) do not suffer autointoxication because they possess anabolic OCTase encoded by *argK* (Hatziloukas and Panopoulos 1992; Mosqueda et al. 1990; Sawada et al. 1997), which are phaseolotoxin resistant and can function even in the presence of phaseolotoxin (Rudolph 1990; Staskawicz et al. 1980; Templeton et al. 1986).

Incidentally, there exists another anabolic OCTase gene, namely, *argF,* on the *P. syringae* genome in addition to *argK* (Fig. 2). It is known that the isozyme encoded by *argF* is phaseolotoxin sensitive, is irrelevant to pathogenicity, and exerts a housekeeping function when the *argK–tox* cluster suspends phaseolotoxin production (Jahn et al. 1985).

Specifically, these OCTase genes (*argK* and *argF*) on the *P. syringae* genome have important functions that are essential for the survival of *P. syringae.* Accordingly, our interest focused on the OCTase genes, and they were analyzed based on various aspects to clarify the features of the OCTase homologues on the *P. syringae* genome, as well as the courses of evolution which they have followed.

argF *Is a Native Gene.* Four findings were obtained with regard to *argF* of *P. syringae,* as follows. (i) In the OCTase tree (Fig. 3), *argF* of *P. syringae* belongs to the Type B lineage, being located extremely close to *argF* of *P. aeruginosa.* This phylogenetic position of *argF* in the OCTase tree (Fig. 3) agrees with that of *P. syringae* in the 16S rDNA tree (Fig. 5), meaning that, unlike *argK,* the course of evolution of *argF* agrees with that of the *P. syringae* genome. (ii) The analytical results obtained by codon usage and 3rd GC showed that both the Z_1 value (0.00) and the value of the 3rd GC (0.678) of *argF* agree with the general tendency of other genes on the *P. syringae* genome (Table 5). (iii) All the strains of *P. syringae* used in this study were proven to possess the *argF* gene (Fig. 6). (iv) The position of *argF* on the genome was confirmed to be conserved in all three pathovars investigated (Fig. 2). As these findings showed that the properties of *argF* are totally opposite those of *argK,* it can be concluded unequivocally that "*argF* is not of foreign origin but is a native gene that has resided on the *P. syringae* genome from the beginning."

Search for the Paralogue of argF (*the Orthologue of* arcB *of* P. aeruginosa): arcB *Has Been Deleted from the* P. syringae *Genome.* The finding that *argK* is foreign and *argF* is native to the *P. syringae* genome made us suspect the existence of one more native OCTase gene on the *P. syringae* genome, because the results of our analysis of OCTase genes proved that paralogues originated from gene duplication exist in pairs (OCTase genes, one each belonging to Types A and B, respectively) on the genomes of *Bacteria* (Figs. 4B and C, Table 6). Specifically, if the same applies to the *P. syringae* genome, there should be an OCTase gene that belongs to Type A, or the paralogue of *argF,* on the *P. syringae* genome, as the other OCTase gene of the pair, *argF* that belongs to Type B, has already been confirmed to exist, as mentioned in the preceding subsection.

Incidentally, *P. aeruginosa* is known to possess two copies of OCTase genes on its genome, i.e., a catabolic OCTase gene (*arcB*) and an anabolic OCTase gene (*argF*) (Cunin et al. 1986; Itoh et al. 1988). As with the analyses discussed in the previous sections, these two copies of genes were investigated based on codon usage (data not shown), 3rd GC (data not shown), and phylogenetic analysis (Figs. 3 and 5). The results suggested that both of them are native genes which have resided on the genome of *P. aeruginosa* from the beginning. This means in turn that a pair of paralogues, which were generated by gene duplication, has undergone functional differentiation on the genome of *P. aeruginosa* or its ancestor: one (Type A) into *arcB* and the other (Type B) into *argF* (Fig. 3, Table 6).

With respect to the OCTase genes of *P. putida, P. mendocina,* and *P. fluorescens,* which are extremely closely related to *P. aeruginosa,* we deduced that a pair of paralogues (*arcB* and *argF*), which have undergone functional differentiation as with *P. aeruginosa,* also exist on the genomes of these bacterial species (Table 6) based on the following three findings: (1) all these bacterial species possess both catabolic and anabolic OCTases (Cunin et al. 1986; Ramos et al. 1967; Stalon et al. 1984); (2) both of these enzymes can catalyze only a unidirectional reaction (the former catalyzes the phosphorolysis of citruline, and the latter the reverse reaction) (Stalon et al. 1977); and (3) the results of Southern blot analysis conducted by us, which are discussed later in this section, showed that both of these bacterial species

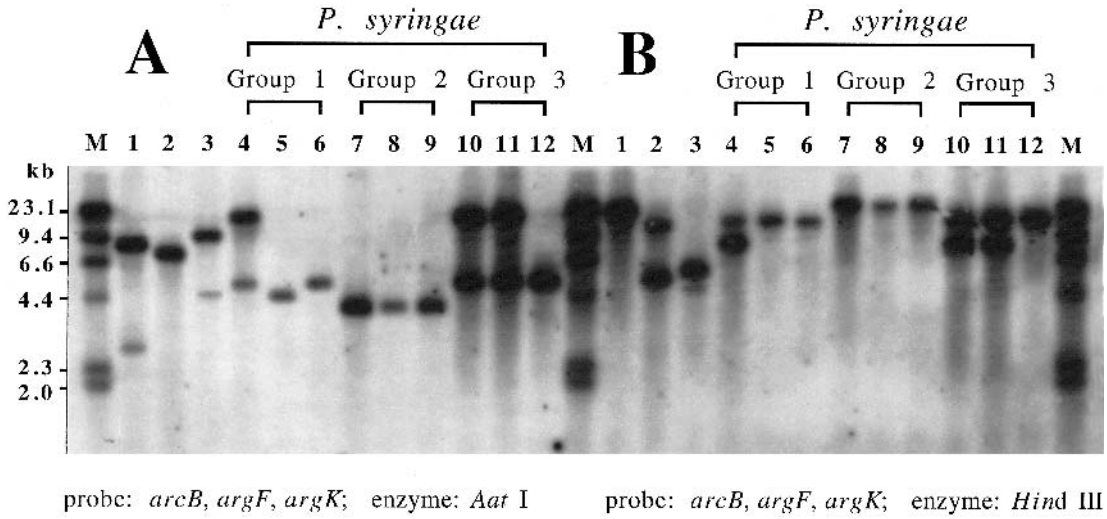

Fig. 6. Southern blot analysis of the hybridization of OCTase-encoding genes (*arcB, argF,* and *argK*) with *Pseudomonas* species. The genomic DNAs of *Pseudomonas* species were digested with either *Aat* I (**A**) or *Hin*d III (**B**), electrophoresed by the conventional method, and analyzed by Southern blot hybridization together with all three of the OCTase gene probes used (*arcB, argF,* and *argK;* see Table 2). Lane 1, *P. aeruginosa* strain PAO1; Lane 2, *P. fluorescens* strain MAFF 302208^{T}; Lane 3, *P. putida* strain IAM 1236; Lane 4, *P. syringae* pv. *actinidiae* strain KW-11^{PT}; Lane 5, *P. s.* pv. *maculicola* strain MAFF 302264^{PT}; Lane 6, *P. s.* pv. *theae* strain PT1^{PT}; Lane 7, *P. s.* pv. *syringae* strain MAFF 302155^{T}; Lane 8, *P. s.* pv. *pisi* strain MAFF 302269^{PT}; Lane 9, *P. s.* pv. *aptata* strain MAFF 302253^{PT}; Lane 10, *P. s.* pv. *phaseolicola* strain MAFF 302282^{PT}; Lane 11, *P. s.* pv. *phaseolicola* strain KUZ 1; Lane 12, *P. s.* pv. *glycinea* strain MAFF 302260^{PT}; and Lane M, DIG-labeled molecular weight marker (λDNA cleaved with *Hin*dIII). Conditions of hybridization and detection: hybridization buffer, DIG Easy Hyb; concentration of DIG-labeled probe in hybridization solution, 25 ng/mL; temperature for hybridization, 40°C; temperature for washing in 0.5 × wash solution, 60°C; and chemiluminescent substrate, CDP-Star. For other conditions for this Southern blot analysis, see Materials and Methods. The bands shown here were detected when a mixture of all three OCTase gene probes (*arcB, argF,* and *argK*) was used. It was confirmed in advance that *Aat*I and *Hin*d III had no site in the coding regions of *arcA, arcD,* or all OCTase genes of *Pseudomonas* (*sensu stricto*) reported up to the present. Accordingly, all native OCTase genes (both Type A and Type B) and *argK* existing on the genomes of the respective strains tested were detected as a single band, respectively. Nevertheless, in both Lane 2 (*P. fluorescens*) in A and in Lane 1 (*P. aeruginosa*) in B, only one band is visible, because two separate bands, which were actually detected, are superimposed at almost the same position.

had two copies of OCTase genes, one each belonging to Type A and Type B, respectively (Figs. 6A and B).

Accordingly, it can be expected that a pair of paralogues, which have undergone a similar functional differentiation (Table 6), exists on the genome of *P. syringae,* because *P. syringae* is extremely closely related to these bacterial species belonging to *Pseudomonas* (*sensu stricto*) (Fig. 5). In other words, as one of the pair of paralogues of *P. syringae* has been revealed to be an anabolic OCTase gene (*argF*) that belongs to Type B, the other paralogue that should belong to Type A is naturally expected to be a catabolic OCTase gene similar to *arcB* of *P. aeruginosa.* Thus, an attempt was made to clarify the OCTase gene composition on the *P. syringae* genome by validating the results obtained by Southern blot analysis.

For the Southern blot analysis, three types of probes were prepared (Table 2): (1) a probe originated from *arcB* of *P. aeruginosa* [for detecting Type A OCTase genes from *Pseudomonas* (*sensu stricto*)], (2) a probe originated from *argF* of pv. *phaseolicola* [for detecting Type B OCTase genes from *Pseudomonas* (*sensu stricto*)], and (3) a probe originated from *argK* of pv. *phaseolicola* [for detecting *argK* that is foreign to *Pseudomonas* (*sensu stricto*)]. Preliminary tests of the specificity of the probes (data not shown) showed that these probes were capable of detecting genes that are orthologous to the respective genes used as probes from *Pseudomonas* (*sensu stricto*) and that they would not hybridize to paralogues under the conditions of hybridization and detection established for this Southern blot analysis (hybridization buffer, DIG Easy Hyb; concentration of DIG-labeled probe in hybridization solution, 25 ng/ml; temperature for hybridization, 40°C; temperature for washing in 0.5× wash solution, 60°C; and chemiluminescent substrate, CDP-Star).

Five strains were selected from each of the three monophyletic groups [Groups 1 to 3 (Sawada et al. 1999)] of *P. syringae* as representative strains, which numbered 15 strains in all (Table 1). In addition, *P. aeruginosa, P. fluorescents,* and *P. putida,* which all belong to *Pseudomonas* (*sensu stricto*), were used as the reference strains. Genomic DNAs prepared from these strains were digested by using one of either *Aat*I, *Hin*dIII, or *Not*I [it had been confirmed in advance that these three restriction enzymes had no cleavage site within the coding regions in any of the genes, such as *arcA* (for arginine deiminase), *arcD* (arginine-ornithine antiporter), and all OCTase genes in *Pseudomonas* (*sensu stricto*)

Table 6. Composition of native OCTase genes presently existing on the genomes of *Pseudomonas* (*sensu stricto*)

Species	Pathovar(s)	No. of native OCTase genes presently existing[a]	Which OCTase type?[a]	
			Type A	Type B
P. aeruginosa		2	*arcB*	*argF*
P. putida		2	(*arcB*)[b]	(*argF*)[b]
P. fluorescens		2	(*arcB*)[b]	(*argF*)[b]
P. syringae	*actinidiae*	1[c,d]	—[c,d]	*argF*
	phaseolicola	1[c,d]	—[c,d]	*argF*
	syringae	1[c]	—[c]	*argF*
	The other pathovars	1[c]	—[c]	*argF*

[a] OCTase genes of Type A and Type B are a pair of paralogues generated due to gene duplication that occurred in the early stage of evolution (see the text and Figs. 3 and 8). Such native OCTase genes which currently remain on the genome of each species are listed here. Accordingly, those which do not remain on the genomes because they have been deleted, and those that are "foreign" because they entered the genomes by conducting horizontal transfer later, are not listed. A — denotes that the gene has been deleted from the genome.

[b] With respect to *P. putida* and *P. fluorescens*, sufficient data are not yet available for concluding definitively which type (Types A or B) has functionally differentiated into which gene [*arcB* (catabolic OCTase gene) or *argF* (anabolic OCTase gene)]. However, given that Type B has differentiated into *argF* (see the text and Fig. 3) in the case of both *P. aeruginosa* and *P. syringae*, it can be assumed that a pair of paralogues, which have undergone functional differentiation similar to that of *P. aeruginosa* and *P. syringae*, also exists on the genomes of *P. putida* and *P. fluorescens*, because these four species are extremely closely related to one another. Thus, it appears that their Type A gene may have differentiated into *arcB* and their Type B gene into *argF*.

[c] The "native" OCTase gene (presumably a catabolic OCTase gene that is orthologous to *arcB* of *P. aeruginosa*) belonging to Type A of *P. syringae* was shown to have been deleted from the genome together with the *arc* operon that contained it (see the text and Figs. 6–8).

[d] "Foreign" *argK* that belongs to Type A entered the genomes of pv. *actinidiae* and pv. *phaseolicola* by conducting horizontal transfer very recently (see the text and Figs. 3, 6, and 8).

that have hitherto been reported], then subjected to Southern blot analysis (Figs. 6 and 7).

Figure 6 shows the results when the mixture of all three OCTase gene probes (*arcB, argF,* and *argK*) was used at the same time. All the native OCTase genes (both Type A and Type B) and *argK* existing on the genomes of the respective strains came out as a single band, respectively. However, only one band is visible in Lane 2 (*P. fluorescens*) in Fig. 6A and Lane 1 (*P. aeruginosa*) in Fig. 6B, because in these lanes two separate bands, which were actually detected when each probe was used separately, are superimposed at almost the same position.

In the case of *P. aeruginosa, P. fluorescens,* and *P. putida* (Lanes 1 to 3 in Figs. 6A and B), which were used as the reference strains, two separate bands could be detected from each strain, i.e., one band for each of the orthologues of *arcB* (Type A) and *argF* (Type B). Regarding *P. syringae,* each of the 15 strains tested was confirmed to have a single copy of *argF* (Type B) (Lanes 4 to 12 in Figs. 6A and B). However, in the case of pv. *actinidiae* (Lane 4 in Figs. 6A and B) and pv. *phaseolicola* (Lanes 10 and 11 in Figs. 6A and B), two distinct bands were detected, which are equivalent to *argF* (Type B) and *argK* (Type A), respectively.

Nevertheless, no band corresponding to an orthologue of *arcB* (Type A) of *P. aeruginosa* could be detected from any of the *P. syringae* pathovars tested (Lanes 4 to 12 in Figs. 6A and B).

Incidentally, it is known that, in the case of *P. aeruginosa,* four genes including *arcB,* which encode the enzymes of the arginine deiminase pathway (the ADI pathway), are organized in an operon called the *arcDABC* operon (Baur et al. 1987, 1989; Gamper et al. 1991). Accordingly, another attempt was made to detect an orthologue of either *arcA* or *arcD* of *P. aeruginosa* from *P. syringae* by conducting a Southern blot analysis (the results shown in Fig. 7 are for *arcA* only; those for *arcD* are not shown); however, no such bands could be detected from any of the *P. syringae* pathovars (Lanes 4 to 12 in Figs. 7A and B). Meanwhile, it is also known that no enzymatic activity of arginine dihydrolase (i.e., arginine deiminase encoded by the *arcA* gene; EC 3.5.3.6) can be detected from *P. syringae* (Palleroni 1984; Sawada et al. 1997). Thus, it can be concluded that no structure of the *arcDABC* operon which contains an orthologue of *arcB* of *P. aeruginosa* exists on the genome of *P. syringae.*

It is known that the *arc* operon in *P. aeruginosa* is involved in the arginine deiminase pathway, which is inducible under anaerobic conditions and serves to generate ATP from arginine (Baur et al. 1987, 1989; Cunin et al. 1986; Gamper et al. 1991). Nonetheless, it is considered unlikely that *P. syringae* is placed under such anaerobic conditions, as it is an epiphyte living mainly on the plant surface, which is in an extremely aerobic environment (Hirano and Upper 2000). Accordingly, the *arc* operon can hardly be an element essential for the survival of *P. syringae.* Thus, it appears that the *arc* operon which contained *arcB* (Type A) was deleted from the genome of the ancestor of *P. syringae* in the phase before the ancestor diverged into the respective pathovars, and hence only the anabolic OCTase gene (*argF*) belonging to Type B remains on the genomes of the respective *P. syringae* pathovars (Table 6).

Conclusions

P. syringae was shown to be extremely rich in genetic diversity in our previous studies (Sawada et al. 1997, 1999). Specifically, the results of our phylogenetic analysis of four genes (*gyrB, rpoD, hrpL,* and *hrpS*) revealed that the ancestor of *P. syringae* (the common ancestor prior to the divergence into pathovars) had so far diverged into at least three monophyletic groups

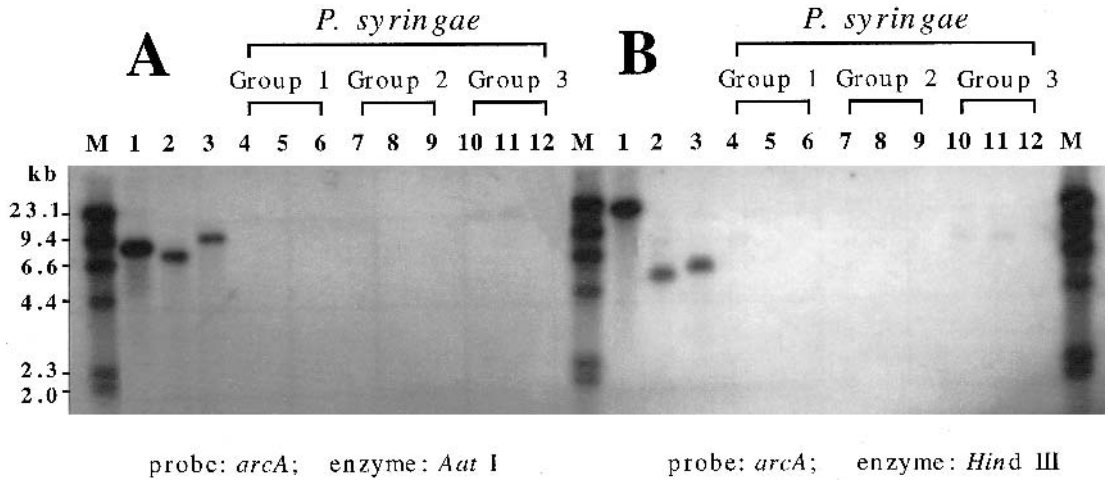

Fig. 7. Southern blot analysis of the hybridization of the *arcA* gene with *Pseudomonas* species. The blot used here is the same as that shown in Fig. 6. For details of the samples electrophoresed in each lane and the conditions of hybridization and detection, see the legend to Fig. 6.

(Groups 1 to 3) in the process of its evolution. In addition, the genome of *P. syringae* was shown to be enormously rich in plasticity in this study. Specifically, the fact that the structure and size of the *P. syringae* genome differ greatly among the three groups was confirmed to be attributable to the active and large-scale genomic rearrangement, including insertion, deletion, and translocation, which occurred while these groups followed different courses of evolution after the ancestor of *P. syringae* diverged into the three groups (Figs. 1 and 2).

The argK–tox *Cluster Is of Foreign Origin and Has Expanded Its Distribution by Conducting Horizontal Transfer*

Furthermore, a phylogenomic analysis was conducted in this study by combining all the information on the molecular evolution, genomic structures, and pathogenicity-related genes (*argK–tox* cluster) to clarify the evolutionary mechanism that caused the pathogenic differentiation of *P. syringae.* The results showed that the *argK–tox* cluster has expanded its distribution by transferring horizontally onto the genomes of the two pathovars (pv. *actinidiae* and pv. *phaseolicola*) of *P. syringae* from bacterial species distantly related to *P. syringae* quite recently in view of the time scale of evolution (i.e., after the ancestor of *P. syringae* diverged into the respective pathovars).

The *argK–tox* cluster is involved in the biosynthesis of phaseolotoxin, which is known to be a virulence factor that causes a disease symptom (chlorosis) by inhibiting the enzymatic activity of anabolic OCTase in host plants (Bender et al. 1999). Accordingly, these two pathovars (pv. *actinidiae* and pv. *phaseolicola*) apparently became able to cause disease symptoms because they obtained the *argK–tox* cluster, which entered their genomes by conducting horizontal transfer, and hence they started attracting attention as phytopathogenic bacteria. Specifically, the results of this study clarified a dynamic evolutionary mechanism in which the pathogenicity-related genes (*argK–tox* cluster), which are present on the genomes, caused a new pathogenic differentiation as a result of conducting horizontal transfer among bacterial species.

The Evolutionary History of the OCTase Genes Unfolded on the P. syringae *Genome: An Unusual Composition of OCTase Genes of Phaseolotoxin-Producing Pathovars*

OCTase genes including *argK,* which are essential for the survival and pathogenicity of *P. syringae* (Jahn et al. 1985; Rudolph 1990; Staskawicz et al. 1980; Templeton et al. 1986), have been the focus of our analyses, reviews, and deliberations throughout our studies. Using all the data we have obtained in this and past studies (Sawada et al. 1997, 1999), the "evolutionary history of OCTase genes unfolded on the genome of *P. syringae*" was reproduced to scrutinize the relationship between the course of evolution of OCTase and the evolution and diversification of *P. syringae.* Figure 8 outlines the evolutionary history of OCTase genes.

It was revealed that gene duplication occurred on the genome of LCA, by which the ACTase gene and OCTase gene were generated (Fig. 8A), then the OCTase gene was duplicated one more time, by which a pair of paralogues (OCTase genes belonging to Types A and B) was generated on the genome of *Bacteria* (Fig. 8B). It is likely that such a pair of paralogues subsequently conducted functional differentiation on the genome of the common ancestor of *Pseudomonas* (*sensu stricto*): one (Type A) into *arcB* (catabolic OCTase gene) and the other (Type B) into *argF* (anabolic OCTase gene) (Fig. 8B). Then the *arcDABC* operon that contained *arcB* (catabolic OCTase gene) was deleted from the genome

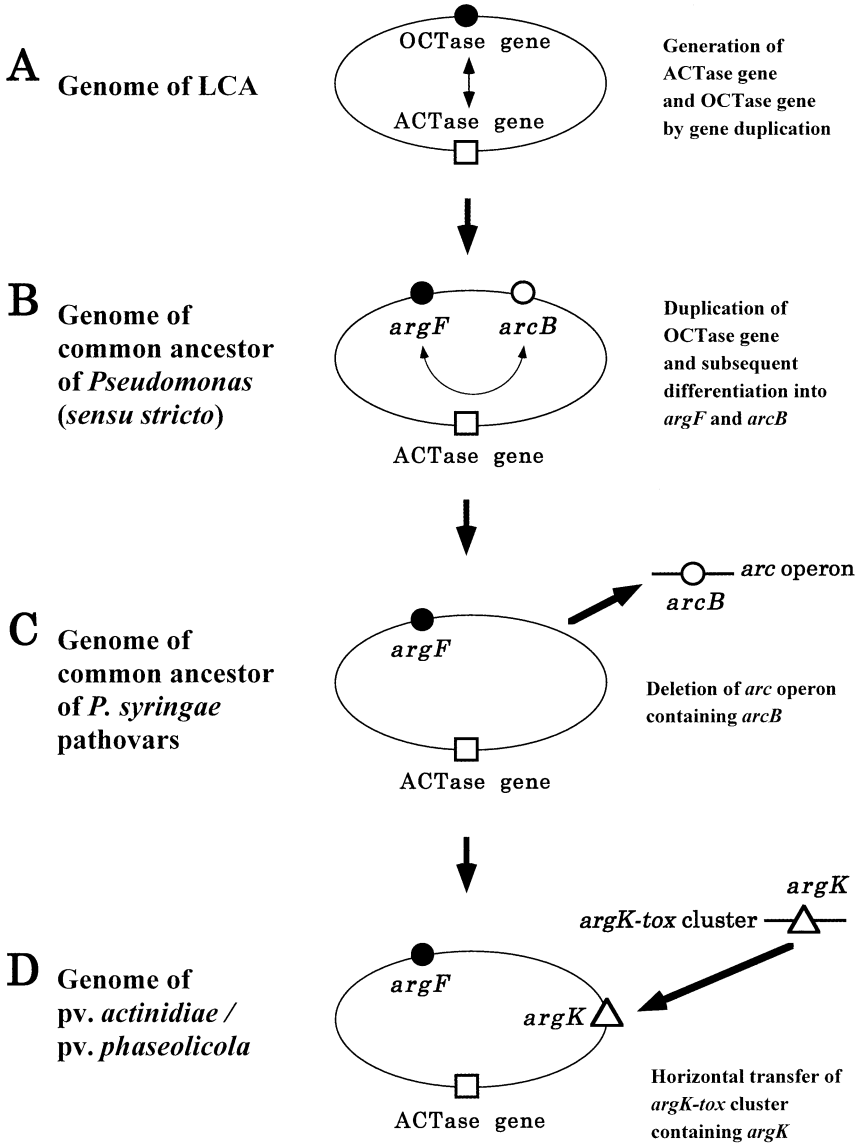

Fig. 8. Evolutionary history of OCTase genes. **A** The ACTase gene and OCTase gene were generated due to gene duplication that took place on the genome of the LCA (last common ancestor). **B** After this, the OCTase gene underwent a second duplication, resulting in the generation of a pair of paralogues (OCTase genes, one each belonging to Types A and B, respectively; see Figs. 3 and 4) on the genome of *Bacteria.* Furthermore, these paralogues conducted functional differentiation on the genome of the common ancestor of *Pseudomonas* (*sensu stricto*): Type A into *arcB* (catabolic OCTase gene) and Type B into *argF* (anabolic OCTase gene) (see Table 6). **C** Then *arcB* on the genome of the ancestor of *P. syringae* (the common ancestor prior to the divergence into pathovars) was deleted along with the *arc* operon that contained it, and only *argF* remained on the genome. **D** Finally, after the ancestor of *P. syringae* diverged into the respective pathovars, the *argK–tox* cluster which contained *argK* (anabolic OCTase gene) entered the genomes of pv. *actinidiae* and pv. *phaseolicola* from bacterial species other than *P. syringae,* creating an unusual gene composition, i.e., "two copies (*argK* and *argF*) of only anabolic OCTase genes exist on the genomes of these two pathovars."

of the ancestor of *P. syringae* (the common ancestor prior to the divergence into pathovars), and only *argF* remained on the genome (Fig. 8C).

Finally, after the ancestor of *P. syringae* diverged into the respective pathovars, the *argK–tox* cluster containing *argK* (anabolic OCTase gene) transferred horizontally onto the genomes of pv. *actinidiae* and pv. *phaseolicola* from bacterial species distantly related to *P. syringae,* and hence an unusual gene composition in which two copies (*argK* and *argF*) of only anabolic OCTase genes exist on the genomes of these two pathovars was created (Fig. 8D). Having obtained these two anabolic OCTase genes on their genomes, these two pathovars became able to select the suitable gene for use at ordinary times (*argF*) and when they exert phytopathogenicity (*argK*). Thus, it was clarified that the genomic rearrangement (duplication, deletion, and insertion) and horizontal transfer of OCTase genes have played an important role in the evolution and diversification of *P. syringae.*

The catabolic gene cluster of xenobiotic-degrading bacteria and the pathogenicity island of animal pathogens have sometimes been observed to reside on mobile genetic elements, such as transposons, phages, and conjugative plasmids; thus, the mobile genetic elements are suspected to have been the cause of the frequent occurrence of genomic rearrangement and horizontal gene transfer, thereby accelerating the evolution and expanding the distribution of the catabolic gene cluster and the pathogenicity island (Hacker and Kaper 1999; Tsuda et al. 1999).

With regard to *P. syringae,* the *in planta* transfer of coronatine-inducing plasmids has already been identified (Sato 1988). As for pathogenicity-related genes existing on the *P. syringae* genomes, the results obtained in this study have opened a new horizon to their dynamic evolution as follows. It is quite likely that the large population of *P. syringae* strains which live saprophytically on the surfaces of various plants (Hirano and Upper 2000),

including not only pathogenic but also nonpathogenic strains, has functioned as the matrix to give birth to new pathovars (e.g., pv. *actinidiae* and pv. *phaseolicola*) by acquiring pathogenicity-related genes (e.g., the *argK–tox* cluster) onto their genomes through horizontal transfer. Such *P. syringae* strains may also have functioned as "factories" for creating new pathogenicity-related genes by generating genetic variations in the process of their genomic rearrangement.

Nevertheless, the most important factor that caused the plasticity of the *P. syringae* genome, i.e., the molecular mechanism which caused genomic rearrangement and horizontal gene transfer, still has not been identified. Thus, in our future studies, we intend to learn when and where such a pathogenicity island-like structure was organized, and how it gained the ability to conduct horizontal transfer and expanded distribution, using the *argK–tox* cluster as a marker. Eventually, we hope to get a more precise understanding of the molecular mechanism involved in the genomic plasticity and diversification of *P. syringae.*

Acknowledgments. We thank Masami Hasegawa, Yoshifumi Itoh, Satoshi Oota, Naoko Takezaki, and Yuichi Takikawa for their helpful suggestions and Junichi Kato, Koushi Nishiyama, Mamoru Sato, and Yuichi Takikawa for providing the bacterial strains used in this study.

References

Adachi J, Hasegawa M (1996) MOLPHY; Programs for molecular phylogenetics, version 2.3. Institute of Statistical Mathematics, Tokyo

Anzai Y, Kim H, Park J, Wakabayashi H, Oyaizu H (2000) Phylogenetic affiliation of the pseudomonads based on 16S rRNA sequence. Int J Syst Evol Microbiol 50:1563–1589

Baur H, Stalon V, Falmagne P, Luethi E, Haas D (1987) Primary and quaternary structure of the catabolic ornithine carbamoyltransferase from *Pseudomonas aeruginosa.* Eur J Biochem 166:111–117

Baur H, Luethi E, Stalon V, Mercenier A, Haas D (1989) Sequence analysis and expression of the arginine-deiminase and carbamatekinase genes of *Pseudomonas aeruginosa.* Eur J Biochem 179:53–60

Bender CL, Alarcon-Chaidez F, Gross DC (1999) *Pseudomonas syringae* phytotoxins: Mode of action, regulation, and biosynthesis by peptide and polyketide synthetases. Microbiol Mol Biol Rev 63: 266–292

Bergthorsson U, Ochman H (1998) Distribution of chromosome length variation in natural isolates of *Escherichia coli.* Mol Biol Evol 15:6–16

Bradbury JF (1986) Guide to plant pathogenic bacteria. CAB International Mycological Institute, Kew, UK

Bringel F, Frey L, Boivin S, Hubert J-C (1997) Arginine biosynthesis and regulation in *Lactobacillus plantarum:* The *carA* gene and the *argCJBDF* cluster are divergently transcribed. J Bacteriol 179: 2697–2706

Cao Y, Adachi J, Janke A, Paabo S, Hasegawa M (1994) Phylogenetic relationships among eutherian orders estimated from inferred sequences of mitochondrial proteins: Instability of a tree based on a single gene. J Mol Evol 39:519–527

Cole ST, Saint-Girons I (1994) Bacterial genomics. FEMS Microbiol Rev 14:139–160

Cunin R, Glansdorff N, Pierard A, Stalon V (1986) Biosynthesis and metabolism of arginine in bacteria. Microbiol Rev 50:314–352

Dye DW, Bradbury JF, Goto M, Hayward AC, Lelliott RA, Schroth MN (1980) International standards for naming pathovars of phytopathogenic bacteria and a list of pathovar names and pathtype strains. Rev Plant Pathol 59:153–168

Eisen JA (1998) A phylogenomic study of the MutS family of proteins. Nucleic Acids Res 26:4291–4300

Esther De Ita M, Marsch-Moreno R, Guzman P, Alvarez-Morales A (1998) Physical map of the chromosome of the phytopathogenic bacterium *Pseudomonas syringae* pv. *phaseolicola.* Microbiology 144:493–501

Falmagne P, Portetelle D, Stalon V (1985) Immunological and structural relatedness of catabolic ornithine carbamoyltransferases and the anabolic enzymes of enterobacteria. J Bacteriol 161:714–719

Felsenstein J (1981) Evolutionary trees from DNA sequences: A maximum likelihood approach. J Mol Evol 17:368–376

Galperin MY, Tatusov RL, Koonin EV (1999) Comparing microbial genomes: How the gene set determines the lifestyle. In: Charlebois RL (ed) Organization of the prokaryotic genome. ASM Press, Washington, DC, pp 91–108

Gamper M, Zimmermann A, Haas D (1991) Anaerobic regulation of transcription initiation in the *arcDABC* operon of *Pseudomonas aeruginosa.* J Bacteriol 173:4742–4750

Glansdorff N (2000) About the last common ancestor, the universal life-tree and lateral gene transfer: A reappraisal. Mol Microbiol 38:177–185

Hacker J, Kaper JB (1999) The concept of pathogenicity islands. In: Kaper JB, Hacker J (eds) Pathogenicity islands and other mobile virulence elements. ASM Press, Washington, DC, pp 1–11

Hasegawa M, Kishino H (1994) Accuracies of the simple methods for estimating the bootstrap probability of a maximum likelihood tree. Mol Biol Evol 11:142–145

Hasegawa M, Kishino H, Yano T (1985) Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. J Mol Evol 22:160–174

Hatziloukas E, Panopoulos NJ (1992) Origin, structure, and regulation of *argK,* encoding the phaseolotoxin-resistant ornithine carbamoyltransferase in *Pseudomonas syringae* pv. phaseolicola, and functional expression of *argK* in transgenic tobacco. J Bacteriol 174: 5895–5909

Hatziloukas E, Panopoulos NJ, Delis S, Prosen DE, Schaad NW (1995) An open reading frame in the approximately 28-kb *tox-argK* gene cluster encodes a polypeptide with homology to fatty acid desaturases. Gene 166:83–87

Hayashi T, Makino K, Ohnishi M, Kurokawa K, Ishii K, Yokoyama K, Han C, Ohtsubo E, Nakayama K, Murata T, Tanaka M, Tobe T, Iida T, Takami H, Honda T, Sasakawa C, Ogasawara N, Yasunaga T, Kuhara S, Shiba T, Hattori M, Shinagawa H (2001) Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genomic comparison with a laboratory strain K-12. DNA Res 8:11–22

Hirano SS, Upper CD (2000) Bacteria in the leaf ecosystem with emphasis on *Pseudomonas syringae*—A pathogen, ice nucleus, and epiphyte. Microbiol Mol Biol Rev 64:624–653

Honeycutt RJ, McClelland M, Sobral BWS (1993) Physical map of the genome of *Rhizobium meliloti* 1021. J Bacteriol 175:6945–6952

Houghton JE, Bencini DA, O'Donovan GA, Wild JR (1984) Protein differentiation: A comparison of aspartate transcarbamoylase and ornithine transcarbamoylase from *Escherichia coli* K-12. Proc Natl Acad Sci USA 81:4864–4868

Hugenholtz P, Goebel BM, Pace NR (1998) Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. J Bacteriol 180:4765–4774

Itoh Y, Soldati L, Stalon V, Falmagne P, Terawaki Y, Leisinger T, Haas D (1988) Anabolic ornithine carbamoyltransferase of *Pseudomonas aeruginosa:* Nucleotide sequence and transcriptional control of the *argF* structural gene. J Bacteriol 170:2725–2734

Jacoby GA (1971) Mapping the gene determining ornithine transcarbamylase and its operator in *Escherichia coli.* J Bacteriol 108:645–651

Jahn O, Sauerstein J, Reuter G (1985) Detection of two ornithine carbamoyltransferases in a phaseolotoxin-producing strain of *Pseudomonas syringae* pv. *phaseolicola.* J Basic Microbiol 25:543–546

Kanaya S, Kudo Y, Nakamura Y, Ikemura T (1996a) Detection of genes in *Escherichia coli* sequences determined by genome projects and prediction of protein production levels, based on multivariate diversity in codon usage. Comput Appl Biosci 12:213–225

Kanaya S, Kudo Y, Suzuki S, Ikemura T (1996b) Systematization of species-specific diversity of genes in codon usage: Comparison of the diversity among bacteria and prediction of the protein production levels in cells. In: Akutsu T, Asai K, Hagiya M, Kuhara S, Miyano S, Nakai K (eds) Genome Informatics Series No. 7. Universal Academy Press, Tokyo, pp 61–71

Kanaya S, Okumura T, Miyauchi M, Fukasawa H, Kudo Y (1997) Assessment of protein coding sequences in *Bacillus subtilis* genome using species-specific diversity of genes in codon usage based on multivariate analysis: Comparison of the diversity between *B. subtilis* and *Escherichia coli.* Res Commun Biochem Cell Mol Biol 1:82–92

Kanaya S, Yamada Y, Kudo Y, Ikemura T (1999) Studies of codon usage and tRNA genes of 18 unicellular organisms and quantification of *Bacillus subtilis* tRNAs: Gene expression level and species-specific diversity of codon usage based on multivariate analysis. Gene 238:143–155

Khetawat G, Bhadra RK, Nandi S, Das J (1999) Resurgent *Vibrio cholerae* O139: rearrangement of cholera toxin genetic elements and amplification of *rm* operon. Infect Immun 67:148–154

Le Bourgeois P, Daveran-Mingot, M, Ritzenthaler P (2000) Genome plasticity among related *Lactococcus* strains: Identification of genetic events associated with macrorestriction polymorphisms. J Bacteriol 182:2481–2491

Liu S-L, Hessel A, Sanderson KE (1993) Genomic mapping with I-*Ceu* I, an intron-encoded endonuclease specific for genes for ribosomal RNA, in *Salmonella* spp., *Escherichia coli,* and other bacteria. Proc Natl Acad Sci USA 90:6874–6878

Liu S-L, Schryvers AB, Sanderson KE, Johnston RN (1999) Bacterial phylogenetic clusters revealed by genome structure. J Bacteriol 181:6747–6755

Maddison WP, Maddison DR (1996) MacClade program and documentation, Version 3.06. Sinauer Associates, Sunderland, MA

Martin-Didonet CCG, Chubatsu LS, Souza EM, Kleina M, Rego FGM, Rigo LU, Yates MG, Pedrosa FO (2000) Genome structure of the genus *Azospirillum.* J Bacteriol 182:4113–4116

Miyata T, Iwabe N, Kuma K, Kawanishi Y, Hasegawa M, Kishino H, Mukohata Y, Ihara K, Osawa S (1991) Evolution of archaebacteria: Phylogenetic relationships among archaebacteria, eubacteria, and eukaryotes. In: Osawa S, Honjo T (eds) Evolution of life: Fossils, molecules, and culture. Springer-Verlag, Tokyo, pp 337–351

Moore ERB, Mau M, Arnscheidt A, Boettger EC, Hutson RA, Collins MD, Van de Peer Y, De Wachter R, Timmis KN (1996) The determination and comparison of the 16S rRNA gene sequences of species of the genus *Pseudomonas* (sensu stricto) and estimation of the natural intrageneric relationships. Syst Appl Microbiol 19:478–492

Mosqueda G, Van den Broeck G, Saucedo O, Bailey AM, Alvarez-Morales A, Herrera-Estrella L (1990) Isolation and characterization of the gene from *Pseudomonas syringae* pv. *phaseolicola* encoding the phaseolotoxin-insensitive ornithine carbamoyltransferase. Mol Gen Genet 222:461–466

Nakayama K, Kanaya S, Ohnishi M, Terawaki Y, Hayashi T (1999) The complete nucleotide sequence of ϕCTX, a cytotoxin-converting phage of *Pseudomonas aeruginosa:* Implications for phage evolution and horizontal gene transfer via bacteriophages. Mol Microbiol 31:399–419

O'Reilly M, Devine KM (1994) Sequence and analysis of the citrulline biosynthetic operon *argC-F* from *Bacillus subtilis.* Microbiology 140:1023–1025

Pace NR (1997) A molecular view of microbial diversity and the biosphere. Science 276:734–740

Palleroni NJ (1984) Genus I. *Pseudomonas* Migula 1894. In: Krieg NR, Holt JG (eds) Bergey's manual of systematic bacteriology, Vol 1. Williams & Wilkins, Baltimore, pp 141–199

Ramos F, Stalon V, Pierard A, Wiame JM (1967) The specialization of the two ornithine carbamoyltransferases of *Pseudomonas.* Biochim Biophys Acta 139:98–106

Rode CK, Melkerson-Watson LJ, Johnson AT, Bloch CA (1999) Type-specific contributions to chromosome size differences in *Escherichia coli.* Infect Immun 67:230–236

Römling U, Duchene M, Essar DW, Galloway D, Guidi-Rontani C, Hill D, Lazdunski A, Miller RV, Schleifer KH, Smith DW, Toschka HY, Tümmler B (1992) Localization of *alg, opr, phn, pho,* 4.5S RNA, 6S RNA, *tox, trp,* and *xcp* genes, *rm* operons, and the chromosomal origin on the physical genome map of *Pseudomonas aeruginosa* PAO. J Bacteriol 174:327–330

Römling U, Fislage R, Tümmler B (1996) Macrorestriction mapping and analysis of bacterial genomes. In: Birren B, Lai E (eds) Nonmammalian genomic analysis—A practical guide. Academic Press, San Diego, pp 165–195

Rudolph K (1990) Toxins as taxonomic features. In: Klement Z, et al. (eds) Methods in phytobacteriology. Akadimeai Kiado, Budapest, pp 251–267

Rudolph K (1995) *Pseudomonas syringae* pathovars. In: Singh US, Singh RP, Kohmoto K (eds) Pathogenesis and host specificity in plant diseases. Elsevier Science, Oxford, UK, pp 47–138

Ruepp A, Muller HN, Lottspeich F, Soppa J (1995) Catabolic ornithine transcarbamylase of *Halobacterium halobium* (*salinarium*): Purification, characterization, sequence determination, and evolution. J Bacteriol 177:1129–1136

Sato M (1988) *In planta* transfer of the genes for virulence between isolates of *Pseudomonas syringae* pv. *atropurpurea.* Ann Phytopathol Soc Jpn 54:20–24

Sawada H, Ieki H, Matsuda I (1995) PCR detection of Ti and Ri plasmids from phytopathogenic *Agrobacterium* strains. Appl Environ Microbiol 61:828–831

Sawada H, Takeuchi T, Matsuda I (1997) Comparative analysis of *Pseudomonas syringae* pv. actinidiae and pv. phaseolicola based on phaseolotoxin-resistant ornithine carbamoyltransferase gene (*argK*) and 16S-23S rRNA intergenic spacer sequences. Appl Environ Microbiol 63:282–288

Sawada H, Suzuki F, Kanaya S, Matsuda I (1998) Origin and evolution of *argK* distributed among *Pseudomonas syringae* pathovars. In: Program and Abstracts at the 1998 Annual Meeting of the Phytopathological Society of Japan. Phytopathological Society of Japan, Tokyo, p 124

Sawada H, Suzuki F, Matsuda I, Saitou N (1999) Phylogenetic analysis of *Pseudomonas syringae* pathovars suggests the horizontal gene transfer of *argK* and the evolutionary stability of *hrp* gene cluster. J Mol Evol 49:627–644

Sawada H, Kanaya S, Suzuki F, Azegami K, Saitou N (2000a) Horizontal transfer of the *argK-tox* gene cluster from gram-positive bacteria onto *Pseudomonas syringae* genomes. In: Proceedings of International Symposium "Evolution 2000," Annual Meeting of the International Society of Molecular Evolution. Center for Information Biology, National Institute of Genetics, Mishima, Japan, pp 44–45

Sawada H, Tsuda M, Suzuki F, Azegami K (2000b) Detection of transposable region from *argK-tox* cluster of *Pseudomonas syringae* pv. *phaseolicola.* In: Program and Abstracts at the 2000 Annual Meeting of the Phytopathological Society of Japan. Phytopathological Society of Japan, Tokyo, p 119

Sawada H, Kanaya S, Tsuda M, Suzuki F, Tsuchiya K, Saitou N (2001) Evolutionary history of OCTase genes on *Pseudomonas syringae* genomes. In: Proceedings of International Symposium "Evolution-

ary Genomics." Center for Information Biology, National Institute of Genetics, Mishima, Japan, p 50

Schurr MJ, Vickrey JF, Kumar AP, Campbell AL, Cunin R, Benjamin RC, Shanley MS, O'Donovan GA (1995) Aspartate transcarbamoylase genes of *Pseudomonas putida:* Requirement for an inactive dihydroorotase for assembly into the dodecameric holoenzyme. J Bacteriol 177:1751–1759

Stalon V, Mercenier A (1984) L-arginine utilization by *Pseudomonas* species. J Gen Microbiol 130:69–76

Stalon V, Legrain C, Wiame JM (1977) Anabolic ornithine carbamoyltransferase of *Pseudomonas:* The bases of its functional specialization. Eur J Biochem 74:319–327

Staskawicz BJ, Panopoulos NJ, Hoogenraad NJ (1980) Phaseolotoxin-insensitive ornithine carbamoyltransferase of *Pseudomonas syringae* pv. *phaseolicola:* Basis for immunity to phaseolotoxin. J Bacteriol 142:720–723

Stibitz S, Yang M (1999) Genomic plasticity in natural populations of *Bordetella pertussis.* J Bacteriol 181:5512–5515

Templeton MD, Sullivan PA, Shepherd MG (1986) Phaseolotoxin-insensitive L-ornithine transcarbamoylase from *Pseudomonas syringae* pv. *phaseolicola.* Physiol Mol Plant Pathol 29:393–403

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673–4680

Thong KL, Puthucheary SD, Pang T (1997) Genome size variation among recent human isolates of *Salmonella typhi.* Res Microbiol 148:229–235

Tsuda M, Tan HM, Nishi A, Furukawa K (1999) Mobile catabolic genes in bacteria. J Biosci Bioeng 87:401–410

Van Vliet F, Cunin R, Jacobs A, Piette J, Gigot D, Lauwereys M, Pierard A, Glansdorff N (1984) Evolutionary divergence of genes for ornithine and aspartate carbamoyl-transferases—Complete sequence and mode of regulation of the *Escherichia coli argF* gene; Comparison of *argF* with *argI* and *pyrB.* Nucleic Acids Res 12: 6277–6289

Ventura L, Perez-Gonzalez JA, Ramon D (1997) Cloning and molecular analysis of the *Aspergillus terreus arg1* gene coding for an ornithine carbamoyltransferase. FEMS Microbiol Lett 149:207–212

Watanabe H, Mori H, Itoh T, Gojobori T (1997) Genome plasticity as a paradigm of Eubacteria evolution. J Mol Evol 44:S57–S64

Young JM, Takikawa Y, Gardan L, Stead DE (1992) Changing concepts in the taxonomy of plant pathogenic bacteria. Annu Rev Phytopathol 30:67–105

Zhang YX, Patil SS (1997) The *phtE* locus in the phaseolotoxin gene cluster has ORFs with homologies to genes encoding amino acid transferases, the AraC family of transcriptional factors, and fatty acid desaturases. Mol Plant–Microbe Interact 10:947–960