J Mol Evol (1999) 49:627–644

JOURNAL OF MOLECULAR EVOLUTION

Phylogenetic Analysis of *Pseudomonas syringae* Pathovars Suggests the Horizontal Gene Transfer of *argK* and the Evolutionary Stability of *hrp* Gene Cluster

Hiroyuki Sawada,[1] Fumihiko Suzuki,[1] Izumi Matsuda,[2] Naruya Saitou[3]

[1] National Institute of Agro-Environmental Sciences, 3-1-1 Kannondai, Tsukuba, Ibaraki 305-8604, Japan
[2] National Agriculture Research Center, Tsukuba, Ibaraki 305, Japan
[3] National Institute of Genetics, Mishima, Shizuoka 411-8540, Japan

Received: 18 January 1999 / Accepted: 25 May 1999

Abstract. *Pseudomonas syringae* are differentiated into approximately 50 pathovars with different plant pathogenicities and host specificities. To understand its pathogenicity differentiation and the evolutionary mechanisms of pathogenicity-related genes, phylogenetic analyses were conducted using 56 strains belonging to 19 pathovars. *gyrB* and *rpoD* were adopted as the index genes to determine the course of bacterial genome evolution, and *hrpL* and *hrpS* were selected as the representatives of the pathogenicity-related genes located on the genome (chromosome). Based on these data, NJ, MP, and ML phylogenetic trees were constructed, and thus 3 trees for each gene and 12 gene trees in total were obtained, all of which showed three distinct monophyletic groups: Groups 1, 2 and 3. The observation that the same set of OTUs constitute each group in all four genes suggests that these genes had not experienced any intergroup horizontal gene transfer within *P. syringae* but have been stable on and evolved along with the *P. syringae* genome. These four index genes were then compared with another pathogenicity-related gene, *argK* (the phaseolotoxin-resistant ornithine carbamoyltransferase gene, which exists within the *argK–tox* gene cluster). All 13 strains of pv. *phaseolicola* and pv. *actinidiae* used had been confirmed to produce phaseolotoxin and to have *argK*, whose sequences were completely identical, without a single synonymous substitution among the strains used (Sawada et al. 1997a). On the other hand, *argK* were not present on the genomes of the other 43 strains used other than pv. *actinidiae* and pv. *phaseolicola.* Thus, the productivity of phaseolotoxin and the possession of the *argK* gene were shown at only two points on the phylogenetic tree: Group 1 (pv. *actinidiae*) and Group 3 (pv. *phaseolicola*). A *t* test between these two pathovars for the synonymous distances of *argK* and the tandemly combined sequence of the four index genes showed a high significance, suggesting that the *argK* gene (or *argK–tox* gene cluster) experienced horizontal gene transfer and expanded its distribution over two pathovars after the pathovars had separated, thus showing a base substitution pattern extremely different from that of the noncluster region of the genome.

Key words: *argK-tox* gene cluster — *hrp* gene cluster — *gyrB* — *rpoD* — Molecular phylogeny — *Pseudomonas syringae* — Horizontal gene transfer

Introduction

Pseudomonas syringae is genetically diverse and is now subclassified into approximately 50 pathovars according to plant pathogenicity and host range (Bradbury 1986; Dye et al. 1980; Rudolph 1995). However, such a pathovar system does not always conform with DNA homology or physiological and biochemical characteristics and may be hindering the analyses of pathogenicity differen-

Correspondence to: H. Sawada; *e-mail:* sawada@niaes.affrc.go.jp

tiation and diversity of *P. syringae* (Bradbury 1986; Rudolph 1995; Young et al. 1992).

To understand the pathogenicity differentiation of *P. syringae*, it is necessary to study not only the phenotypic expression of pathogenicity but also the genes that code for the phenotype. In general, plant pathogenic bacteria must have various factors that are related to host plant affinity and directly involved in destroying plant tissues to intrude into and infect plants and cause disease. Thus, the number of pathogenicity-related genes should be enormous (Baker et al. 1997; Daniels et al. 1988; Gopalan and He 1996; Rich and Willis 1997; Rudolph 1995). Moreover, there are data suggesting that dynamic evolutionary mechanisms such as horizontal gene transfer and genome rearrangement are involved in the diversification of pathogenicity-related genes existing on the *P. syringae* genome ("genome" refers to "bacterial chromosome" in this paper) and have accelerated and complicated the pathogenicity differentiation (Hatziloukas and Panopoulos 1992; Sawada et al. 1995a, 1996, 1997a, b; Sawada, unpublished data). Therefore, to understand the complicated pathogenicity differentiation of *P. syringae*, we should systematically analyze (1) what pathogenicity-related genes are present and functioning in each strain, (2) how each gene has mutated and differentiated, (3) whether the gene has experienced dynamic evolutionary changes such as horizontal gene transfer and genome rearrangement, (4) what its origin and molecular mechanisms of its dynamic evolution are, and (5) what new gene combinations have appeared on the genome through the dynamic evolution.

Based on this viewpoint, we started to investigate the diversity of *P. syringae* by comparing its genes. We first selected the 16S–23S rRNA intergenic spacer sequences in the *rrn* operon as the index and discovered that *P. syringae* pathovars have diverse spacer sequences (Sawada et al. 1995a, 1996, 1997a). At the same time, we also found that their 16S–23S spacers are not appropriate as an index for phylogenetic analyses since the insertion and deletion of several to dozens nucleotides frequently occur in the 16S–23S spacer (Sawada et al. 1995a, 1996, 1997a; Sawada, unpublished data). Thus, a protein-coding gene in which base substitution is a basis of mutation and horizontal gene transfer seldom takes place should be found and used as a new index.

The objectives of this study are to analyze (1) the evolutionary course of the *P. syringae* genome, (2) when and which pathogenicity-related genes arose or were introduced by horizontal gene transfer on each genome, and (3) how those genes mutated and differentiated, using the newly adopted index genes. The index genes adopted here to determine the course of *P. syringae* genome evolution were *gyrB* (the DNA gyrase B subunit gene) (Huang 1996; Yamamoto and Harayama 1995; Yamamoto and Harayama 1996) and *rpoD* (the primary sigma factor gene, or the group 1 σ^{70}-type sigma factor gene) (Gruber and Bryant 1997; Lonetto et al. 1992), which are indispensable single-copy genes on which horizontal gene transfer seldom occurs, and widely accepted indices for phylogenetic analyses. We used 56 strains belonging to 19 pathovars, including their pathotype strains, and sequenced the two adopted index genes to clarify the course of genome evolution. Then we selected *hrpL* and *hrpS* as the representatives of the pathogenicity-related genes, which exist in all *P. syringae* pathovars and are believed to be involved in the basic process of determining the pathogenicity and host range (Baker et al. 1997; Gopalan and He 1996), and compared the course of genome evolution determined from *gyrB* and *rpoD* genes with the data on *hrpL* and *hrpS* genes. We also compared the data of these four genes with those of another pathogenicity-related gene, *argK* [the phaseolotoxin-resistant ornithine carbamoyltransferase (ROCT) gene, which exists within the *argK–tox* gene cluster], whose sequence we analyzed previously (Sawada et al. 1995a, 1997a; Sawada, unpublished data).

The study revealed that these *P. syringae* pathovars used differentiated into three monophyletic groups, the *argK* gene (*argK–tox* gene cluster) had been distributed in two pathovars through horizontal gene transfer, and the *hrp* gene cluster (*hrpL* and *hrpS*) had always been stable on the genome and never underwent horizontal gene transfer between groups within *P. syringae*. Preliminary reports of this study have been presented orally at the annual meetings of the Phytopathological Society of Japan (Sawada et al. 1995a, 1996, 1997b).

Materials and Methods

Bacterial Strains

Fifty-six strains belonging to 19 *P. syringae* pathovars, including their pathotype strains, were used (Table 1). The sources of the strains used and their relevant characteristics are also shown in Table 1. Taxonomic positions of *P. syringae* strains were confirmed by identifying their phenotypic features according to standard methods (Hildebrand et al. 1988; Takikawa et al. 1989).

Gene Amplification and Nucleotide Sequencing

Template preparation, PCR amplification, and nucleotide sequencing were performed according to the methods described previously (Sawada et al. 1995b, 1997a).

Oligonucleotide primers amplifying the target regions were designed based on the published sequence: for *gyrB* (gyr-F, 5′-CGCCAGGGTTTTCCCAGTCACGACCMGGCGGYAAGTTCGATGACAAYTC-3′; and gyr-R, 5′-TTTCACACAGGAAACAGCTATGACTRATBKCAGTCARACCTTCRCGSGC-3′); for *rpoD* (rpo-F, 5′-CGCCAGGGTTTTCCCAGTCACGACAAGGCGARATCGAAATCGCCAAGCG-3′; and rpo-R, 5′-TTTCACACAGGAAACAGCTATGACGGAACWKGC-

Table 1. Strains of *Pseudomonas syringae* used and data bank accession numbers for DNA sequences determined in this study

Taxon (pathovar)	OTU name[a]	Strain designation[b]	Species name of host plant (common name)	Geographic origin[c]	Source[b]	DDBJ accession No.[d]			
						gyrB	*rpoD*	*hrpL*	*hrpS*
pv. *aceris*	AR1	MAFF 302273^{PT}	*Acer* sp. (maple)	USA	MAFF	AB016299	AB016300	AB016301	AB016302
pv. *actinidiae*	AC30	KW–11^{PT}	*Actinidia chinensis* (kiwi fruit)	Shizuoka, Japan	Y. Takikawa	AB016303	AB016304	AB016305	AB016306
	AC30	MAFF 302091	*Actinidia chinensis* (kiwi fruit)	Kanagawa, Japan	MAFF	AB016303	AB016304	AB016305	AB016306
	AC30	MAFF 302092	*Actinidia chinensis* (kiwi fruit)	Kanagawa, Japan	MAFF	AB016303	AB016304	AB016305	AB016306
	AC30	MAFF 302145	*Actinidia chinensis* (kiwi fruit)	Wakayama, Japan	MAFF	AB016303	AB016304	AB016305	AB016306
	AC30	MAFF 302133	*Actinidia arguta* (tara vine)	Kanagawa, Japan	MAFF	AB016303	AB016304	AB016305	AB016306
pv. *aptata*	AP1	MAFF 302253^{PT}	*Beta vulgaris* var. *saccharifera* (sugarbeet)	USA	MAFF	AB016307	AB016308	AB016309	AB016310
pv. *broussonetiae*	BR1	KOZ 8101^{PT}	*Broussonetia kazinoki* x *B. papyrifera* (paper mulberry)	Tottori, Japan	M. Sato	AB016311	AB016312	AB016313	AB016314
pv. *castaneae*	CA1	C–3^{PT}	*Castanea crenata* (chestnut)	Shiga, Japan	K. Takanashi	AB016315	AB016316	AB016317	AB016318
pv. *eriobotryae*	ER1	MAFF 302259^{PT}	*Eriobotrya japonica* (loquat)	USA	MAFF	AB016319	AB016320	AB016321	AB016322
pv. *glycinea*	GL1	MAFF 302260^{PT}	*Glycine max* subsp. *max* (soybean)	NZ	MAFF	AB016323	AB016324	AB016325	AB016326
	GL1	MAFF 301683	*Glycine max* subsp. *max* (soybean)	Ibaraki, Japan	MAFF	AB016323	AB016324	AB016325	AB016326
pv. *japonica*	JA1	MAFF 301072^{PT}	*Hordeum vulgare* (barley)	Tochigi, Japan	MAFF	AB016327	AB016328	AB016329	AB016330
pv. *lachrymans*	LA1	MAFF 302278^{PT}	*Cucumis sativus* (cucumber)	USA	MAFF	AB016331	AB016332	AB016333	AB016334
	LA2	MAFF 301315	*Cucumis sativus* (cucumber)	Gunma, Japan	MAFF	AB016335	AB016336	AB016337	AB016338
	LA2	MAFF 301321	*Cucumis sativus* (cucumber)	Chiba, Japan	MAFF	AB016335	AB016336	AB016337	AB016338
pv. *maculicola*	MA1	MAFF 302264^{PT}	*Brassica oleracea* var. *botrytis* (cauliflower)	NZ	MAFF	AB016339	AB016340	AB016341	AB016342
	MA2	MAFF 301174	*Brassica pekinensis*	Kanagawa, Japan	MAFF	AB016343	AB016344	AB016345	AB016346
	MA2	H–1	*Brassica campestris* (Chinese cabbage)	?[c]	M. Sato	AB016343	AB016344	AB016345	AB016346
	MA2	H 3–6	*Brassica campestris* (Chinese cabbage)	?	M. Sato	AB016343	AB016344	AB016345	AB016346
	MA2	H 3–9	*Brassica campestris* (Chinese cabbage)	?	M. Sato	AB016343	AB016344	AB016345	AB016346
pv. *mori*	MR1	MAFF 302279^{PT}	*Morus alba* (mulberry)	Hungary	MAFF	AB016347	AB016348	AB016349	AB016350
	MR2	MAFF 301020	*Morus alba* (mulberry)	Nagano, Japan	MAFF	AB016351	AB016352	AB016353	AB016354
	MR6	S 6801	*Morus alba* (mulberry)	Gunma, Japan	M. Sato	AB016355	AB016356	AB016357	AB016358
	MR6	S 6806	*Morus alba* (mulberry)	Tokyo, Japan	M. Sato	AB016355	AB016356	AB016357	AB016358
	MR6	S 7133–2 (Sm2)	*Morus alba* (mulberry)	Shimane, Japan	M. Sato	AB016355	AB016356	AB016357	AB016358
	MR6	S 7233–1 (Sm4)	*Morus alba* (mulberry)	Shimane, Japan	M. Sato	AB016355	AB016356	AB016357	AB016358
pv. *morsprunorum*	MP1	MAFF 302280^{PT}	*Prunus domestica* (European plum)	?	MAFF	AB016359	AB016360	AB016361	AB016362
	MP2	MAFF 301436	*Prunus mume* (mume, Japanese apricot)	Shiga, Japan	MAFF	AB016363	AB016364	AB016365	AB016366
	MP3	W 8010	*Prunus salicina* (plums)	Aomori, Japan	K. Takanashi	AB016367	AB016368	AB016369	AB016370
pv. *myricae*	MY1	MAFF 301464^{PT}	*Myrica rubra* (myrica)	Okinawa, Japan	MAFF	AB016371	AB016372	AB016373	AB016374
pv. *phaseolicola*	PA1	MAFF 302282^{PT}	*Phaseolus vulgaris* (kidney bean)	Canada	MAFF	AB016375	AB016376	AB016377	AB016378
	PA1	MAFF 301673	*Phaseolus vulgaris* (kidney bean)	Ibaraki, Japan	MAFF	AB016375	AB016376	AB016377	AB016378
	PA1	MAFF 301022	*Phaseolus vulgaris* (kidney bean)	Hokkaido, Japan	MAFF	AB016375	AB016376	AB016377	AB016378
	PA1	MAFF 301616	*Phaseolus vulgaris* (kidney bean)	Hokkaido, Japan	MAFF	AB016375	AB016376	AB016377	AB016378
	PA1	TMR 327	*Pueraria lobata* (kudzu vine)	Shizuoka, Japan	Y. Takikawa	AB016375	AB016376	AB016377	AB016378
	PA1	MAFF 301766	*Pueraria lobata* (kudzu vine)	Ibaraki, Japan	MAFF	AB016375	AB016376	AB016377	AB016378
	PA1	KUZ 8	*Pueraria lobata* (kudzu vine)	Yamagata, Japan	M. Sato	AB016375	AB016376	AB016377	AB016378
	PU4	KUZ 1	*Pueraria lobata* (kudzu vine)	Ibaraki, Japan	M. Sato	AB016379	AB016380	AB016381	AB016382
pv. *pisi*	PI1	MAFF 302269^{PT}	*Pisum sativum* (pea)	NZ	MAFF	AB016383	AB016384	AB016385	AB016386
	PI2	MAFF 301208	*Pisum sativum* (pea)	Shizuoka, Japan	MAFF	AB016387	AB016388	AB016389	AB016390

Table 1. Continued

Taxon (pathovar)	OTU name[a]	Strain designation[b]	Species name of host plant (common name)	Geographic origin[c]	Source[b]	DDBJ accession No.[d]			
						gyrB	*rpoD*	*hrpL*	*hrpS*
pv. *syringae*	SY1	MAFF 302155^{T}	*Syringa vulgaris* (common lilac)	UK	MAFF	AB016391	AB016392	AB016393	AB016394
	SY2	MAFF 301861	*Syringa vulgaris* (common lilac)	Nagano, Japan	MAFF	AB016395	AB016396	AB016397	AB016398
	SY2	LOB 2–5	*Syringa vulgaris* (common lilac)	Nagano, Japan	Y. Takikawa	AB016395	AB016396	AB016397	AB016398
	SY5	W 7835	*Prunus persica* (peach)	Fukushima, Japan	K. Takanashi	AB016399	AB016400	AB016401	AB016402
	SY7	T–1	*Citrus iyo* (Iyo)	Yamaguchi, Japan	S. Shigeta	AB016403	AB016404	AB016405	AB016406
pv. *tabaci*	TB1	MAFF 302270^{PT}	*Nicotiana tabacum* (tobacco)	Hungary	MAFF	AB016407	AB016408	AB016409	AB016410
	TB1	MAFF 301612	*Nicotiana tabacum* (tobacco)	Akita, Japan	MAFF	AB016407	AB016408	AB016409	AB016410
	TB1	PT–13	*Nicotiana tabacum* (tobacco)	?	T. Miyoshi	AB016407	AB016408	AB016409	AB016410
pv. *theae*	TH3	PT1^{PT}	*Camellia sinensis* (tea)	Shizuoka, Japan	Y. Takikawa	AB016411	AB016412	AB016413	AB016414
	TH3	S 2	*Camellia sinensis* (tea)	Shizuoka, Japan	Y. Nishi	AB016411	AB016412	AB016413	AB016414
	TH3	Tea 632	*Camellia sinensis* (tea)	Shizuoka, Japan	Y. Nishi	AB016411	AB016412	AB016413	AB016414
	TH2	K 9301	*Camellia sinensis* (tea)	Kagoshima, Japan	Y. Nishi	AB016415	AB016416	AB016417	AB016418
	TH2	K 9302	*Camellia sinensis* (tea)	Kagoshima, Japan	Y. Nishi	AB016415	AB016416	AB016417	AB016418
	TH2	1	*Camellia sinensis* (tea)	Kagoshima, Japan	Y. Nishi	AB016415	AB016416	AB016417	AB016418
pv. *tomato*	TO1	MAFF 302272^{PT}	*Lycopersicon esculentum* (tomato)	UK	MAFF	AB016419	AB016420	AB016421	AB016422

[a] If two strains had one nonidentical gene sequence, both strains were kept for further analyses and given the respective OTU names regardless of whether the other three gene sequences were identical. If strains showed all four gene sequences to be identical, one strain was selected from one pathovar as an OTU. In this way, we selected 31 strains of 56 as OTUs, gave each strain the OTU name, and used the OTU names in the figures, tables, and text.

[b] MAFF—Microorganisms Section of the MAFF Gene Bank, National Institute of Agrobiological Resources, Tsukuba, Ibaraki 305-8602, Japan; T. Miyoshi—Takanori Miyoshi, Ehime Fruit Tree Experiment Station, Japan; Y. Nishi—Yatsuka Nishi, Kagoshima Tea Experiment Station, Japan; M. Sato—Mamoru Sato, National Institute of Sericultural and Entomological Science, Japan; S. Shigeta—Susumu Shigeta, Yamaguchi Prefectural Nishiki Agricultural Extension Center, Japan; K. Takanashi—Kazuo Takanashi, Hokkou Chemical Co., Ltd., Japan; Y. Takikawa—Yuichi Takikawa, Faculty of Agriculture, Shizuoka University, Japan. PT, pathotype strain; T, type strain.

[c] ?, information was not available.

[d] All the listed sequences were determined in this study. The same sequences within each OTU are represented by one accession number.

GCAGGAAGTCGGCACG-3′); for *hrpL* (hrpL-F, 5′-CGCCAGGGTTTTCCCAGTCACGACTTTTGGCTGGCAYGGTTATCGCTATA-3′; and hrpL-R, 5′-TTTCACACAGGAAACAGCTATGACTGTGGTTTTGCGTGCGAGTTGGTTCC-3′); and for *hrpS* (hrpS-F, 5′-CGCCAGGGTTTTCCCAGTCACGACCTSCAGGCCAAGCTGCTGAGGGTGC-3′; and hrpS-R, 5′-TTTCACACAGGAAACAGCTATGACTTGAGCTCRCGGATATTGCCGGGCC-3′).

Both the forward (F) and the reverse (R) primers contained the sequences of Cy5-labeled primers (ALFred M13-40 and ALFred M13 Reversal) for cycle sequencing reactions, which are indicated by underscores (the sequence of primer ALFred M13-40 was added to forward primers, and that of primer ALFred M13 Reversal to reverse primers). Both strands were sequenced directly by cycle sequencing, using these Cy5-labeled primers.

DNA Sequence Alignments

Raw sequence data were analyzed and compiled using DNASIS-Mac version 3.6 (Hitachi Software Engineering) and MacClade version 3.06 (Maddison and Maddison 1996) softwares, and sequence alignments were facilitated using CLUSTAL W version 1.7 (Thompson et al. 1994). Several typical DNA sequences were selected from raw data and converted into amino acid sequences, and multiple alignments among these were inferred. These results and the published alignments data were used as templates to conduct multiple alignments of all the DNA sequences determined and to correct them manually. Functional informations on the gene products were also used to aid in the multiple alignments. The multiple alignments are available in machine-readable form. For information, e-mail sawada@niaes.affrc.go.jp. Positions where gaps are present in any one of the aligned sequences were excluded from the following analysis.

Phylogenetic Analysis

Phylogenetic trees were constructed using three tree-building methods: the neighbor-joining (NJ) method (Saitou and Nei 1987), the maximum-parsimony (MP) method (Fitch 1977), and the maximum-likelihood (ML) method (Felsenstein 1981).

Evolutionary distances (number of base substitutions) were estimated using the one-parameter (Jukes and Cantor 1969) and two-parameter (Kimura 1980) methods, and these distances were used for constructing NJ trees using MEGA version 1.0 (Kumar et al. 1993), CLUSTAL W, and PHYLIP 3.572c (Felsenstein 1996). Nonsynonymous and synonymous substitutions per site were calculated by the method of Nei and Gojobori (1986) using MEGA.

PAUP version 3.1.1 (Swofford 1993) was used for MP analysis, and heuristic and branch-and-bound searches were used to ensure finding the most-parsimonious trees. When two or more parsimonious trees were obtained, we constructed a strict consensus tree and a 50% majority-rule consensus tree.

In looking for the ML tree topology, the local rearrangement searches of NucML and ProtML (contained in the program package MOLPHY version 2.3) (Adachi and Hasegawa 1996) were carried out, starting from the NJ tree topology and/or the topology obtained by the quick add OTUs search option as the initial trees. The HKY85 model (Hasegawa et al. 1985) was used for the base substitution process, and the JTT-F model ("F" option of the JTT model) (Cao et al. 1994) for the amino acid substitution process. When the heterogeneity in rate was detected among the three codon positions of protein-coding genes, a realistic model distinguishing among rates at different codon positions, namely, the rate-heterogeneous model (Hasegawa and Adachi 1996), was adopted. When a single gene did not contain sufficient phylogenetic information to resolve the problem at hand, we collectively used four index genes (*gyrB, rpoD, hrpL,* and *hrpS*), evaluated the total support for a particular tree topology by summing up the estimated log-likelihoods of individual genes for that topology, and compared the total log-likelihoods for different topologies, using TotalML (Adachi and Hasegawa 1996).

Reliability Test for Inferred Tree Topology

To evaluate the reliability of the inferred tree topology, the bootstrap probability (Felsenstein 1985) was calculated for trees inferred by the NJ and MP methods by repeating the bootstrap resampling procedure 1000 or 10,000 times. For the ML method, the local bootstrap probability (LBP) was calculated using the RELL method (Adachi and Hasegawa 1996; Hasegawa and Kishino 1994; Kishino et al. 1990) with 1000 replications.

DNA Sequence Accession Numbers

The DNA sequences which we determined in this study have been deposited in the DDBJ/EMBL/GenBank international nucleotide sequence database under the accession numbers shown in Table 1. The same sequences within each OTU are represented by one accession number.

Results

Sequence Determination and Alignment

To confirm that the *gyrB* and *rpoD* genes, which were selected as the indices for determining the evolutionary course of the *P. syringae* genomes, are actually single copies on bacterial genomes, we studied the databases of five types of eubacteria [*Haemophilus influenzae, Mycoplasma genitalium, Mycoplasma pneumoniae, Synechocystis* sp. (PCC 6803), *Escherichia coli* K-12] for which the sequences of the entire genomes are determined and found that there was only one copy identified or estimated to be *gyrB* or *rpoD* on the genome of each bacterium (data not shown). A Southern analysis also confirmed that these two genes are both single copies on the *P. syringae* genomes used in this study (Sawada, unpublished data). Moreover, the GC contents of these genes, their GC contents at the third codon positions, and the frequency of codon usage showed that both genes had characteristics similar to those of other genes existing on the *P. syringae* genome (Sawada, unpublished data). Hence, both genes likely have never experienced horizontal gene transfer or gene duplication but had evolved on and along with the *P. syringae* genome. Therefore, we concluded that *gyrB* and *rpoD* are appropriate indices for analyzing the evolutionary course of *P. syringae* genomes.

gyrB. We determined the partial sequences of *gyrB* genes (612 bp) for 56 strains, which belong to 19 pathovars of *P. syringae* (Table 1). The sequence determined corresponds from position 1501 to position 2115 on *E. coli* K12 numbering (accession number X04341). We

used 615 bp of the *gyrB* sequences of *P. aeruginosa* (Yamamoto, personal communication) and *E. coli* (accession number X04341), which correspond to the *P. syringae gyrB* partial sequences determined in this study, as outgroups (OTU names: PAE and ECO, respectively). (Unless stated otherwise, sequence data are expressed in this paper with the OTU names shown in Table 1.)

As a preliminary test, two sequences (BR1 and PAE) were converted into amino acid sequences with DNASIS and were aligned with ClustalW. Then, we conducted multiple alignment for all the DNA sequences of *gyrB* used in this study with ClustalW and corrected the results manually based on the amino acid sequence alignment of BR1 and PAE.

rpoD. We determined the *rpoD* partial sequences of 56 strains used in this study (Table 1). For most strains, the sequence of 516 bp in the coding region was determined, but we obtained data on 522 bp for pv. *eriobotryae* (ER1) due to inserted nucleotides in this region. The sequences determined correspond to positions 3362–3874 of *E. coli* K-12 numbering (accession number J01687) and cover most of the "nonconserved insertion" located between region 1 and region 2 of RpoD (Lonetto et al. 1992). Since this section is less conserved than the other regions of *rpoD,* we expected that it would be suitable for the index of intraspecific comparison. The *rpoD* sequences of *P. aeruginosa* (accession number D90118) and *E. coli* (J01687) of 519 and 513 bp, respectively, which correspond to the *P. syringae* data determined here, were used as outgroups (OTU names: PAE and ECO, respectively).

As a preliminary test, five sequences (AC30, ER1, BR1, PAE, and ECO) were converted into amino acid sequences and were multiple aligned. Then we conducted multiple alignment for all the DNA sequences of *rpoD* used in this study, and corrected the results manually based on the amino acid sequence alignment obtained by the preliminary test and on that reported by Tanaka and Takahashi (1991).

hrpL. We determined the sequence of the entire *hrpL* coding region (555 bp) for 56 strains used. The sequence of the entire *algU* coding region of *P. aeruginosa* (582 bp; U49151) and the sequence of the entire *rpoE* coding region of *E. coli* (576 bp; U37089) were used as outgroups. AlgU, RpoE, and HrpL are all alternative sigma factors of the ECF subfamily that belongs to the group 3 σ^{70}-type sigma factor (Lonetto et al. 1994). After multiple alignment of these DNA sequences, we corrected the results manually based on the amino acid sequence alignments reported by Martin et al. (1994) and Wei and Beer (1995).

hrpS. We determined the partial sequence (240 bp) of the *hrpS* coding region for 56 strains used. This 240-bp region corresponds to positions 577–816 of the *P. syringae* pv. *phaseolicola* numbering (M28524) and constitutes a part of domain D of prokaryotic regulatory proteins that interact with σ^{54}-RNA polymerase holoenzyme (Drummond et al. 1986; Grimm and Panopoulos 1989). We used 240 bp of *P. aeruginosa fleR* (L41213) and *E. coli tyrR* (M12114), which correspond to the *P. syringae hrpS* partial sequences determined here, as outgroups (all of these code for prokaryotic regulatory proteins).

Selection of OTUs. We compared the sequences of 56 strains and found strains that have identical sequences. If two strains had one nonidentical gene sequence, both the strains were kept for further analyses regardless of whether the other three gene sequences were identical. If strains showed all four gene sequences being identical, one strain was selected from one pathovar as an OTU. In this way, we selected 31 strains of 56 as OTUs, gave each strain the OTU name shown in Table 1, and used them for further analyses with two outgroups (ECO and PAE).

Since the object of this study, *P. syringae,* is one species, the analyses were intraspecific comparison. Moreover, evolutionary distances between any two OTUs of *P. syringae* should be very small (Figs. 1 and 2). Thus, the saturation of base substitution should scarcely affect the analyses. In addition, we determined the base composition of each gene and confirmed that the compositions did not much vary among *P. syringae* OTUs (data not shown). Therefore, discussions in the following sections are based mainly on DNA sequence analyses and use amino acid sequences for reference.

Analyses of Each Gene by Three Phylogenetic Tree Drawing Methods

gyrB. We conducted NJ bootstrap analysis of finally determined *gyrB* sequence alignments, each of 612 bp, from which the insertion/deletion sites had been removed. The resulting NJ tree showed three large groups of *P. syringae* (Fig. 1): Group 1, which contained 9 OTUs (TO1, LA1, MA1, MA2, SY7, MP1, TH2, TH3, AC30); Group 2, with 8 OTUs (AR1, SY1, SY2, SY5, JA1, AP1, PI1, PI2); and Group 3, with 14 OTUs (GL1, PA1, PU4, MR1, MR2, MR6, BR1, MP2, MP3, MY1, TB1, ER1, CA1, LA2). These groups were formed with bootstrap probabilities of 100, 100, and 75.9%, respectively. The outgroups (ECO and PAE) formed a root between Group 1 and the other two groups, and thus the topology that shows the group branching order was (1,(2,3)) (Fig. 1, Table 2). The bootstrap probability for determining the group branching order was relatively high (94.5%).

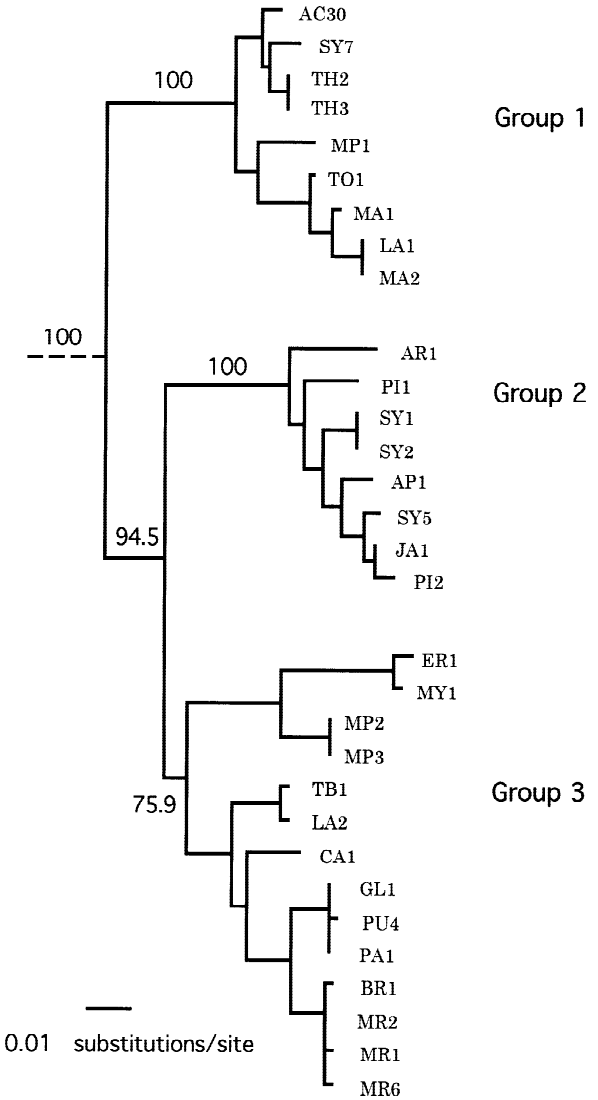

Fig. 1. NJ tree expressing the evolution of the *P. syringae gyrB* gene, based on 612 bp of the *gyrB* gene sequence, from which sites that include gaps in more than one sequence had been excluded, constructed using the distance matrix derived from Kimura's (1980) two-parameter method (see text for details). The tree was rooted using *gyrB* sequences of *P. aeruginosa* (Yamamoto, personal communication) and *E. coli* (X04341) as the outgroups. Horizontal branch lengths are proportional to the estimated number of nucleotide substitutions, and bootstrap probabilities (as percentages), as determined for 1000 resamplings, are given *above* or *beside* the internal branches. Abbreviations for strains (OTU names) are as listed in Table 1.

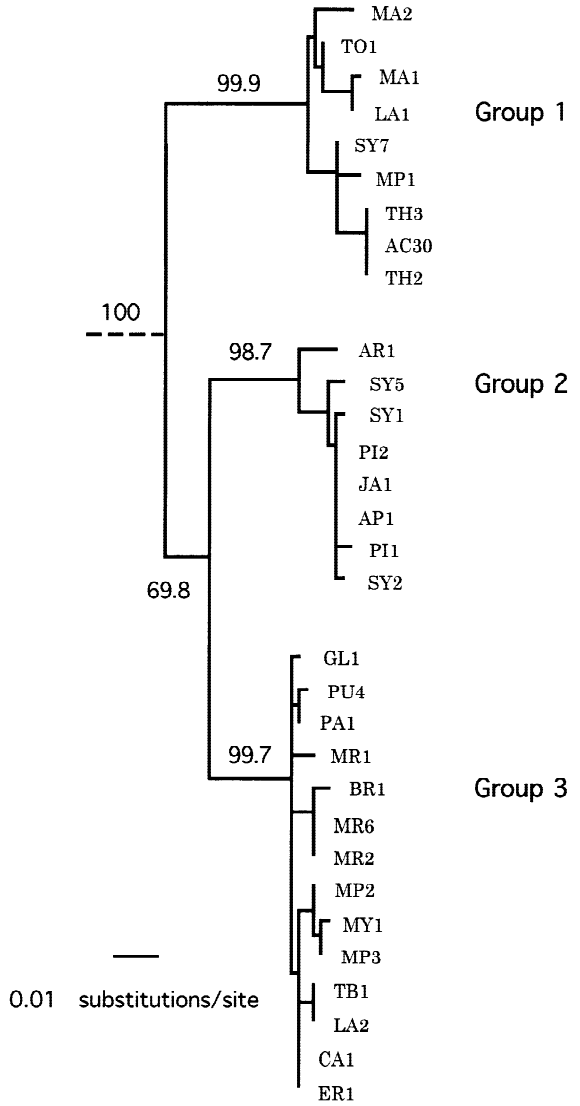

Fig. 2. NJ tree expressing the evolution of the *P. syringae rpoD* gene, based on 510 bp of the *rpoD* gene sequence (see text and legend to Fig. 1 for details). The tree was rooted using *rpoD* sequences of *P. aeruginosa* (D90118) and *E. coli* (J01687) as the outgroups. Horizontal branch lengths and bootstrap probabilities are as described in the legend to Fig. 1. Abbreviations for strains (OTU names) are as listed in Table 1.

The topology of the ML tree also showed three groups, Groups 1, 2, and 3 (data not shown). The OTUs forming each group were identical to those of the NJ tree. The branching order of these groups was also identical to that of the NJ tree and was (1,(2,3)) (Table 2). However, the local bootstrap probability determining the branching order was lower (77.6%).

The MP tree also showed three groups (data not shown). The OTUs constituting each group and the branching order of the three groups were identical to the NJ and ML results, but the bootstrap probability determining the branching order was lower (65.3%) (Table 2).

rpoD. We analyzed *rpoD* sequence alignments each of 510 bp, from which insertion/deletion sites had been removed, using the NJ and MP methods. Like the *gyrB* analyses, these analyses also showed three groups (Fig. 2). The OTUs constituting each group and the branching order of the three groups were identical to the *gyrB* results (Table 2). The bootstrap probabilities for the branching orders were both low, 69.8 and 73.0%, respectively.

The ML tree also showed three groups and OTU constitutions identical to those of the NJ and MP trees (data not shown). However, the branching order of these three groups, (2,(1,3)), was different (Table 2). The local bootstrap probability determining the branching order was low, 72.8%.

Table 2. Tree topology (group branching order) obtained by three phylogenetic tree drawing methods

Sequence	Method	Tree topology (group branching order)[a] (3,(1,2))	(2,(1,3))	(1,(2,3))	Bootstrap probability (%)[b]
gyrB	NJ			X	94.5
	ML			X	77.6
	MP[c]			X	65.3
rpoD	NJ			X	69.8
	ML		X		72.8
	MP[c]			X	73.0
hrpL	NJ			X	53.4
	ML		X		55.1
	MP[c]		X		88.1
hrpS	NJ			X	54.0
	ML		X		51.6
	MP[c]			X	55.4
Tandemly combined sequence of all 4 genes	NJ			X	86.9
	ML		X		67.0
	MP[c]		X		60.9

[a] Group 1 contained 9 OTUs (TO1, LA1, MA1, MA2, SY7, MP1, TH2, TH3, AC30); Group 2 contained 8 OTUs (AR1, SY1, SY2, SY5, JA1, AP1, PI1, PI2); and Group 3 contained 14 OTUs (GL1, PA1, PU4, MR1, MR2, MR6, BR1, MP2, MP3, MY1, TB1, ER1, CA1, LA2) (see Figs. 1–4).
[b] The bootstrap probability that determines the group branching order. For the ML method, the local bootstrap probability (LBP) was calculated using the RELL method (Hasegawa and Kishino 1994; Kishino et al. 1990) with 1000 replications.
[c] When two or more parsimonious trees were obtained, the topology of a strict consensus tree was used.

hrpL. The NJ, ML, and MP trees determined from the *hrpL* data, each of 555 bp, from which gaps had been eliminated, showed the same three groups as the above analyses (data not shown). The OTUs constituting each group were also identical to the above results. The branching order of the groups was (1,(2,3)) in the NJ tree but was (2,(1,3)) in the ML and MP trees (Table 2). The bootstrap probabilities for the orders were all low, 53.4, 55.1, and 88.1%, respectively.

hrpS. All the NJ, ML, and MP trees showed three groups as in the other three genes (data not shown). The OTUs constituting each group were also identical to the other results. The branching order of the groups was (1,(2,3)) in the NJ and MP methods and (2,(1,3)) in the ML method (Table 2). The bootstrap probabilities for the orders were all low, 54.0, 55.4, and 51.6%.

Confirmation of Monophyly of the Three Groups that Appeared in the Gene Trees

All 12 phylogenetic trees (gene trees) described in the previous section showed an identical grouping pattern of 31 *P. syringae* OTUs: all trees showed three groups (Groups 1, 2, and 3). We confirmed the monophyly of each group by checking each gene.

Confirmation by Bootstrap Probability. The bootstrap probabilities, which show the unity of each group, were all relatively low (87.4–91.2%) for *rpoD* in the MP method (Table 3). The *gyrB* results also showed relatively low unity of Group 3 (54.9–81.0%). However, the other combinations of genes and methods all demonstrated that each group (Groups 1, 2, and 3) is united with a high bootstrap probability (Table 3).

Table 3. Bootstrap probability that shows the unity of each group

Sequence	Method	Bootstrap probability (%)[a] Group 1	Group 2	Group 3
gyrB	NJ	100.0	100.0	75.9
	ML[b]	96.1	100.0	81.0
	MP	93.5	99.4	54.9
rpoD	NJ	99.9	98.7	99.7
	ML[b]	100.0	99.4	99.4
	MP	89.7	91.2	87.4
hrpL	NJ	98.6	98.1	100.0
	ML[b]	100.0	97.0	100.0
	MP	100.0	99.9	100.0
hrpS	NJ	99.2	99.4	100.0
	ML[b]	100.0	98.5	100.0
	MP	100.0	99.9	100.0
Tandemly combined sequence of all 4 genes	NJ	100.0	100.0	100.0
	ML[b]	100.0	98.6	100.0
	MP	100.0	100.0	100.0

[a] The bootstrap probability showing the unity of each group, which appeared in the gene trees (Figs. 1 and 2).
[b] For the ML method, the local bootstrap probability (LBP) was estimated by the RELL method (Hasegawa and Kishino 1994; Kishino et al. 1990) with 1000 replications.

Table 4. Average synonymous distance within each group and between groups[a]

Sequence	Average distance within each group			Average distance between groups		
	Group 1	Group 2	Group 3	Group 1–Group 2	Group 1–Group 3	Group 2–Group 3
gyrB	0.123	0.098	0.199	0.575	0.490	0.420
rpoD	0.063	0.029	0.033	0.438	0.344	0.219
hrpL	0.052	0.048	0.016	0.999	0.856	0.875
hrpS	0.125	0.087	0.028	1.179	0.979	0.888
Tandemly combined sequence of all 4 genes	0.086	0.063	0.075	0.690	0.581	0.506

[a] Synonymous substitutions were estimated by the method of Nei and Gojobori (1986) using MEGA.

Confirmation by Synonymous Distances. The mean synonymous distance of *gyrB* was relatively large (0.199) within Group 3 members, but those of the other combinations of genes and groups were all small within each group (Table 4). On the contrary, the mean synonymous distances among groups were all very large. Therefore, the variation within a group was always much smaller than those between groups, except for Group 3 in *gyrB.*

Monophyly of the Three Groups. As for Group 3 of the *gyrB* gene, members may have started branching earlier than the other groups (Fig. 1), and thus Group 3 may be less united (Tables 3 and 4). Nevertheless, the OTUs constituting Group 3 of *gyrB* are all identical to those of Group 3 appeared in the other three gene trees (Figs. 1 and 2).

We therefore concluded that all three groups that had been confirmed by each gene are monophyletic and that these four genes of *P. syringae* have differentiated into three groups in the course of evolution. The observation that the same set of OTUs constitutes each group in all four genes (Figs. 1 and 2) suggests that these genes had not experienced any intergroup horizontal gene transfer within *P. syringae* but have been stable on and evolved along with the genome. Therefore, we concluded that there is no need to separate and compare the pathogenicity-related *hrpL* and *hrpS* genes from the housekeeping *gyrB* and *rpoD* genes, which had been selected for comparison with pathogenicity-related genes, but decided to use all four of these genes collectively as an integrated index for studying the evolutionary course of the *P. syringae* genome in the following analyses.

Study of the Branching Order

We could not specify the branching order of the three groups since the 12 gene trees obtained topologically showed no (3,(1,2)) but both (2,(1,3)) and (1,(2,3)) at a ratio of 4:8 (Table 2). The bootstrap probabilities for determining the branching orders were all low except for the NJ analysis for *gyrB* (94.5%). These results suggest that such analyses based on each gene are insufficient for determining the branching order, and we conducted other analyses.

Drawing Phylogenetic Trees of the Tandemly Combined Sequence (Concatenated Sequence). As described in the previous section, the four index genes had not experienced any intergroup horizontal gene transfer, and the variation within a group was always smaller than the difference between groups (Tables 3 and 4). Therefore, investigation of the group branching order should not cause a serious error even by integrating the sequence data of the four index genes into one. We drew phylogenetic trees of the tandemly combined sequence using three methods (NJ, ML, and MP) and compared them.

For each OTU, we integrated its *gyrB, rpoD, hrpL,* and *hrpS* data into one set and conducted an NJ bootstrap analysis. As in the analyses of each gene, three groups were shown, all with 100% bootstrap probability (Table 3). The OTUs constituting each group were all identical to those determined by the analyses of each gene. The branching order of the groups was (1,(2,3)), with a bootstrap probability of 86.9% (Table 2). On the other hand, the ML tree for the tandemly combined sequence showed a different branching order, (2,(1,3)), with a low local bootstrap probability of 67.0% (Table 2). The MP tree also showed a branching order of (2,(1,3)) but, again, with a low bootstrap probability of 60.9%.

The analyses with tandemly combined sequences did not give uniform results or high bootstrap probabilities or could not determine whether the branching order of the three groups was (2,(1,3)) or (1,(2,3)).

ML Analysis Based on More Realistic Models and Parameters. In the second method, we thoroughly investigated the characteristics of the data of the four index genes, selected models and parameters appropriate for each datum, analyzed each gene with the ML method, and comprehensively evaluated the four results so obtained. In other words, we comprehensively evaluated the four data sets using the TotalML of Molphy version 2.3 (Adachi and Hasegawa 1996) and compared the log-likelihoods among the three topologies concerning the group branching order.

Table 5. Base composition and α/β ratio for each codon position of the four index genes

Sequence	Codon position	Length[a]	G + C[b]	α/β[c]
gyrB	1st	203	0.56	3.7
	2nd	203	0.36	
	3rd	203	0.69	8.6
	Total	609	0.54	
rpoD	1st	170	0.73	2.2
	2nd	170	0.40	1.7
	3rd	170	0.61	5.8
	Total	510	0.58	
hrpL	1st	183	0.62	2.1
	2nd	183	0.36	1.6
	3rd	183	0.66	4.7
	Total	549	0.54	
hrpS	1st	80	0.65	1.9
	2nd	80	0.42	1.9
	3rd	80	0.71	3.5
	Total	240	0.60	
Tandemly combined sequence of all 4 genes	1st	636	0.63	2.2
	2nd	636	0.38	1.8
	3rd	636	0.66	5.7
	Total	1908	0.56	

[a] Insertion/deletion sites were excluded.
[b] Mean values of 31 *P. syringae* OTUs. The base composition for each codon position did not much vary among *P. syringae* OTUs.
[c] The α/β ratios were optimized for each codon position using NucML (HKY 85 model and rate-heterogeneous model). Because of the small number of substitutions, ML estimates cannot be obtained for the second codon position of *gyrB.*

(a) *Selection of Models Based on Characteristics of Sequence Data.* We determined the base compositions of each gene at three codon positions and found that the composition bias varied by codon position (Table 5). The optimum α/β ratios also varied by codon position, with the α/β ratio of the second codon position being especially low (Table 5). This phenomenon, which had been known for the mitochondrial genome (Hasegawa and Adachi 1996; Yang 1996), was shown to occur in bacterial genomes in this study. Since any gene is heterogeneous in terms of codon positions, we used the rate-heterogeneous model (Hasegawa and Adachi 1996), in which the α/β ratio, branch length, and log-likelihood are separately estimated for each of the three codon positions and the topologies are evaluated by adding the three log-likelihood values obtained. As for the base substitution model, we adopted the HKY85 model (Hasegawa et al. 1985), since the base composition is biased to a certain base and the optimum α/β ratio is not 1 at all codon positions (Table 5).

(b) *ML Analysis Using a Rate-Heterogeneous Model.* The analyses based on bootstrap probability (Table 3) and synonymous substitution distance (Table 4) showed large differences between groups and small variation within a group, with each group being very homogeneous. Thus, we created a user tree for determining the branching order not of the members constituting each group but of the groups themselves only. Using this user tree, we compared the log-likelihoods of three topologies, (3,(1,2)), (2,(1,3)), and (1,(2,3)), with TotalML and conducted a comprehensive evaluation of the group branching order.

For *gyrB,* a comprehensive evaluation of the analyses of each codon position resulted in an ML tree of (2,(1,3)) (Table 6). A *rpoD* analysis also showed (2,(1,3)) to be the most likely. A comprehensive evaluation of these two data sets also supported (2,(1,3)) most. The ML trees for *hrpL* and *hrpS* were also (2,(1,3)) (Table 6). A comprehensive evaluation of *hrpL* and *hrpS* data also supported (2,(1,3)).

A comprehensive evaluation of these four genes strongly supported (2,(1,3)) (Table 6). On the other hand, the topology of (3,(1,2)) was rejected at the 5% level of significance, with a log-likelihood lower by 12.6 ± 6.2 (0.87% BP). The (1,(2,3)) topology was not rejected, although its bootstrap probability was very low, 9.9%. A rate-heterogeneous model analysis of the tandemly combined sequence, in which the sequence data of the four index genes were concatenated, also strongly supported (2,(1,3)) and rejected (3,(1,2)), with a log-likelihood lower by 14.5 ± 6.8 (0.29% BP). (Table 6).

An amino acid sequence analysis with the JTT-F model using the ProtML and TotalML programs (Adachi and Hasegawa 1996) also showed (3,(1,2)) to have the worst likelihood and the lowest bootstrap probability (Table 7). (2,(1,3)) was comprehensively best in likelihood, but the likelihood difference with (1,(2,3)) was not as large as that shown by the DNA sequence analyses (Table 6).

Branching Order of the Three Groups. Among the three topologies for the group branching order, (3,(1,2)) was statistically rejected by the TotalML analysis of DNA sequences (Table 6). None of the analyses conducted for each gene supported this topology either (Table 2). Therefore, (3,(1,2)) should be excluded from the branching order candidates. On the other hand, (2,(1,3)) was most supported by both the DNA sequence and the amino acid sequence analyses (Tables 6 and 7). The order (1,(2,3)) showed a worse log-likelihood than (2,(1,3)) but could not be totally excluded (Tables 6 and 7). We concluded that the group branching order of *P. syringae* should not be determined as (2,(1,3)) or (1,(2,3)) with only the data used in this study. The difficulty may be attributable to these three groups having continuously branched within a very short period in the evolutionary time scale and/or to the inadequacy of the selection of the outgroups. More *P. syringae* relatives and new indices should be analyzed to accumulate data, and analytical methods should be further investigated to determine the branching order.

Table 6. Comprehensive evaluation of tree topologies (group branching order) based on NucML and TotalML analyses of the four index genes

Sequence	Codon position	Length	(3,(1,2))		(2,(1,3))		(1,(2,3))	
			Δl[a]	BP[b]	Δl[a]	BP[b]	Δl[a]	BP[b]
gyrB[c]	1st	207	−1.0 ± 1.7	0.0121	⟨−825.7⟩	0.6567	−1.0 ± 1.7	0.3312
	3rd	207	⟨−2488.5⟩	0.4742	−0.7 ± 1.4	0.1210	−0.2 ± 1.8	0.4048
	Total	414	−0.3 ± 1.7	0.3184	⟨−3314.8⟩	0.4313	−0.5 ± 2.5	0.2503
rpoD	1st	175	⟨−596.7⟩	0.8274	−1.7 ± 1.7	0.0840	−1.7 ± 1.7	0.0886
	2nd	175	−2.7 ± 2.6	0.0568	⟨−546.0⟩	0.8666	−2.7 ± 2.6	0.0766
	3rd	175	−4.0 ± 3.2	0.0273	⟨−1324.0⟩	0.6547	−1.8 ± 4.2	0.3180
	Total	525	−5.0 ± 4.1	0.0778	⟨−2468.4⟩	0.7711	−4.5 ± 5.2	0.1511
Total for *gyrB* and *rpoD*		939	−4.6 ± 4.4	0.1282	⟨−5783.5⟩	0.7289	−4.7 ± 5.8	0.1429
hrpL	1st	194	−3.0 ± 1.8	0.0176	⟨−897.0⟩	0.8014	−1.8 ± 2.1	0.1810
	2nd	194	−0.2 ± 0.6	0.2757	⟨−709.3⟩	0.5985	−0.2 ± 0.6	0.1258
	3rd	194	−0.3 ± 1.1	0.3393	−0.4 ± 0.9	0.1036	⟨−1478.1⟩	0.5571
	Total	582	−3.1 ± 2.2	0.0278	⟨−3084.8⟩	0.7218	−1.6 ± 2.2	0.2504
hrpS	1st	80	−1.9 ± 2.1	0.0682	⟨−512.2⟩	0.7984	−1.9 ± 2.1	0.1334
	2nd	80	−2.9 ± 2.4	0.0314	⟨−376.3⟩	0.7417	−2.1 ± 2.9	0.2269
	3rd	80	−1.2 ± 2.1	0.1924	−1.3 ± 2.0	0.1307	⟨−870.7⟩	0.6769
	Total	240	−4.7 ± 3.8	0.0144	⟨−1760.6⟩	0.7290	−2.6 ± 3.6	0.2566
Total for *hrpL* and *hrpS*		822	−7.8 ± 4.4	0.0034	⟨−4845.4⟩	0.8169	−4.2 ± 4.2	0.1797
Total for all 4 genes		1761	−12.6 ± 6.2[d]	0.0087	⟨−10628.9⟩	0.8926	−8.9 ± 7.2	0.0987
Tandemly combined sequence of all 4 genes	1st	656	−8.1 ± 4.8	0.0112	⟨−3032.5⟩	0.8355	−5.5 ± 5.5	0.1533
	2nd	656	−5.1 ± 3.7	0.0203	⟨−2230.9⟩	0.8677	−4.4 ± 4.0	0.1120
	3rd	656	−1.4 ± 2.9	0.1743	⟨−6548.3⟩	0.5203	−0.8 ± 3.1	0.3054
	Total	1968	−14.5 ± 6.8[d]	0.0029	⟨−11811.7⟩	0.9261	−10.7 ± 7.5	0.0710

[a] The log-likelihood of the ML tree is given in angle braces, and the differences in log-likelihood of alternative topologies from that of the ML tree (Δl) are shown with their SEs (following ±), which were estimated by Kishino and Hasegawa's formula (Adachi and Hasegawa 1996).
[b] The bootstrap probabilities (BP) were estimated by the RELL method (Hasegawa and Kishino 1994; Kishino et al. 1990) with 10,000 replications.
[c] Because of the small number of substitutions, ML estimates cannot be obtained for the second codon position of *gyrB*, and this position was not used in the analysis.
[d] The relationship of (3,(1,2)) was rejected at the 5% level of significance.

Table 7. Comprehensive evaluation of tree topologies (group branching order) based on ProtML (JTT-F model) and TotalML analyses of the four index genes

Sequence	Length	(3,(1,2))		(2,(1,3))		(1,(2,3))	
		Δl[a]	BP[b]	Δl[a]	BP[b]	Δl[a]	BP[b]
gyrB	207	−2.6 ± 2.6	0.0478	⟨−1182.9⟩	0.5998	−1.2 ± 3.0	0.3524
rpoD	175	−3.4 ± 3.1	0.0505	−1.9 ± 3.8	0.3050	⟨−1207.3⟩	0.6445
Total for *gyrB* and *rpoD*	382	−4.8 ± 4.0	0.0110	−0.7 ± 3.8	0.4325	⟨−2391.4⟩	0.5565
hrpL	193	−3.0 ± 1.6	0.0040	⟨−1610.5⟩	0.9724	−2.9 ± 1.7	0.0236
hrpS	80	−2.2 ± 2.4	0.0398	−0.8 ± 3.6	0.3874	⟨−898.9⟩	0.5728
Total for *hrpL* and *hrpS*	273	−4.4 ± 2.9	0.0021	⟨−2510.2⟩	0.7137	−2.2 ± 1.7	0.2842
Total for all 4 genes	655	−8.4 ± 5.0	0.0001	⟨−4902.3⟩	0.5862	−1.4 ± 3.5	0.4137
Combined for all 4 genes	655	−8.7 ± 5.0	0.0003	⟨−5183.9⟩	0.7695	−4.9 ± 6.6	0.2302

[a] The log-likelihood of the ML tree is given in angle braces, and the differences in log-likelihood of alternative topologies from that of the ML tree (Δl) are shown with their SEs (following ±), which were estimated by Kishino and Hasegawa's formula (Adachi and Hasegawa 1996).
[b] The bootstrap probabilities (BP) were estimated by the RELL method (Hasegawa and Kishino 1994; Kishino et al. 1990) with 10,000 replications.

Construction of a Phylogenetic Tree Expressing the Genome Evolution of P. syringae

We attempted to construct a phylogenetic tree (genome tree) that expresses the genome evolution of *P. syringae* by collectively using the data on all four index genes. We used the tandemly combined sequence of the four genes as indices based on the reasons described in the previous section. Since we could not determine the group branching order with the analyses described in the previous

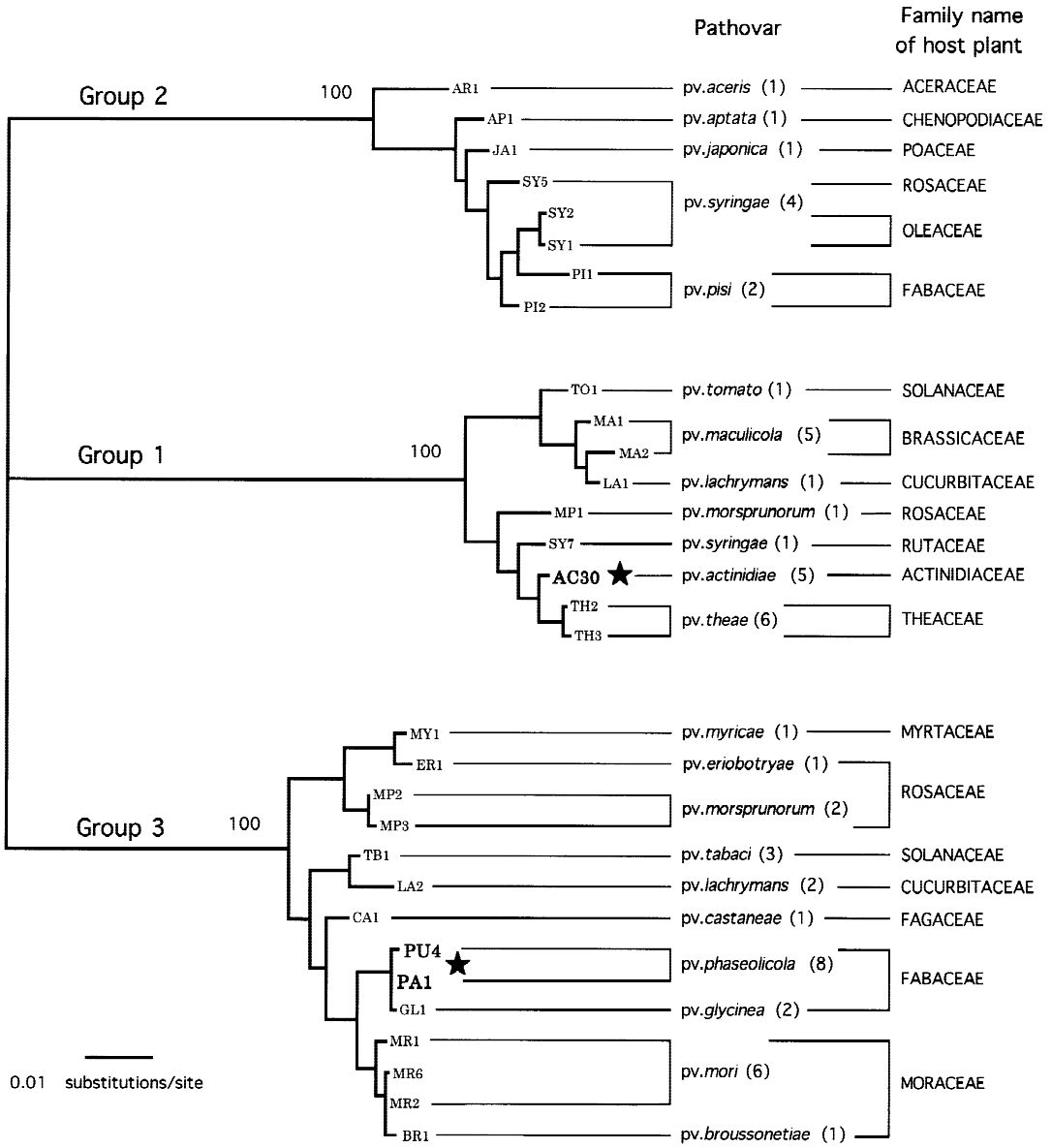

Fig. 3. ML tree expressing the evolution of the *P. syringae* genome, based on the tandemly combined sequence (concatenated sequence) of the four index genes (*gyrB, rpoD, hrpL,* and *hrpS*) of 31 *P. syringae* OTUs excluding outgroups (PAE and ECO), constructed using NucML [HKY85 model (Hasegawa et al. 1985); $\alpha/\beta = 4.5$, $\ln L = -6694.01$] (see text for details). The group branching order of this tree is either (2,(1,3)) or (1,(2,3)), with (2,(1,3)) being much more likely (see Tables 6 and 7 and the text). An identical topology was found by MP analysis. Horizontal branch lengths are proportional to the estimated number of nucleotide substitutions, and local bootstrap probabilities (as percentages), as determined for 1000 resamplings, are given *above* the internal branches. Abbreviations for strains (OTU names) are as listed in Table 1. *Stars* indicate the strains which are capable of producing phaseolotoxin and have the identical *argK* gene on their genomes. Each strain is labeled with the pathovar name to which it belongs and the family name of the host plant from which it was isolated. The *number in parentheses* after each pathovar name indicates the number of strains used which belong to that pathovar and are included in that clade.

sections [although (2,(1,3)) is most likely], we excluded the data on outgroups, *E. coli* (ECO) and *P. aeruginosa* (PAE), and drew an unrooted tree with only *P. syringae* data.

The resulting ML tree clearly showed distinction among Groups 1, 2, and 3, all with 100% bootstrap probability (Fig. 3). The topologies within each group were all identical to those of the ML tree determined from the tandemly combined sequences of 33 OTUs, including the outgroups (ECO and PAE), which is described in the previous section. The MP tree also showed clear separation of Groups 1, 2, and 3, all with 100% bootstrap probability (data not shown). The topologies within each group were all identical to those of the ML tree (Fig. 3). The NJ tree also showed Groups 1, 2, and 3, all with 100% bootstrap probability (Fig. 4). The topologies within each group were all identical to those of the NJ tree determined from 33 OTUs, including the outgroups (see the previous section). This correspondence is likely attributable to each group being constituted by a suffi-

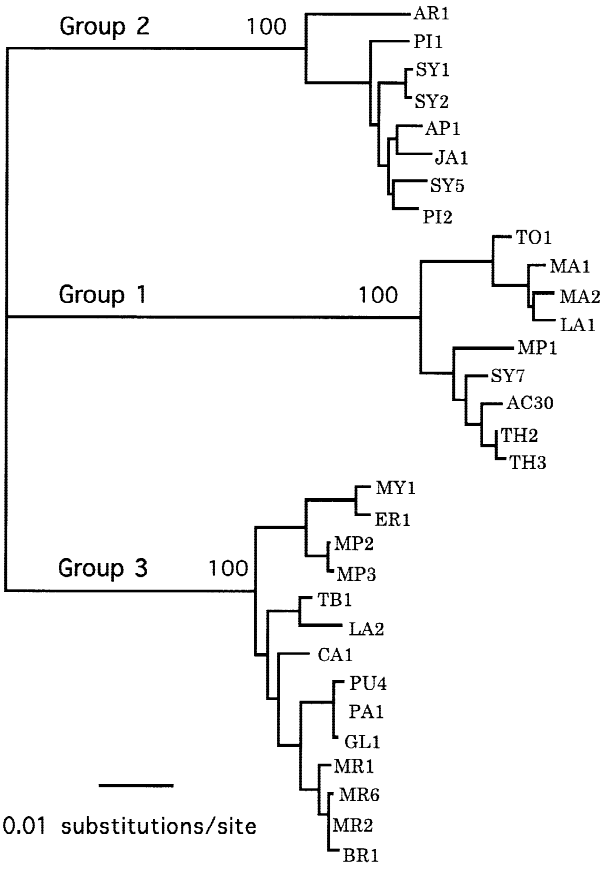

Fig. 4. NJ tree expressing the evolution of the *P. syringae* genome, based on the tandemly combined sequence of the four index genes, constructed using the distance matrix derived from the two-parameter method (Kimura 1980) (see the text and the legend to Fig. 3 for details). The group branching order of this tree is either (2,(1,3)) or (1,(2,3)), with (2,(1,3)) being much more likely (see Tables 6 and 7 and text). The branching order of the OTUs shallower than AR1 in Group 2 is slightly different from those of the MP and ML trees (Fig. 3). Horizontal branch lengths and bootstrap probabilities are as described in the legend to Fig. 1. Abbreviations for strains (OTU names) are as listed in Table 1.

cient number of OTUs (around 10), thus stabilizing the topological estimation. A comparison of newly determined NJ, ML, and MP trees (Figs. 3 and 4) showed a slight difference of OTU branching order between NJ tree and the others in Group 2 but totally identical topologies concerning Groups 1 and 3. This topological difference in Group 2 was insignificant for the major conclusions of this study.

Discussion

With consideration of the results obtained, we collectively used the data on all four index genes (*gyrB*, *rpoD*, *hrpL*, and *hrpS*) and created a genome tree that expresses the evolutionary course of the genome in the final section under Results (Figs. 3 and 4). Since we could not determine the group branching order (see Results), we constructed an unrooted tree with *P. syringae* data (31 OTUs) only and assumed a tentative root for the unrooted tree based on the likeliest (2,(1,3)) topology. Information concerning various phenotypes and pathogenicity-related genes was added to the genome tree obtained (Fig. 3), whose evolutionary mechanisms in *P. syringae* are discussed below.

Relationship with the Pathovar to Which a Strain Belongs

We located the pathovar name to which each strain belongs in the genome tree (Fig. 3). Pathovar *tomato* (OTU name of the corresponding strains: TO1), pv. *maculicola* (MA1, MA2), pv. *lachrymans* (LA1), pv. *morsprunorum* (MP1), pv. *syringae* (SY7), pv. *actinidiae* (AC30), and pv. *theae* (TH2, TH3) were located in Group 1; pv. *aceris* (AR1), pv. *aptata* (AP1), pv. *japonica* (JA1), pv. *syringae* (SY1, SY2, SY5), pv. *pisi* (PI1, PI2) were in Group 2; and pv. *myricae* (MY1), pv. *eriobotryae* (ER1), pv. *morsprunorum* (MP2, MP3), pv. *tabaci* (TB1), pv. *lachrymans* (LA2), pv. *castaneae* (CA1), pv. *phaseolicola* (PA1, PU4), pv. *glycinea* (GL1), pv. *mori* (MR1, MR2, MR6), and pv. *broussonetiae* (BR1) were in Group 3. A phylogenetic analysis based on the *hrpJKL*-region sequence (Cournoyer et al. 1996) had confirmed that pv. *pisi* and pv. *syringae* are close relatives and agreed with the results of this study (Figs. 3 and 4). PCR-RFLP analysis of the *rrn* operon (Manceau and Horvais 1997) and AFLP and RAPD analyses (Clerc et al. 1998) had confirmed the close relationship between pv. *tomato* and pv. *maculicola*, also agreeing with the results of this study (Figs. 3 and 4).

Three pathovars, pv. *lachrymans*, pv. *morsprunorum*, and pv. *syringae*, are distributed over two groups (Figs. 3 and 4). The pathotype strains of pv. *lachrymans* (LA1) and pv. *morsprunorum* (MP1) are included in Group 1, but all the other OTUs (LA2 for pv. *lachrymans*, MP2 and MP3 for pv. *morsprunorum*) representing Japanese strains used (Table 1) belong to Group 3. For pv. *syringae*, only a Japanese citrus strain (SY7) (Shigeta and Nakata 1995) belongs to Group 1, while all the other OTUs, representing the pathotype strain (SY1) and another three Japanese strains (SY2 and SY5), are included in Group 2.

The cause for these three pathovars being distributed over two groups and being genetically nonuniform may be associated with incorrect or incomplete pathovar definition or strain identification (Bradbury 1986; Rudolph 1995; Young et al. 1992). Various indices suggest that pv. *syringae* is especially heterogeneous and is a collection of multiple kinds of strains (Cameron 1962; Denny et al. 1988; Legard et al. 1993; Young 1991; Young et al. 1992), which may represent different pathovars or even different species (Young 1991). On the other hand, it is also suggested that nine pathovars (pv. *aceris*, pv. *aptata*, pv. *atrofaciens*, pv. *dysoxyli*, pv. *japonica*, pv. *lapsa*, pv. *panici*, pv. *papulans*, and pv. *pisi*) may be synonyms of

pv. *syringae* (Gardan et al. 1991, 1994; Young 1992; Young et al. 1992). Of these nine pathovars, this study tested four [pv. *aceris* (AR1), pv. *aptata* (AP1), pv. *japonica* (JA1), and pv. *pisi* (PI1, PI2)] and revealed that all of them and pv. *syringae* (SY1, SY2, SY5) formed one tight cluster, Group 2 (Figs. 3 and 4). Our results agree with studies based on the DNA–DNA hybridization technique (Gardan et al. 1994) and may provide gene-level support for the synonymy of these pathovars. As shown above, the pathovar system of *P. syringae* is complicated and confused in terms of classification (Bradbury 1986; Rudolph 1995; Young et al. 1992). Thus, the classification of *P. syringae* should be carefully investigated by testing many strains of diverse histories and by conducting and cross-checking gene-level analyses and phenotype studies.

Of the 16 pathovars other than pv. *lachrymans*, pv. *morsprunorum*, and pv. *syringae*, 8 were tested for two or more strains (Table 1); pv. *maculicola* (number of strains tested, five; OTU names, MA1 and MA2), pv. *actinidiae* (five; AC30), pv. *theae* (six; TH2 and TH3), pv. *pisi* (two; PI1 and PI2), pv. *phaseolicola* (eight; PA1 and PU4), pv. *mori* (six; MR1, MR2, and MR6), pv. *tabaci* (three; TB1), and pv. *glycinea* (two; GL1). As far as these eight pathovars were concerned, all strains that belong to the same pathovar were always closely located in the genome tree (Figs. 3 and 4), suggesting their homogeneity. Pathovar *actinidiae* is especially homogeneous as we reported previously (Sawada et al. 1995a, 1996, 1997a), since the sequences of the four index genes investigated in this study were identical in all five strains tested.

Relationship Between Strains and Plant Sources (Host Plants)

We then located the family names of host plants, from which the strains were isolated, in the genome tree (Fig. 3). Group 1 includes many diverse host plant families, such as *Solanaceae* (OTU name of the corresponding strains: TO1), *Brassicaceae* (MA1 and MA2), *Cucurbitaceae* (LA1), *Rosaceae* (MP1), *Rutaceae* (SY7), *Actinidiaceae* (AC30), and *Theaceae* (TH2 and TH3), and showed no special characteristics or correlation between the phylogenetic relationship of OTUs and that of their host plants.

The host range of pv. *syringae* included in Group 2 is considered to be very wide, including *Pisum sativum*, *Triticum aestivum*, and *Beta vulgaris* (Bradbury 1986; Young 1991), which are also hosts for pv. *pisi*, pv. *japonica*, and pv. *aptata* of the same group, respectively (Bradbury 1986). Thus, Group 2 members appear to have common host plants with pv. *syringae*, but this interpretation must be carefully examined by further studying the definition of pv. *syringae* as discussed in the previous section.

In Group 3, pv. *mori* (MR1, MR2, MR6) and pv. *broussonetiae* (BR1), which infect *Moraceae* plants, are closely located, as are pv. *phaseolicola* (PA1, PU4) and pv. *glycinea* (GL1) from *Fabaceae* and pv. *morsprunorum* (MP2, MP3) and pv. *eryobotryae* (ER1) from *Rosaceae*, respectively (Figs. 3 and 4). The three strains of pv. *tabaci* (TB1) tested were from *Nicotiana tabacum*, whose host range also includes *Cucumis sativus* (Bradbury 1986), and are located in a tree near pv. *lachrymans* (LA2) isolated from *Cucumis sativus*. Therefore, we may conclude that in Group 3 there are four subgroups, whose members were isolated from host plants that are closely related to each other.

Groups 2 and 3 showed a correlation between the phylogenetic relationships of OTUs and those of their host plants. To study these relationships in terms of coevolution, we should further elucidate the phylogenetic relationships of host plants and the molecular mechanisms of pathogenicity.

Evolutionary Stability of the hrp *Gene Cluster Within* P. syringae

The *hrp/avr* genes are considered to exist in all *P. syringae* pathovars and to be involved in the pathogenicity and the determination of host range (reviewed by Baker et al. 1997; Gopalan and He 1996). Among these genes, the ones that are involved in the fundamental processes such as gene expression control and secretion of gene products form a supraoperon clustering structure (the so-called "*hrp* gene cluster") on the genome. The cluster is believed to contain at least 25 genes that are organized into seven separate transcription units. The *hrpL* and *hrpS* genes that were selected to represent pathogenicity-related genes in this study are located near the left and right ends of the *hrp* gene cluster, respectively, and are involved in the expression control of the *hrp/avr* genes. To investigate the evolutionary mechanism of the *hrp* gene cluster, we compared the phylogenetic analyses for the inside of the *hrp* gene cluster (*hrpL* and *hrpS*; representing the cluster) and those for the outside section (*gyrB* and *rpoD*; representing the genome).

NJ, ML, and MP analyses showed that the *gyrB* and *rpoD* genes located outside of the cluster had both differentiated into three independent groups, as had the *hrpL* and *hrpS* genes located in the cluster (Tables 2, 3, and 4). The OTUs constituting each group were identical for genes in and out of the cluster. Total ML analysis with a rate-heterogeneous model showed the highest log-likelihood when the branching order of the three groups was (2,(1,3)) for both *gyrB* and *rpoD* (Table 6). Similarly, (2,(1,3)) was the ML tree for both *hrpL* and *hrpS*. Therefore, the data for both the outside and the inside of the cluster showed the same tendency. We, thus, concluded that the *hrp* gene cluster of the strains tested in this study had not experienced any intergroup horizontal gene transfer within *P. syringae*, and the *hrp* gene cluster and the *P. syringae* genome, on which the cluster is

located, likely followed the same course of evolution. It was recently postulated that *hrp* genes may have been introduced into the ancestor of *P. syringae* from an unknown enteric animal pathogen by horizontal gene transfer (Brown et al. 1998). At least after that event, they should have been stable on the *P. syringae* genome.

Pathogenicity-related genes, whose expression is controlled and/or whose products are secreted outside of the bacteria by the function of *hrp* genes, were found also to exist on non-*hrp* gene cluster sections of the genome and plasmids (Baker et al. 1997; Gopalan and He 1996; Yuan and He 1996). Thus, to understand fully the evolution of pathogenicity associated with the *hrp* gene cluster, we should systematically analyze not only the so-called core *hrp* gene cluster itself, which is the regulation and secretion system, but also the data for various pathogenicity-related genes under its control.

Expansion of the argK (argK–tox *Gene Cluster*) *Distribution in* P. syringae

Phaseolotoxin is a non-host-specific toxin that is active not only in plants but in microorganisms including *E. coli* (reviewed by Rudolph 1990). The enzyme that is the target of the toxin is anabolic ornithine carbamoyltransferase (OCTase; EC 2.1.3.3) in the biosynthetic pathway of arginine. This means, in turn, that the biosynthesis of L-citrulline from L-ornithine and carbamoylphosphate is inhibited by the toxin. This is thought to be the cause of both the accumulation of L-ornithine and the deficiency of L-citrulline and L-arginine (Mitchell and Bieleski 1977; Mitchell 1984; Patil et al. 1970; Rudolph and Stahmann 1966).

Pathovar *phaseolicola*, which produces phaseolotoxin, has both phaseolotoxin-sensitive OCTase (SOCT) and phaseolotoxin-resistant OCTase (ROCT). It is known that pv. *phaseolicola* is resistant to phaseolotoxin produced by itself since ROCT is not inhibited by phaseolotoxin but remains active (Ferguson et al. 1980; Jahn et al. 1985; Staskawicz et al. 1980; Templeton et al. 1986). The gene that codes for ROCT, *argK* (Hatziloukas and Panopoulos 1992; Mosqueda et al. 1990), is contained in the *tox* cluster (Hatziloukas and Panopoulos 1992; Peet, and Panopoulos 1987), which is involved in the biosynthesis of phaseolotoxin (Peet et al. 1986; Zhang et al. 1993).

Phaseolotoxin, which was first discovered as a toxin of pv. *phaseolicola* (Rudolph 1990), is also produced by pv. *actinidiae* and has been proved actually to cause chlorosis in pv. *actinidiae*-infected kiwifruit leaves (Tamura et al. 1989, 1997). We also confirmed that all strains of these two pathovars tested (13 strains that are expressed with three OTU names: AC30, PA1, and PU4) (Table 1) produced phaseolotoxin and had an *argK* gene on their genome, using Southern blot analysis (Sawada et al. 1995a, 1997a; Sawada, manuscript in preparation). On the contrary, we found that the other 43 strains used in this study (Table 1), which do not belong to these two pathovars, were not capable of producing phaseolotoxin and did not have the *argK* gene (Sawada et al. 1995a, 1997a; Sawada, manuscript in preparation). Thus, the productivity of phaseolotoxin and the possession of the *argK* gene are shown at only two separate points in the genome tree (Fig. 3): Group 1 (AC30; pv. *actinidiae*) and Group 3 (PA1 and PU4; pv. *phaseolicola*). It is known that the *argK–tox* gene cluster is located on the genome (Peet et al. 1986; Sawada, manuscript in preparation; Zhang et al. 1993). The following paragraphs discuss the reason for the *argK–tox* gene cluster on the genome, not on a conjugative plasmid, being distributed over two phylogenetic groups (Groups 1 and 3) and the expansion mechanisms of its distribution.

Hatziloukas and Panopoulos (1992) analyzed *argK* of pv. *phaseolicola* isolated from *Phaseolus vulgaris* in the United States and found that the GC content of the *argK* gene and that of the whole genome of pv. *phaseolicola* were very different. Thus, they hypothesized that the *argK* gene originated from another prokaryote. Narumi and Takikawa (1996) compared the *argK–tox* gene cluster structures of pv. *phaseolicola* and pv. *actinidiae* using restriction enzyme maps and found that the structures were similar for the inside section of the cluster but were different for the outside between the two pathovars. We revealed that the 13 strains of pv. *phaseolicola* and pv. *actinidiae* used (Table 1) have *argK*, whose sequences (D86356) are completely identical to each other, without a single synonymous substitution (Sawada et al. 1995a, 1997a; Sawada, manuscript in preparation), also coincide with that (M94049) of pv. *phaseolicola* isolated in the United States (Hatziloukas and Panopoulos 1992) and are quite similar (only three substitutions exist) to that (X55520) of pv. *phaseolicola* isolated in Mexico (Mosqueda et al. 1990). On the other hand, we found that the genomes of pv. *phaseolicola* and pv. *actinidiae* belong to different and separate phylogenetic groups (Groups 3 and 1, respectively) (Tables 3 and 4, Figs. 3 and 4).

Therefore, the following three hypotheses may be derived for the origin and expanding distribution of *argK* or the *argK–tox* gene cluster. (1) *argK* (*argK–tox* gene cluster) was introduced by horizontal gene transfer from the original organism into the genome of a common ancestor of pv. *actinidiae* (Group 1) and pv. *phaseolicola* (Group 3). Afterward, as the ancestor evolved into pv. *actinidiae* and pv. *phaseolicola*, the *argK* gene took the corresponding evolutionary course along with and stably on the genome (no horizontal gene transfer of *argK* occurred after the pathovars separated). (2) After pv. *actinidiae* and pv. *phaseolicola* separated from the common ancestor, *argK* was introduced by horizontal gene transfer from the original organism into either pv. *actinidiae* or pv. *phaseolicola*, from which the gene was introduced by

the second horizontal gene transfer into the other. (3) After pv. *actinidiae* and pv. *phaseolicola* separated, *argK* was introduced from the original organism into pv. *actinidiae* and pv. *phaseolicola* by two independent horizontal gene transfers. We examined these three hypotheses based on the results of this study.

The tandemly combined sequence (index for genome evolution) of *gyrB, rpoD, hrpL,* and *hrpS* was used to derive the synonymous distance and the standard error between AC30 (pv. *actinidiae*) and PA1 (pv. *phaseolicola* isolated from *Phaseolus vulgaris* or *Pueraria lobata*) to be 0.5316 ± 0.0463. The distance between AC30 and PU4 (pv. *phaseolicola* isolated from *Pueraria lobata*) was 0.5308 ± 0.0462, a similar value. On the contrary, there is no synonymous substitution in the *argK* gene between pv. *actinidiae* (AC30) and pv. *phaseolicola* (PA1 and PU4). Thus, a *t* test between pv. *actinidiae* and pv. *phaseolicola* for the synonymous distances of the tandemly combined sequence and *argK* showed $t = 0.53/0.046 = 11.5$, which is highly significant. This result suggests that *argK* of these two pathovars had taken evolutionary courses different from those of the four index genes selected to represent their genomes. This and other analyses of ours (Sawada et al. 1995a, 1997a; Sawada, unpublished data) revealing the lack of *argK* on the genomes of the 43 strains used other than pv. *actinidiae* or pv. *phaseolicola* strongly suggest that the *argK* gene (*argK–tox* gene cluster) experienced horizontal gene transfer after the pathovars had separated, thus showing a base substitution pattern extremely different from that of the noncluster region of the genome. Hence, the first of our three hypotheses mentioned above is rejected. The *argK* gene was likely introduced very recently in terms of evolutionary time scale, either (1) by horizontal gene transfer from the original organism to pv. *actinidiae* or pv. *phaseolicola,* and from each to the other by the second transfer, or (2) by two independent horizontal gene transfers from the original organism to pv. *actinidiae* and pv. *phaseolicola.*

To verify these two hypotheses, we should (1) test more strains of pv. *actinidiae* and pv. *phaseolicola* with diverse histories, (2) compare the DNA sequences of the *tox* cluster other than the *argK* region and the structures of the genome near the cluster boundary, (3) identify the original organisms from which the *argK–tox* gene cluster was derived, and so on. When and how the gigantic "pathogenicity island"-like structure of the *argK–tox* gene cluster was formed is another important topic.

This study revealed that the *P. syringae* strains used differentiated into three different monophyletic groups. Our phylogenetic analysis of the evolutionary mechanisms of *argK* (*argK–tox* gene cluster) showed that it had expanded its distribution over two pathovars by horizontal gene transfer. On the other hand, the *hrp* gene cluster had not experienced horizontal gene transfer between groups of *P. syringae* but had been stable on the genome. We are now conducting studies from other viewpoints based on the results of this study: we will further investigate the origins and evolutionary mechanisms of the *argK–tox* gene cluster and other pathogenicity-related genes by analyzing the frequency of codon usage, comparing the genome structures, and conducting verification tests concerning the evolutionary mechanisms, such as horizontal gene transfer and genome rearrangement.

Acknowledgments. We thank Masami Hasegawa, Satoshi Oota, Naoko Takezaki, Satoshi Yamamoto, and Yuichi Takikawa for their helpful suggestions and Koushi Nishiyama, Mamoru Sato, Yuichi Takikawa, Takanori Miyoshi, Yatsuka Nishi, Susumu Shigeta, and Kazuo Takanashi 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

Baker B, Zambryski P, Staskawicz B, Dinesh-Kumar SP (1997) Signaling in plant-microbe interactions. Science 276:726–733

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

Brown EW, Allard MW, van der Zwet T (1998) Phylogenetic characterization of the eubacterial *lcrD* gene family: Molecular evolutionary aspects of pathogen-induced hypersensitivity in plants. Cladistics 14:45–62

Cameron HR (1962) Diseases of deciduous fruit trees incited by *Pseudomonas syringae* van Hall. Techn Bull 66:1–64

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

Clerc A, Manceau C, Nesme X (1998) Comparison of randomly amplified polymorphic DNA with amplified fragment length polymorphism to assess genetic diversity and genetic relatedness within genospecies III of *Pseudomonas syringae.* Appl Environ Microbiol 64:1180–1187

Cournoyer B, Arnold D, Jackson R, Vivian A (1996) Phylogenetic evidence for a diversification of *Pseudomonas syringae* pv. *pisi* race 4 strains into two distinct lineages. Phytopathology 86:1051–1056

Daniels MJ, Dow JM, Osbourn AE (1988) Molecular genetics of pathogenicity in phytopathogenic bacteria. Annu Rev Phytopathol 26: 285–312

Denny TP, Gilmour MN, Selander RK (1988) Genetic diversity and relationships of two pathovars of *Pseudomonas syringae.* J Gen Microbiol 134:1949–1960

Drummond M, Whitty P, Wootton J (1986) Sequence and domain relationships of *ntrC* and *nifA* from *Klebsiella pneumoniae:* Homologies to other regulatory proteins. EMBO J 5:441–447

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

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

Felsenstein J (1985) Confidence limits on phylogenies: An approach using the bootstrap. Evolution 39:783–791

Felsenstein J (1996) PHYLIP (phylogeny inference package), version 3.57c. Department of Genetics, University of Washington, Seattle

Ferguson AR, Johnston JS, Mitchell RE (1980) Resistance of *Pseudo-*

monas syringae pv. *phaseolicola* to its own toxin, phaseolotoxin. FEMS Microbiol Lett 7:123–125

Fitch WM (1977) On the problem of discovering the most parsimonious tree. Am Nat 111:223–257

Gardan L, Cottin S, Bollet C, Hunault G (1991) Phenotypic heterogenicity of *Pseudomonas syringae* van Hall. Res Microbiol 142:995–1003

Gardan L, Shafik H, Bollet C (1994) Taxonomy of some pathovars of *Pseudomonas syringae*. In: Lemattre M, et al. (eds) Plant pathogenic bacteria, 8th Int Conf, INRA, ORSTOM, Paris, pp 419–421

Gopalan S, He SY (1996) Bacterial genes involved in the elicitation of hypersensitive response and pathogenesis. Plant Dis 80:604–610

Grimm C, Panopoulos NJ (1989) The predicted protein product of a pathogenicity locus from *Pseudomonas syringae* pv. *phaseolicola* is homologous to a highly conserved domain of several procaryotic regulatory proteins. J Bacteriol 171:5031–5038

Gruber M, Bryant DA (1997) Molecular systematic studies of eubacteria, using σ^{70}-type sigma factors of group 1 and group 2. J Bacteriol 179:1734–1747

Hasegawa M, Adachi J (1996) Phylogenetic position of cetaceans relative to artiodactyls: Reanalysis of mitochondrial and nuclear sequences. Mol Biol Evol 13:710–717

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

Hildebrand DC, Schroth MN, Sands DC (1988) Pseudomonas. In: Schaad NW (ed) Laboratory guide for identification of plant pathogenic bacteria, 2nd ed. Am Phytopathol Soc Press, St. Paul, MN, pp 60–80

Huang WM (1996) Bacterial diversity based on type II DNA topoisomerase genes. Annu Rev Genet 30:79–107

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

Jukes TH, Cantor CR (1969) Evolution of protein molecules. In: Munro HN (ed) Mammalian protein metabolism. Academic Press, New York, pp 21–132

Kimura M (1980) A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. J Mol Evol 16:111–120

Kishino H, Miyata T, Hasegawa M (1990) Maximum likelihood inference of protein phylogeny, and the origin of chloroplasts. J Mol Evol 31:151–160

Kumar S, Tamura K, Nei M (1993) MEGA: Molecular evolutionary genetics analysis, version 1.0. Pennsylvania State University, University Park

Legard DE, Aquadro CF, Hunter JE (1993) DNA sequence variation and phylogenetic relationships among strains of *Pseudomonas syringae* pv. *syringae* inferred from restriction site maps and restriction fragment length polymorphism. Appl Environ Microbiol 59:4180–4188

Lonetto M, Gribskov M, Gross CA (1992) The sigma 70 family: Sequence conservation and evolutionary relationships. J Bacteriol 174:3843–3849

Lonetto MA, Brown KL, Rudd KE, Buttner M (1994) Analysis of the *Streptomyces coelicolor sigE* gene reveals the existence of a subfamily of eubacterial RNA polymerase σ factors involved in the regulation of extracytoplasmic functions. Proc Natl Acad Sci USA 91:7573–7577

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

Manceau C, Horvais A (1997) Assessment of genetic diversity among strains of *Pseudomonas syringae* by PCR-restriction fragment length polymorphism analysis of rRNA operons with special emphasis on *P. syringae* pv. tomato. Appl Environ Microbiol 63:498–505

Martin DW, Schurr MJ, Yu H, Deretic V (1994) Analysis of promoters controlled by the putative sigma factor AlgU regulating conversion to mucoidy in *Pseudomonas aeruginosa:* Relationship to σ^E and stress response. J Bacteriol 176:6688–6696

Mitchell RE (1984) The relevance of non-host-specific toxins in the expression of virulence by pathogens. Annu Rev Phytopathol 22: 215–245

Mitchell RE, Bieleski RL (1977) Involvement of phaseolotoxin in halo blight of beans: Transport and conversion to functional toxin. Plant Physiol 60:723–729

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

Narumi M, Takikawa Y (1996) Comparative analysis of phaseolotoxin production genes in *Pseudomonas syringae* pv. *phaseolicola* and pv. *actinidiae*. In: Abstracts of the Annual Meeting of the Phytopathological Society of Japan 1996, Phytopathological Society of Japan, Tokyo, p 100 (in Japanese)

Nei M, Gojobori T (1986) Simple methods for estimating the numbers of synonymous and non-synonymous nucleotide substitutions. Mol Biol Evol 3:418–426

Patil SS, Kolattukudy PE, Diamond AE (1970) Inhibition of ornithine carbamoyltransferase from bean plants by the toxin of *Pseudomonas phaseolicola*. Plant Physiol 46:752–753

Peet RC, Panopoulos NJ (1987) Ornithine carbamoyltransferase genes and phaseolotoxin immunity in *Pseudomonas syringae* pv. *phaseolicola*. EMBO J 6:3585–3591

Peet RC, Lindgren PB, Willis DK, Panopoulos NJ (1986) Identification and cloning of genes involved in phaseolotoxin production by *Pseudomonas syringae* pv. "*phaseolicola*." J Bacteriol 166:1096–1105

Rich JJ, Willis DK (1997) Multiple loci of *Pseudomonas syringae* pv. syringae are involved in pathogenicity on bean: Restoration of one lesion-deficient mutant requires two tRNA genes. J Bacteriol 179: 2247–2258

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, pp 47–138

Rudolph K, Stahmann MA (1966) The accumulation of L-ornithine in halo-blight infected bean plants (*Phaseolus vulgaris* L.) induced by the toxin of the pathogen *Pseudomonas phaseolicola* (Burkh.) Dowson. Phytopathol Z 57:29–46

Saitou N, Nei M (1987) The neighbor-joining method: A new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425

Sawada H, Takeuchi T, Matsuda I (1995a) Nucleotide sequence analysis of phaseolotoxin-resistant OCTase gene and 16S-23S ribosomal DNA spacer from *Pseudomonas syringae* pv. *phaseolicola* and pv. *actinidiae*. In: Abstracts of the Annual Meeting of the Phytopathological Society of Japan 1995. Phytopathological Society of Japan, Tokyo, p 110 (in Japanese)

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

Sawada H, Suzuki F, Matsuda I (1996) Comparison among *Pseudomonas syringae* pathovars based on sequence analysis of 16S-23S spacer and *hrpL*. In: Abstracts of the Annual Meeting of the Phy-

topathological Society of Japan 1996. Phytopathological Society of Japan, Tokyo, p 97 (in Japanese)

Sawada H, Takeuchi T, Matsuda I (1997a) 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, Matsuda I, Saitou N (1997b) Analysis of molecular evolution of *Pseudomonas syringae.* In: Abstracts of the Annual Meeting of the Phytopathological Society of Japan 1997, Phytopathological Society of Japan, Tokyo, p 13 (in Japanese)

Shigeta S, Nakata E (1995) Bacterial brown spot of citrus caused by *Pseudomonas syringae* pv. *syringae* van Hall 1902. Ann Phytopathol Soc Jpn 61:150–157

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

Swofford DL (1993) PAUP: Phylogenetic analysis using parsimony, version 3.1.1. Illinois Natural History Survey, Champaign

Takikawa Y, Serizawa S, Ichikawa T, Tsuyumu S, Goto M (1989) *Pseudomonas syringae* pv. *actinidiae* pv. nov.: The causal bacterium of canker of kiwifruit in Japan. Ann Phytopathol Soc Jpn 55:437–444

Tamura K, Takikawa Y, Tsuyumu S, Goto M (1989) Characterization of the toxin produced by *Pseudomonas syringae* pv. *actinidiae,* the causal bacterium of kiwifruit canker. Ann Phytopathol Soc Jpn 55:512 (in Japanese)

Tamura K, Imamura M, Kohno Y, Yoneyama K, Yamaguchi I, Takahashi H (1997) Phaseolotoxin production and its role for pathogenicity in *Pseudomonas syringae* pv. *actinidiae.* In: Abstracts of the Annual Meeting of the Phytopathological Society of Japan 1997. Phytopathological Society of Japan, Tokyo, p 149 (in Japanese)

Tanaka K, Takahashi H (1991) Cloning and analysis of the gene (*rpoDA*) for the principal σ factor of *Pseudomonas aeruginosa.* Biochim Biophys Acta 1089:113–119

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

Wei Z, Beer SV (1995) *hrpL* activates *Erwinia amylovora hrp* gene transcription and is a member of the ECF subfamily of σ factors. J Bacteriol 177:6201–6210

Yamamoto S, Harayama S (1995) PCR amplification and direct sequencing of *gyrB* genes with universal primers and their application to the detection and taxonomic analysis of *Pseudomonas putida* strains. Appl Environ Microbiol 61:1104–1109

Yamamoto S, Harayama S (1996) Phylogenetic analysis of *Acinetobacter* strains based on the nucleotide sequences of *gyrB* genes and on the amino acid sequences of their products. Int J Syst Bacteriol 46:506–511

Yang Z (1996) Maximum-likelihood models for combined analyses of multiple sequence data. J Mol Evol 42:587–596

Young JM (1991) Pathogenicity and identification of the lilac pathogen, *Pseudomonas syringae* pv. *syringae* van Hall 1902. Ann Appl Biol 118:283–298

Young JM (1992) *Pseudomonas syringae* pv. *japonica* (Mukoo 1955) Dye et al. 1980 is a junior synonym of *Ps. syringae* pv. *syringae* van Hall 1902. Lett Appl Microbiol 15:129–130

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

Yuan J, He SY (1996) The *Pseudomonas syringae* Hrp regulation and secretion system controls the production and secretion of multiple extracellular proteins. J Bacteriol 178:6399–6402

Zhang Y, Rowley KB, Patil SS (1993) Genetic organization of a cluster of genes involved in the production of phaseolotoxin, a toxin produced by *Pseudomonas syringae* pv. phaseolicola. J Bacteriol 175: 6451–6458