OMICS A Journal of Integrative Biology
Volume 9, Number 1, 2005

Comparative Analyses of *Xanthomonas* and *Xylella* Complete Genomes

***LEANDRO M. MOREIRA,[1,] *ROBSON F. DE SOUZA,[1] LUCIANO A. DIGIAMPIETRI,[2] ANA C.R. DA SILVA,[3] and JOÃO C. SETUBAL[2,4]**

ABSTRACT

Computational analyses of four bacterial genomes of the *Xanthomonadaceae* family reveal new unique genes that may be involved in adaptation, pathogenicity, and host specificity. The *Xanthomonas* genus presents 3636 unique genes distributed in 1470 families, while *Xylella* genus presents 1026 unique genes distributed in 375 families. Among *Xanthomonas*-specific genes, we highlight a large number of cell wall degrading enzymes, proteases, and iron receptors, a set of energy metabolism genes, second copy of the type II secretion system, type III secretion system, flagella and chemotactic machinery, and the xanthomonadin synthesis gene cluster. Important genes unique to the *Xylella* genus are an additional copy of a type IV pili gene cluster and the complete machinery of colicin V synthesis and secretion. Intersections of gene sets from both genera reveal a cluster of genes homologous to *Salmonella's* SPI-7 island in XAC and Xf-9a5c, which might be involved in host specificity. Each genome also presents important unique genes, such as an HMS cluster, the *kdgT* gene, and O-antigen in *Xanthomonas axonopodis* pv *citri*; a number of *avrBS* genes and a distinct O-antigen in *Xanthomonas campestris* pv *campestris*, a type I restriction-modification system and a nickase gene in *Xylella fastidiosa* 9a5c, and a type II restriction-modification system and two genes related to peptidoglycan biosynthesis in *Xylella fastidiosa* temecula 1. All these differences imply a considerable number of gene gains and losses during the divergence of the four lineages, and are associated with structural genome modifications that may have a direct relation with the mode of transmission, adaptation to specific environments and pathogenicity of each organism.

[1]Departamento de Bioquímica, Instituto de Química, Universidade de São Paulo, São Paulo, Brazil.
[2]Instituto de Computação, Universidade Estadual de Campinas, Campinas, Brazil.
[3]Alellyx Applied Genomics, Campinas, Brazil.
[4]Virginia Bioinformatics Institute and Department of Computer Science, Virginia Polytechnic Institute and State University, Blacksburg, Virginia.
*These authors contributed equally to this work.

INTRODUCTION

Xylella and *Xanthomonas* are two genera in the gamma subdivision of the proteobacteria that belong to the same family, the *Xanthomonadaceae*. All known species in these genera are phytopathogens. In this paper, we compare four species in this family, the only ones for which complete genomes are publicly available.

In the *Xylella* genus, two genomes have been published: *Xylella fastidiosa* 9a5c (Xf-9a5c), the causal agent of citrus variegated chlorosis (CVC) (Simpson et al., 2000), and *Xylella fastidiosa* Temecula 1 (Xf-temecula), the causal agent of Pierce's disease, which affects grapevine (Van Sluys et al., 2003). Two additional species have been partially sequenced: *Xylella fastidiosa* Ann-1 (infects oleander) and *Xylella fastidiosa* Dixon (infects almond) (Bhattacharyya et al., 2002a,b). In the *Xanthomonas* genus, for two species there are complete genomes publicly available: *Xanthomonas axonopodis* pv *citri* (XAC) and *Xanthomonas campestris* pv *campestris* (XCC) (da Silva et al., 2002).

Some previous comparative analyses involving these species have been done. Bhattacharyya et al. (2002a,b) compared Xf-9a5c with the *X. fastidiosa* Ann-1 and Dixon strains. Van Sluys et al. (2003) compared Xf-9a5c and Xf-temecula. Da Silva et al. (2002) and Lima et al. (2005) compared the two *Xanthomonas* species. Van Sluys et al. (2002) compared the genomes of eight plant-associated bacteria, including Xf-9a5c, XAC, and XCC. Finally, Moreira et al. (2004), using the common citrus host as a reference, compared Xf-9a5c and XAC. These analyses have shown that the genomes of all four organisms are closely related, while highlighting significant differences. In this paper, we present the first analysis of all four completely sequenced representatives of the *Xanthomonadaceae*. We expand on previous analyses and report several new results. These results shed new light on the pathogenicity mechanisms used by these organisms and bring understanding of the evolution of these related pathogens to a new level.

General features of Xylella

Xylella fastidiosa is a fastidious and xylem-limited bacteria that causes a great variety of diseases in distinct plant hosts (over 100 plant species), such as Pierce's disease of grape, periwinkle wilt, phony peach disease, citrus variegated chlorosis, and almond, plum, elm, maple, oak, and sycamore leaf scorches (Hendson et al., 2001; Hopkins, 1989; Purcell, 1980; Wells et al., 1987). Although these strains are generally classified as a single species, some genetic studies suggested the existence of multiple species (Schaad et al., 2004). *Xylella* cells are approximately 0.3–0.5 μm in diameter and 3–5 μm in length, with a single chromosome of approximately 2.6 Mbp, and may contain plasmids (Holt, 1994; Wells et al., 1981). Transmission of *Xylella* involves the obligatory participation of specific insect vectors (leafhoppers), entering the oral apparatus of insects while they feed on phloem sap from plants (Lopes et al., 1996). Inside the xylem, these bacteria cause vessel occlusion through cellular agglutination, a process that seems critical for the development of symptoms, which include chlorotic areas on the upper side of leaves, reduction of fruit size and hardened consistence (Chang, 1993; Lee, 1993; Rossetti, 1990). The two isolates we analyze here, Xf-9a5c and Xf-temecula, are responsible for reduction of citrus and wine production in São Paulo state (Brazil) and California (USA), respectively. Xf-9a5c has a single chromosome of 2.67 Mbp and two plasmids, with a total of 2249 coding sequences (CDSs), 58.4% of which have been assigned putative functions (Simpson et al., 2000; Van Sluys et al., 2002). Xf-temecula carries one chromosome of 2.51 Mbp and one plasmid, with 2066 predicted CDSs, 68.1% of which were assigned putative functions (Van Sluys et al., 2003) (Table 1).

T1

General features of Xanthomonas

Infections induced by *Xanthomonas* spp. have been described for 124 monocotyledons and 268 dicotyledons, generally inducing tissue necrosis, vascular or parenchymal diseases, and damage of leaves and fruits that induce plant decline (Long and Staskawicz, 1993). *Xanthomonas* cells are rod shaped, with a diameter of 0.2–0.6 μm and a length of 0.8–2.9 μm, with a single chromosome of slightly more than 5 Mbp that may be associated with plasmids (Swings and Civerolo, 1993).

XAC is the causal agent of citrus canker, a disease characterized by the formation of canker associated with water-soaked lesions on leaves, generally surrounded by chlorotic haloes (Gottwald and Graham, 2000).

Table 1. General Features of the Four *Xanthomonadaceae* Genomes Analyzed

Features	*Xf-9a5c*	*Xf-temecula*	*XAC*	*XCC*
General features				
Compatible host	Citrus plant	Grapevine plants	Citrus plant	*Brassica* sp
Disease	Citrus variegated chlorosis (CVC)	Pierce's disease (PD)	Citrus canker (CC)	Black rot of cruciferous (BR)
Tissues infected	Xylem	Xylem	Mesophyll	Mesophyll and vessels
Symptoms	Conspicuous variegations on older leaves; chlorotic areas on the upper side; corrresponding light brown lesions; reduced and hardened fruits	Scorched leaves that detach; bare petioles attached to the canes; wilted, shriveled and raisin-like fruit	Canker lesions; abscission of fruit and leaves; general tree decline	Marginal leaf chlorosis; darkening of vascular tissue; extensive wilting and necrosis
Vector(s)	Sharpshooter leafhoppers	Sharpshooter leafhoppers	Animals, ambient conditions and utensils	Animals, ambient conditions and utensils
Importance	CVC is considered the most important citrus disease in the State of São Paulo and is responsible for large annual losses.	PD is a major threat to the viability of the California wine industry	CC causes major economic losses to the citrus industry and is a nuisance to people with ornamental citrus trees.	BR is the most serious disease of crucifer crops world wide when environmental conditions (high temperature and humidity) are favorable
Genome features				
Genome status	Complete (Simpson et al., 2000)	Complete (Van Sluys et al., 2003)	Complete (da Silva et al., 2002)	Complete (da Silva et al., 2002)
Website	www.lbi.ic.unicamp.br	www.lbi.ic.unicamp.br	http://genoma4.iq.usp.br	http://genoma4.iq.usp.br
Genome size	2,679,305 bp	2,519,802 bp	5,175,554 bp	5,076,187 bp
Plasmid(s) (CDS)	pXF51 (64) pXF1.3 (2)	pXFPD1.3 (2)	pXAC33 (42) pXAC64 (73)	—
CDS number	2249	2066	4313	4182
With assigned function (%)	1314 (58.42%)	1408 (68.15%)	2770 (64.22%)	2708 (64.75%)
GC %	52.7%	51.8%	64.7%	65.0%
Ribosomal RNA operons	2	2	2	2
Transfer RNA (AA)	49 (20)	49 (20)	54 (20)	53 (20)
Phage proteins (%)	82 (3.64%)	139 (6.72%)	35 (0.81%)	43 (1.03%)
Transposases (%)	7 (0.31%)	7 (0.33%)	84 (1.94%)	108 (2.58%)
Comparative features				
Unique genes (%)	283 (12.58%)	104 (5.98%)	665 (15.41%)	555 (13.27%)
With classified function (%)	87 (30.74%)	74 (71.15%)	207 (31.12%)	215 (38.73%)
Unique Intersection	XF-temecula 1026; XAC 47; XCC 8	XF-9A5C 1026; XAC 2; XCC 38	XCC 3636; XF-9A5C 47; XF-temecula 2	XAC 3636; XF-9A5C 8; XF-temecula 38
Plant-associated genome Comparisons Involving Xanthomonadaceae genomes	(Bhattacharyya et al., 2002a; Bhattacharyya et al., 2002b; Koide et al., 2004; Moreira et al, 2004; Nunes et al., 2003; Van Sluys et al., 2003; Van Sluys et al., 2002)	(Nunes et al., 2003 Van Sluys et al., 2003)	(da Silva et al., 2002; Lima et al., 2004; Moreira et al, 2004; Van Sluys et al., 2002)	(da Silva et al., 2002; Lima et al., 2004)

CDSs, coding sequences; AA, amino acids.

The surface penetrating necrotic lesions on fruit lead to abscission of fruit and general tree decline, causing great losses in citrus production (Gottwald and Graham, 2000). These bacteria can survive and multiply outside the host as epiphytic organisms, but in contact with stomates, hydathodes, water pores or lesions of plant tissues, they may colonize the mesophyll producing the classical symptoms of the disease in compatible hosts. Dissemination and transmission generally occur through bacteria exudates derived from lesions under wet weather and by splash dispersal at short range, windblown rain at medium to long range, and human-assisted movement at all ranges (Graham et al., 2004). XAC contains one chromosome of 5.17 Mbp and two plasmids with a total of 4313 CDSs, 64.7% of which have been assigned a putative function (da Silva et al., 2002) (Table 1).

XCC is the causal agent of black rot in cruciferous plants (*Brassica* sp) and also infects weeds, including *Arabidopsis thaliana*. XCC propagates from host to host through the same mechanisms as XAC, but causes a systemic infection (Schaad and Alvarez, 1993). In compatible hosts, XCC induces marginal leaf chlorosis, darkening of vascular tissue, extensive wilting, and necrosis. Its cells contain a single chromosome of 5.07 Mbp with 4182 CDSs, 65% of them with assigned putative functions (da Silva et al., 2002) (Table 1).

MATERIALS AND METHODS

Gene clusterization

We clustered genes into a set of gene families, using results from an all-against-all BLAST analysis (Altschul et al., 1997) for all proteins coded by the four genomes. The clustering was performed in three steps: first, we selected all pairs of bidirectional best hits with e-values of $\leq 10^{-20}$, creating mutually exclusive families of two members each. In the second step, a singleton gene g was added to an already existing family of size n if at least the first k BLAST hits of g belonged to that family ($k = 0.8n$; the value 0.8 was empirically derived). The last step joined families. Two families f_1 and f_2 were joined if each had a sufficiently high number of genes with a sufficiently high number of hits in the other family. "Sufficiently high" threshold was chosen to be 80%. These criteria were derived from similar ones used by Digiampietri et al. (2003).

We call a gene g unique (or specific) to genome x if it is a singleton or if it belongs to a gene family that does not include members from other genomes (a paralogous family). This definition is extended in the obvious way to declare genes unique to two genomes with respect to the other reference genomes. The set of genes shared by two or more genomes (orthologous families) are all genes in the families that have at least one gene from each reference genome. The set of paralogous genes from genome x is comprised of the genes in families that contain at least two genes from genome x.

Phylogenetic analysis

Selection of homologous genes for phylogenetic analysis was performed through different methodologies for isolated and concatenated genes. Isolated genes were compared to sequences in the UNIPROT database (Apweiler et al., 2004) using BLAST (Altschul et al., 1997), and valid matches were selected based on varying degrees of sequence identity, alignment coverage and e-value. Concatenated alignments were built for putative operons by selecting, for each gene in a well defined seed cluster from a certain organism (usually XAC or XF-9a5c), all homologous sequences from the KEGG (Kanehisa et al., 2004) and COG (Tatusov et al., 2001) databases, and later selecting all homologous CDSs in different genomes that were separated by no more than ten CDSs from other homologous genes of the same seed cluster. Some genes and organisms were then dropped in order to avoid missing genes in the final alignment.

Once a set of homologous sequences was chosen for each gene, selected sequences were aligned using CLUSTALW, version 1.74 (Thompson et al., 1994), with default parameters, and conserved regions were selected using the program GBLOCKS (Castresana, 2000). When dealing with clusters, alignments were concatenated at this point. All alignments were then analyzed by the PROTML program (Adachi and Hasegawa, 1996) for maximum likelihood inference of phylogenies, using the Poisson model of amino acid substitution.

Integrase classification and putative insertion/deletion islands (PinDels)

Phage insertions and transposases are known mediators of rearrangements and horizontal gene transfer events, and we searched putative insertion/deletion events (PinDels) associated with phages. We defined PinDels as regions flanked by phage integrase genes and rich in hypothetical and phage-related genes. In a second step, we also accepted, as PinDels, regions without flanking integrases but with significant GC and codon biases, and also rich in hypothetical and phage-related genes. The GC and codon biases were computed following Karlin's methodology (Karlin, 2001).

As a complement to the analysis above, integrases were classified based on two criteria: (i) sequence similarity and (ii) integration site class (Table 2). Based on sequence similarity, four distinct paralogous groups were formed for integrases (I, II, III, and IV), and an independent group was defined for recombinases xerC and xerD. Three integration site classes were defined: (A) integrases flanked by t-RNA genes, (B) integrases flanking PinDels but not t-RNAs, and (C) integrases with no flanking t-RNAs or located inside a PinDel. T2

Determination of putative terminus of replication was based on maximum value for the cumulative strand bias (Lobry, 1996).

RESULTS

Structural comparisons and PinDels of the four genomes

Figure 1 illustrates the rearrangements of relevant gene clusters among the genomes of the four *Xanthomonadaceae*. Graphs below each genome map represent variations in genome composition, measured by codon and GC biases. F1

In Xf-9a5c, four complete phages were found (Simpson et al., 2000), which could have been acquired from soil-inhabiting bacteria or directly by siphophage (Bhattacharyya et al., 2002a; Bhattacharyya et al., 2002b). In addition, 22 CDSs coding for phage proteins were found dispersed on the Xf-9a5c genome and in 11 PinDels (Fig. 1). Nunes et al. (2003) described eight of these Xf-9a5c PinDels as regions found only in citrus-related *Xylella* strains (in parenthesis the designation given by Nunes et al.): PinDel-1 (pGI2), PinDel-2 (GI1), PinDel-3 (GI3), PinDel-4 and PinDel-5 (GI2), PinDel-6 (GI4), PinDel-7 (pI1), PinDel-8 (GI_5), and PinDel-11 (pGI_6). Nunes et al. (2003) described a strong correlation of these regions with unique genes of *Xylella*. We detected two new PinDels, which we called PinDel-9 and PinDel-10. Interestingly, PinDel-6 is an inverted repeat, that is, it is composed by hypothetical and phage-related proteins that are duplicated and in reverse order. Xf-9a5c integrases flank all four phages and six PinDels (Table 2). All integrases of group II flank a phage or a PinDel (B-II). The majority of group I and all group IV Xf-9a5c integrases are flanked by tRNA genes (A-I and A-IV in Table 2). Also, six Xf-9a5c integrases (3, 4, 7, 13, 16, and 17) are positioned at the border of regions that characterize rearrangements when compared to Xf-temecula, and not just four, as described by Van Sluys et al. (2003). Therefore, sequence similarity and integration site classes are correlated, and all rearrangements between the *Xylella* strains could have been induced by phage insertion.

Although no complete phage insertion was found in Xf-temecula, 139 CDS's are grouped in 10 PinDels, 8 of which were previously described (Van Sluys et al., 2003). Xf-temecula PinDel-2 is located very close to the replication terminus (Fig. 1). Previous analysis has suggested that Xf-temecula PinDel-6 could have originated from a fusion of two different phages (Van Sluys, et al., 2003).

In *Xanthomonas* the number of genes coding for phage proteins is much smaller, with 35 CDSs in XAC and 43 CDSs in XCC (Moreira et al., 2004). No complete phage is present in XAC and a single phage ((Lf) is found in the XCC genome. We identified thirteen PinDels in XAC and seven in XCC, all flanked by phage integrases, except PinDel-3 and PinDel-7 from XCC, which were identified based on anomalous GC content. PinDels 1, 2, 5, 6, 7, 8, 9, 10, 12, and 13 in XAC and 3, 4, 5, and 6 in XCC might have been inserted or lost after divergence of the two strains. PinDel-3 of XAC is homologous to PinDel-1 of XCC and the same holds for PinDel-4 and PinDel-11 of XAC and PinDel-2 and PinDel-6 of XCC, respectively. While *Xanthomonas* presents a smaller number of integrases and phage-related proteins than *Xylella*, it has many genes coding for transposases, which are not so abundant in *Xylella*.

Table 2. *Xylella* Phage Integrases Classification

Integrase number	Genome	ORF number	Integrase Group					Insertion				Rearrangements		Nearest homologues	
			Insertion	Blast	Strand	AA	Location	Name	Location	Start	End	Number	Location	XAC	XCC
1	XF-9a5c	XF0480	A	IV	←	N	Up	PInDel-1	Up	XF0480	XF0558	—	—	XAC2682	XCC3012
2	XF-9a5c	XF0631	C	III	→	—	—	PInDel-3	Inside	XF0623	XF0648	—	—	XAC2222/XAC2183	—
3	XF-9a5c	XF0678	A	I	←	V	Up	XFP1	Up	XF0678	XF0733	I	Down	—	—
4	XF-9a5c	XF0968	A	I	←	V	Up	—	—	XF0968	—	II	Down	—	—
5	XF-9a5c	XF1425	*xerD*	Recomb	←	—	—	—	—	XF1425	—	—	—	XAC3551	XCC0654
6	XF-9a5c	XF1483	*xerC*	Recomb	→	—	—	—	—	XF1483	—	—	—	XAC0636	XCC3497
7	XF-9a5c	XF1555/XF1556	B	II	←	—	—	XFP3	Up	XF1555	XF1596	III	Down	—	—
8	XF-9a5c	XF1642	B	II	←	—	—	XFP4	Up	XF1642	XF1711	—	—	—	—
9	XF-9a5c	XF1718	A	III	→	G	Up	PInDel-7	Up	XF1718	XF1754	—	—	—	XCC2110
10	XF-9a5c	XF1754	B	III	→	—	—	PInDel-7/8	Dwn/Up	XF1718	XF1754	—	—	—	XCC2110?
11	XF-9a5c	XF1789	A	I	←	T	Down	PInDel-8	Down	XF1754	XF1793	—	—	—	—
12	XF-9a5c	XF2028	B	Singlet	←	—	—	PInDel-9	Up	XF2028	XF2081	—	—	XACb0061	XCC1630
13	XF-9a5c	XF2131/XF2132	B	II	→	—	—	PInDel-11	Down	XF2108	XF2132	IV	Down	—	—
14	XF-9a5c	XF2288	B	II	←	—	—	PInDel-12	Up	XF2288	XF2309	—	—	—	—
15	XF-9a5c	XF2309	A	I	→	G	Down	PInDel-12	Down	XF2288	XF2309	—	—	—	—
16	XF-9a5c	XF2478	B	I	→	—	—	XFP2	Up	XF2478	XF2530	V	Down	—	—
17	XF-9a5c	XF2530	B	II	→	—	—	XFP2	Down	XF2478	XF2530	V	Down	—	—
18	XF-9a5c	XF2761	A	IV	←	PRHK	Down	PInDel-13	Up	XF2761	XF2773	—	—	XAC2628	XCC3012
8	XF-temecula	PD0384	B	II	→	—	—	PInDel-1	Down	PD0363	PD0384	—	—	—	—
5	XF-temecula	PD0652	*xerD*	Recomb	←	—	—	—	—	—	—	—	—	XAC3551	XCC0654
6	XF-temecula	PD0700	*xerC*	Recomb	→	—	—	—	—	—	—	—	—	XAC0636	XCC3497
8	XF-temecula	PD0764	B	II	←	—	—	PInDel-2	Down	PD0764	PD0772	III	Down	—	—
12	XF-temecula	PD0789	C	Singlet	→	—	—	DePo #	Up	—	—	—	—	XACb0061	XCC1630
9	XF-temecula	PD0990	B	III	←	—	—	PInDel-5	Up	PD0990	PD1019	—	—	XAC2183	—
13	XF-temecula	PD1019	B	II	→	—	—	PInDel-5	Down	PD0990	PD1019	—	—	—	—
1	XF-temecula	PD1075	A	IV	←	T	Up	PInDel-6	Up	PD1075	PD1139	—	—	XAC2628	XCC3012
11	XF-temecula	PD1078	A	I	→	G	Down	PInDel-6	Up	PD1075	PD1139	—	—	—	—
8	XF-temecula	PD1139	B	II	→	—	—	PInDel-6	Down	PD1075	PD1139	—	—	—	—
8	XF-temecula	PD1196	B	II	→	—	—	PInDel-7	Down	PD1166	PD1196	—	—	—	—
14	XF-temecula	PD1320*	C	II	←	—	—	—	—	PD1320	PD1323	—	—	—	—
16	XF-temecula	PD1495	A	I	→	V	Down	—	—	—	—	V	Down	—	—
1	XF-temecula	PD1605*	A	IV	→	N	Down	PInDel-8	Down	PD1592	PD1605	—	—	XAC2628	XCC3012
16	XF-temecula	PD1732*	A	I	→	V	Down	PInDel-9	Down	PD1714	PD1732	I	Down	—	—

Integrase number 7 and 13 of Xf-9a5c is fragmented into two genes. Up, upstream; Down, downstream. *Gene with frameshift or point mutation. #Although DePo is not a PinDel, it is discussed in the text and therefor3e was highlighted. Recomb, recombinases; Xf-temecula integrase number was defined using the number of homologous integrase genes in XF-9a5c.

The location of one copy of the type II secretion system (T2SS) genes, a cluster known as xcs (Fig. 1), is interesting. Although this system is absent from *Xylella*, the regions in *Xylella* homologous to the xcs flanking regions from XAC (A and B in Fig. 1) are joined together (Moreira et al., 2004). The pattern is the same as the one found for the flanking clusters (C and D) of genes in the type III secretion system (T3SS), which are also neighbors in *Xylella* (Moreira et al., 2004). Although the copies of T2SS xps and xcs are located at different regions in the XAC genome, these clusters belong to homologous translocated regions containing recombinases xerC and xerD (Fig. 1), also observed in Xf-9a5c. These two proteins are important to separate concatenated chromosomes during replication and have been involved in genome rearrangements (Hendricks et al., 2000). Region E (*fimT*, *uvrB* and a tRNA-Val gene) flanks genes in the Type IV secretion system in XAC, and is found between rearrangements II and III in Xf-9a5c, and upstream to rearrangement II in Xf-temecula.

For XAC, we have classified the integrases in three groups of sequence similarity and identified some that are homologous to the *Xylella* integrases (Table 3). Although we have not found any correlation between sequence similarity group and integration site class in XAC, eleven of eighteen integrases flank PinDels. In XCC nine of twelve integrases flank PinDels.

T3

Analysis based on presence/absence of genes

Although many unique genes were described in previous work (da Silva et al., 2002; Moreira et al., 2004; Simpson et al., 2000; Van Sluys et al., 2002; 2003), a careful and detailed analyses of the four *Xanthomonadaceae* revealed new unique genes and systems, which seem to be involved in the processes of adaptation, colonization, and pathogenicity. These genes were classified into four distinct major groups: (i) genes unique to each strain; (ii) genes unique to each genus; (iii) genes unique to the intersection of strains of different genera; and (iv) genes unique to *Xanthomonas* and *Xylella* genera, that is, genes not found in organisms that do not belong to one of these genera. In the presentation that follows, these categories were divided into two subclasses: genes arranged in tandem, that is, close enough to be considered members of a gene cluster, and scattered unique genes.

Unique genes in each genome

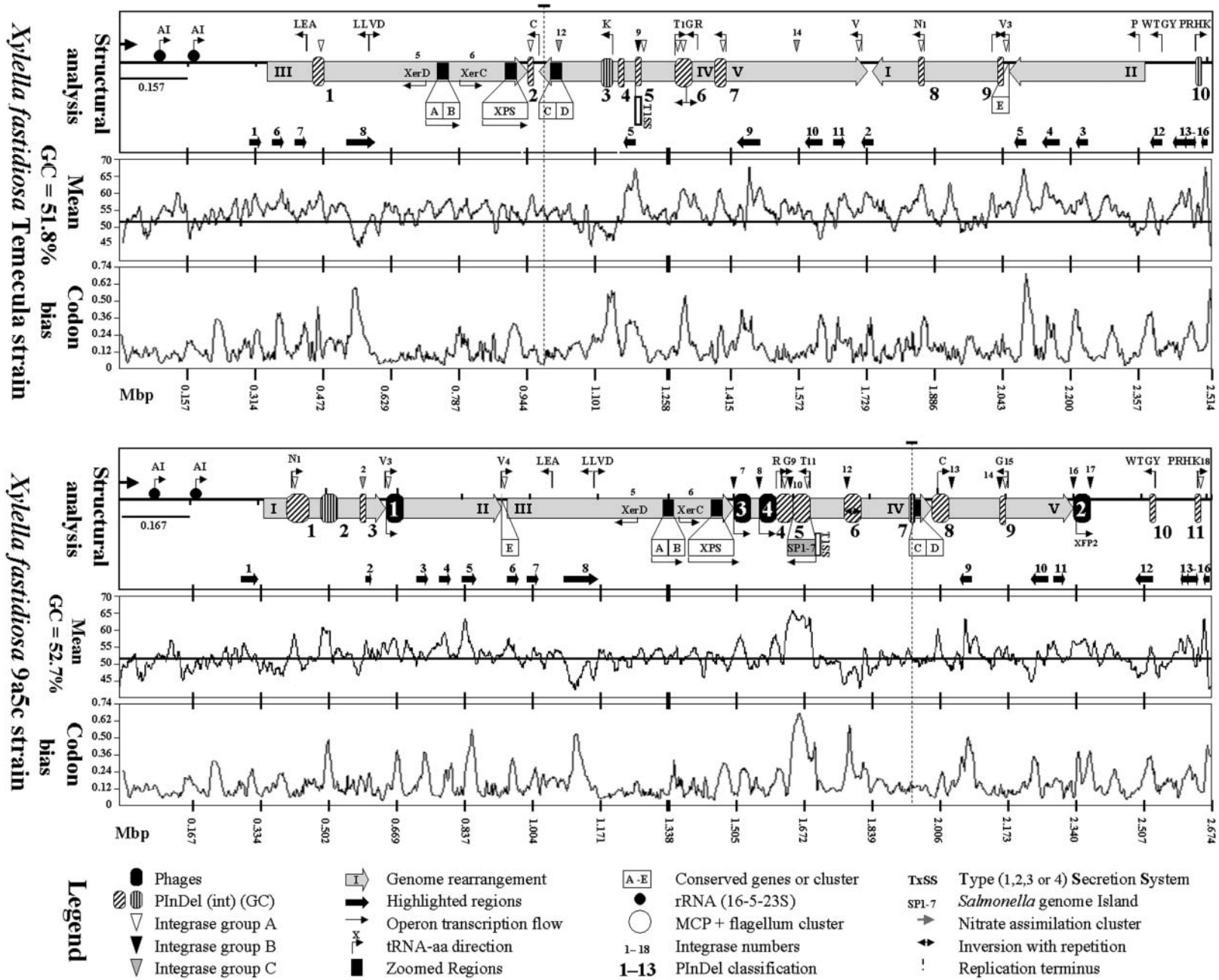
Table 4 gives the number of unique genes in each genome with respect to the others as well as unique genes in combinations of genomes. The composition of these gene sets is analyzed next.

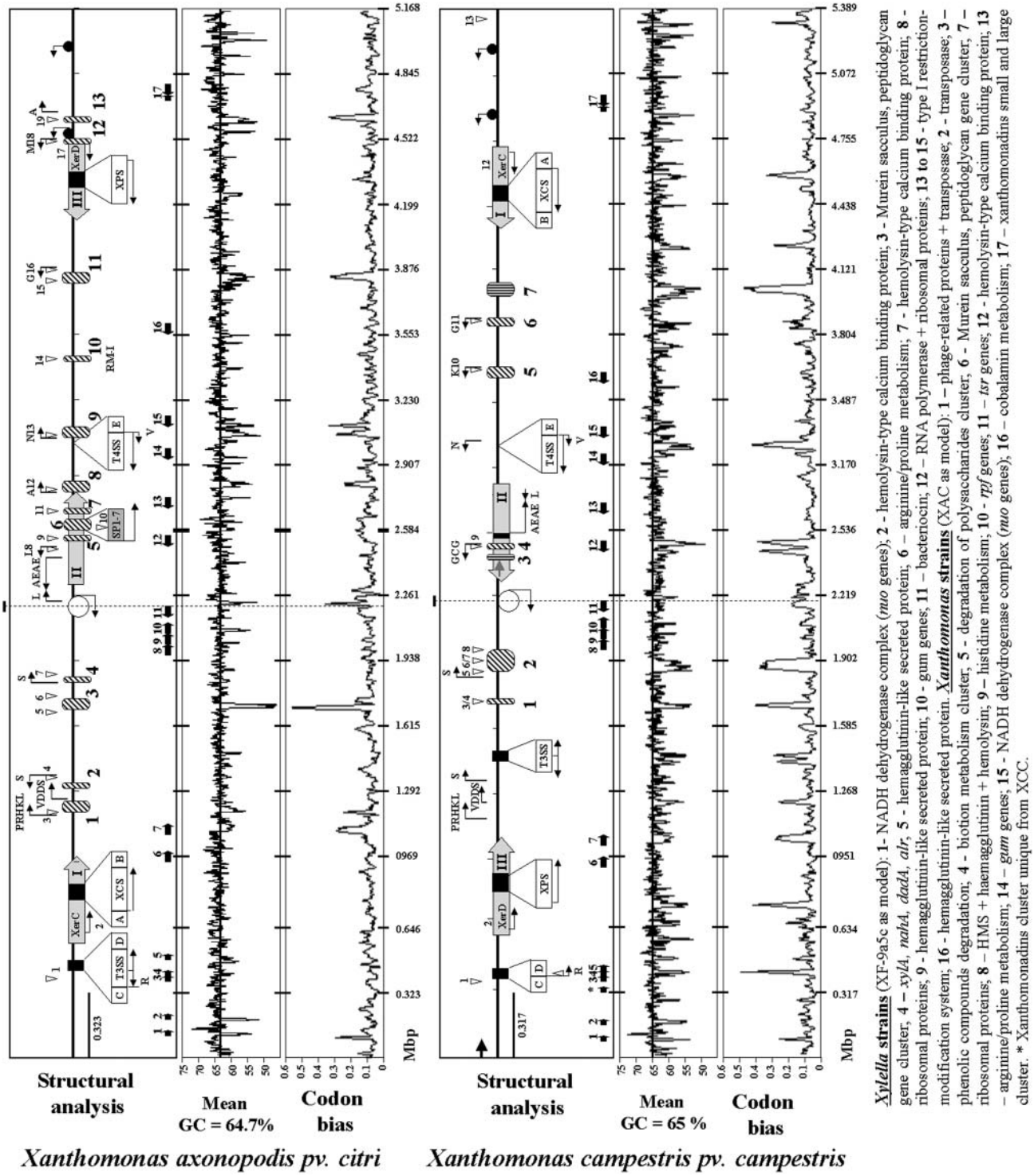
T4

XAC-specific genes. There are 18 clusters of unique genes in XAC, which we call **XaUC***n*, where the first **Xa** stands for *Xanthomonas axonopodis*, **UC** highlights that this is a unique cluster and *n* is a number from 1 to 18 (Table 5). XaUC1 (XAC0855-XAC0860) includes an ABC transporter permease complex, previously described as unique by da Silva et al. (2002), composed by the subunits *oppD*, *oppC*, and *oppB*, and an associated oligopeptide binding protein *oppA* (Fig. 2, square 8C). The cointegrate resolution protein subunits (SST) form a gene cluster (XAC3227-XAC3229) inside the region XaUC12, with a Tn5044 transposase gene (XAC3226) located upstream of the cluster and another transposase downstream, after three hypothetical genes. We note that XAC3227 and XAC3229 are homologous to integrase and cointegrase genes found in the pXAC64 plasmid (XACb0009 and XACb0010, respectively) and in a T3SS cluster from *X. campestris* pv. *vesicatoria* genome (Noel et al., 2003). The entire sequence of XAC3228, on the other hand, is highly similar to the N-terminal region of XAC3227, and thus seems to be a truncated copy of this gene. HMS proteins (Pendrak and Perry, 1991) form a single cluster comprising the subunits R, F, and H (XAC1811-XAC1813), known as pigmentation locus *pmg* in *Yersinia pestis* (Yp) (Fig. 2, square 7C). This cluster is inserted inside the XaUC15 region and is involved in uptake and storage of exogenous hemin groups (Pendrak and Perry, 1991; Schubert et al., 1998). XAC also has a XamI adenine-specific methyltransferase (XAC2436) and a type II XamI restriction enzyme (XAC2437) (da Silva et al., 2002; Gomez et al., 1997) (Fig. 2, squares 12F and 13F). These genes are located inside region XaUC9, which is composed by 29 CDSs, including 21 unique genes flanked by a tRNA-A, and six transposases, one phage and four plasmid-related proteins. The richness of plasmid-related and transposase genes suggests the involvement of plasmids in the acquisition of this region. The cluster formed by four pyridine nucleotide transhy-

T5

F2


Xylella fastidiosa Temecula strain
Structural analysis
Mean GC = 51.8%
Codon bias
Mbp
Xylella fastidiosa 9a5c strain
Structural analysis
Mean GC = 52.7%
Codon bias
Mbp
Legend
Phages
PInDel (int) (GC)
Integrase group A
Integrase group B
Integrase group C
Genome rearrangement
Highlighted regions
Operon transcription flow
tRNA-aa direction
Zoomed Regions
Conserved genes or cluster
rRNA (16-5-23S)
MCP + flagellum cluster
Integrase numbers
PInDel classification
Type (1,2,3 or 4) Secretion System
Salmonella genome Island
Nitrate assimilation cluster
Inversion with repetition
Replication terminus



***Xylella* strains** (XF-9a5c as model): **1**- NADH dehydrogenase complex (*nuo* genes); **2** - hemolysin-type calcium binding protein; **3** - Murein sacculus, peptidoglycan gene cluster; **4** – *xylA, nahA, dadA, alr*; **5** - hemagglutinin-like secreted protein; **6** – arginine/proline metabolism; **7** - hemolysin-type calcium binding protein; **8** - ribosomal proteins; **9** - hemagglutinin-like secreted protein; **10** - gum genes; **11** – bacteriocin; **12** – RNA polymerase + ribosomal proteins; **13 to 15** - type I restriction-modification system; **16** - hemagglutinin-like secreted protein. ***Xanthomonas* strains** (XAC as model): **1** – phage-related proteins + transposase; **2** - transposase; **3** – phenolic compounds degradation; **4** - biotion metabolism cluster; **5** - degradation of polysaccharides cluster; **6** - Murein sacculus, peptidoglycan gene cluster; **7** – ribosomal proteins; **8** – HMS + haemagglutinin + hemolysin; **9** – histidine metabolism; **10** - *rpf* genes; **11** – *tsr* genes; **12** - hemolysin-type calcium binding protein; **13** – arginine/proline metabolism; **14** – *gum* genes; **15** - NADH dehydrogenase complex (*nuo* genes); **16** – cobalamin metabolism; **17** – xanthomonadins small and large cluster. * Xanthomonadins cluster unique from XCC.

FIG. 1. Structure and composition of the four genomes. Each genome is depicted based on three parameters: structural composition, GC content variation, and codon bias variation. The key to symbols used are given below the *Xylella* section. Gene clusters related to pathogenicity or adaptation are depicted by numbered black arrows just above the GC content graph; the key to numbers is given below the *Xanthomonas* section. The vertical dashed line represents our estimate for the location of the terminus of replication.

Table 3. XAC and XCC Unique Gene Clusters

Island	XAC ORF number			PinDel	XAC CHP/HP	Transposases HXAC/UXAC	Phage proteins (int) HXAC/UXAC	tRNA XAC	ORF with problem HXAC/UXAC
	Initial	Final	T (U)						
XaUC01	0843	0860	18 (13)	—	—/3	—/—	—/—	—	2/—
XaUC02	0918	0924	07 (04)	—	—/—	—/—	—/—	—	—/2
XaUC03	1101	1108	08 (05)	2	—/2	2/—	(1)/1	S	—/—
XaUC04	1489	1511	23 (13)	3	3/7	2/—	(2)/1	tm	1/—
XaUC05	1809	1818	10 (05)	—	—/2	—/—	—/—	R	—/—
XaUC06	2174	2204	31 (15)	5	4/6	2/—	(1)/2	—	—/1
XaUC07	2214	2234	31 (13)	6	2/3	2/—	(1)/2	—	—/—
XaUC08	2269	2286	18 (13)	7	4/7	—/—	(1)/1	—	—/—
XaUC09	2417	2245	29 (21)	8	5/6	5/1	1+1#/3#	A	—/1
XaUC10	2901	2904	04 (04)	10	3/—	—/—	—/(1)	—	—/—
XaUC11	3018	3025	08 (08)	—	—/8	—/—	—/—	—	—/—
XaUC12	3221	3234	11 (09)	—	—/4	3/1	—/—	—	—/3
XaUC13	3251	3299	48 (23)	11	2/18	3/—	—/(2)+1#	G	1/2
XaUC14	3503	3531	29 (08)	—	1/5	2/—	—/—	—	4/—
XaUC15	3702	3732	31 (08)	—	—/—	—/—	—/—	—	1/1
XaUC16	3763	3786	24 (15)	12	6/7	2/—	(1)/—	M	2/1
XaUC17	3932	3989	58 (28)	13	10/11	3/2	—/(1)	A	1/—
XaUC18	4112	4148	37 (27)	—	16/5	1/1	—/—	—	1/2
Total	—	—	425 (232)	10	56/94	27/5	8(7)+1#/11(4)+4#	7	13/13

Island	XCC ORF number			PinDel	XCC CHP/HP	Transposases HXCC/UXCAC	Phage proteins (int) HXCC/UXCC	tRNA XCC	ORF with problem HXCC/UXCC
	Initial	Final	T (U)						
XcUC01	0319	0347	29 (20)	—	5/8	—/—	—/—	—	—/—
XcUC02	0524	0545	22 (08)	—	3/1	3/—	—/—	—	—/—
XcUC03	0599	0618	20 (10)	—	—/1	4/1	—/—	—	1/1
XcUC04	0735	0751	17 (08)	—	3/4	—/—	—/—	—	—/—
XcUC05	1307	1321	13 (06)	—	3/3	—/—	—/—	—	—/—
XcUC06	1446	1463	18 (09)	1	4/3	4/	2(1)/1	—	—/—
XcUC07	2091	2113	22 (16)	4	3/2	3/1	/(1)	GGLC	—/—
XcUC08	2413	2429	17 (07)	—	1/2	—/—	—/—	F	—/—
XcUC09	4048	4051	04 (04)	—	1/—	—/—	—/—	—	—/—
Total	—	—	162 (88)	2	23/24	14/2	2(1)/2(1)	2	1/1

XaUC*n*, **X**anthomonas **a**xonopodis **U**nique **C**luster number *n*; **XcUC***n*, **X**anthomonas **c**ampestric **U**nique **C**luster number *n*; T, Total; U, Unique; **PinDel, P**utative **in**sertion or **De**letion; CHP, Conserved Hypothetical Proteins; HP, Hypothetical Proteins; **H**XAC or **H**XCC, homologous; **U**XAC or **U**XCC, Unique; **Int,** integrases. #Plasmid-related protein.

drogenases (XAC0918-XAC0924) was named XaUC2 and has two *pntA* subunits (both with frameshifts) that promote the reversal conversion of NAD in NADP, and two *pntB* subunits that promote the transhydrogenation of NADH and NADPH. XaUC6 is formed by 31 genes and has a unique hemolysin-type calcium binding protein, placed side by side with *hlyD* and *hlyB* genes, coding for the T1SS machinery. Other eleven **XaUC***n* regions are clusters of unique genes composed almost entirely of hypothetical proteins, always accompanied by phage proteins or transposases and usually flanked by tRNAs (Table 3).

Lima et al. (2005), based on atypical GC content and codon bias, described five other unique regions in XAC. Although no genes in these five clusters presented relevant functions, some isolated genes have known

TABLE 4. NUMBER OF UNIQUE GENES AND FAMILIES FOR ALL INTERSECTIONS OF THE FOUR GENOMES

		XAC	*XCC*	*Xf-9a5c*	*Xf-temecula*
XAC	P	198 208	2046	16	2
	H	453 457	1590	31	0
XCC	P	752	188 215	2	6
	H	718	335 340	6	2
Xf-9a5c	P	8	1	74 82	419
	H	15	3	165 179	607
Xf-temecula	P	1	2	147	74 82
	H	0	1	228	30 36

Gray background cells show the number of unique gene families per genome or genome intersection. White background cells show the number of unique genes. P, genes with a putative function; H, hypotetical and conserved hypothetical genes.

functions, like an antirestriction protein that inhibits both restriction and modification by each of the four type I restriction systems in *Escherichia coli* (Delver et al., 1991), two drug resistance genes (*matE*) that mediate resistance against specific drugs, two genes related to pectin degradation, an important step in plant tissue colonization and pathogenicity, and two components of the type V autotransporter secretion system involved in translocation of specific proteins across the outer membrane via a transmembrane pore.

Among scattered unique genes of XAC, we highlight the gene that codes for 2-keto-3-deoxy-D-gluconate transport system, named *kdgT* (XAC0337), involved in uptake and transport of degraded pectin into the cell (Fig. 2, square 9C), which is placed near two copies of ISXac3 transposases. Although XCC does not have a copy of *kdgT*, genes flanking *kdgT* in XAC show a conserved order in XCC. Two genes involved in phytoene metabolism (XAC2744 and XAC3594), an important precursor of carotenoid biosynthesis, were also found in XAC and may be components of a photoprotection system that is not mediated by xanthomonadins. Three copies of XAC peptidases have homology to pseudomonapepsin and xanthomonapepsin carboxypeptidases, involved in oligopeptide absorption and degradation (da Silva et al., 2002) (Fig. 2, square 8C). XAC also contains one hrp-associated gene, *hpaF* (XAC0391), which is part of T3SS and is not found in XCC.

XCC-specific genes. As was done above for XAC, we named genomic regions in XCC enriched with unique genes as **XcUC***n* (Table 5). XcUC1 (XCC0319-XCC0347) includes a cluster of genes that code for proteins involved in xanthomonadin biosynthesis. This region has a total of 29 genes, 20 of which are found only in XCC, among the four bacteria. Six genes between XCC0332 and XCC0342 align to the amino terminal region of the *Xanthomonas oryzae pv. oryzae pigH* gene, with high similarity, suggesting a tandem duplication event. XcUC3 has 10 unique genes involved in biosynthesis of LPS O-antigen and the LPS core (XCC0599–XCC0618), previously described by Vorholter *et al*. (2001). According to these authors, knockout of genes in this region may impact on XCC pathogenic interactions with the host. Interestingly, this cluster is positioned downstream of the *xanAB*, *rmcABCD* and *ispJI* gene clusters, all related to LPS biosynthesis (Steinmann et al., 1997; Vorholter et al., 2001). Comparing XcUC3 with the genome of XAC, we note that a unique XAC gene cluster, named as XAC5 by Lima et al. (2005), is placed at a locus homologous to XcUC3 and has two ABC transporters (*rmd* and *gmd*) in common (Fig. 3a). Other genes that are inserted into XAC5 have functions that might be related to LPS biosynthesis. Therefore, a detailed functional analysis of this cluster in XAC may reveal new genes involved in LPS biosynthesis (Fig. 3b).

F3

XCC has a cluster of seven genes that may be required to assimilate and convert nitrate and nitrite into ammonium (da Silva et al., 2002) (Fig. 2, squares 7I to 7K). This cluster is placed at the end of replication and upstream of other three XCC unique regions (XCC2, XCC3, and XcUC7) (da Silva et al., 2002). XcUC7 and XcUC9 have a total of four copies of avirulence/effector protein-coding genes, which are very important to plant-host interaction and mediate bacteria colonization of compatible plants. Another important unique gene cluster is composed by three enzymes of the type III restriction-modification system (Fig. 2,

TABLE 5. *XANTHOMONAS* UNIQUE GENE FAMILIES, ORGANIZED BY PUTATIVE FUNCTION

Product	*XAC*[a]	*XCC*[b]	*No. of families*[c]	*Largest*[d]
50S Ribossomal proteins	6	6	6	—
ABC transporters	6	6	6	—
Acethyltransferases	8	8	8	—
Acyl proteins	21	20	13	683 (5)
Chemotaxis proteins	**35**	**36**	**12**	**1029 (19)**
Conserved Hypothetical proteins	775	784	713	5393/1329 (9)
Cytochromes	17	19	13	132/133 (3)
Dehydrogenase	7	6	5	2112 (3)
Flagellar proteins	**30**	**30**	**30**	**—**
Glutathione S-transferase	12	9	9	1254/1743/4689 (2)
Glycosyl proteins	8	8	8	—
Gum proteins	**4**	**4**	**4**	**—**
Hpa/Hrc/Hrp	**27**	**26**	**23**	**1104/1120/717/409 (2)**
Hydrolase	12	12	11	4994 (2)
Hypothetical	22	10	5	51/5392 (7)
Inner membrane protein	4	6	4	5210 (3)
Integral membrane protein	5	5	5	—
Integrase	**4**	**2**	**2**	**5236 (3)**
Transposase	**85**	**70**	**5**	**161/4727 (21), 2670 (17)**
Membrane protein	7	6	6	5727 (2)
MFS transporter	11	12	9	1884/1902/2097 (2)
Molybdenum/molybdopterin proteins	9	9	8	1656 (2)
Outer membrane protein	22	17	10	2025 (9)
Oxidoreductase	14	16	11	1266 (3)
Phage-related protein	**15**	**14**	**14**	**4912 (2)**
Protocatechuate degradation	**4**	**4**	**4**	**—**
TonB receptors/like proteins	**12**	**14**	**6**	**5486/5151 (4), 2396 (3)**
Transcriptional regulator	56	55	49	1952 (4)
Two-component system	14	14	10	4958/432 (2)
Type II secretion system protein	**8**	**8**	**8**	**—**
Vanillate degradation	**4**	**4**	**3**	**3488 (2)**
VirB proteins	**8**	**10**	**6**	**5446 (5)**
Total	**1272**	**1250**	**1026**	**—**

[a]Number of *X. axonopodis* pv. *cit*ri genes.

[b]Number of *X. campestris* pv. *campestris* genes.

[c]Number of orthologous/paralogous gene families.

[d]Gene family id for this function with largest number of paralogous genes (the number of genes is given in parenthesis immediately after the gene family id; when more than one family has the same number of genes, the families are separated by slashes).

The grouping was based either on the predicted protein participation in metabolic pathways and/or the predicted protein molecular function. Lines in bold type highlight systems discussed in the text.

squares 12G and 13G), which are inserted into the conserved region number 9 of XCC and XAC genomes, described previously by Lima et al. (2005). The three genes of this restriction-modification system are the only CDSs inside this region that are absent from XAC. This system is composed of a helicase subunit (XCC1067), whose gene is located upstream of a possible DNA methylase (XCC1068), and another helicase (XCC1069), which may be involved in endonucleolytic cleavage of DNA, recognition of specific short DNA sequences and cleavage of sites distant from the recognition sequence.

Previous analyses involving XCC and XAC revealed other five regions unique to XCC (Lima et al., 2005). The most interesting among these clusters is XcUC7 (region 4 of Lima et al.), which includes several transposases and a gene that codes for a tannase protein (Fig. 2, square 8I to 8K). This protein is

important to prevent damage induced by plant tannins produced as a response to infection (da Silva et al., 2002), and is located near some transposases and phage-related genes. XCC PinDel-4 is located inside XcUC7, and is one of the regions in XCC with largest deviations in GC content and codon bias (Fig. 1).

Among the scattered unique genes of XCC an avr protein, named *avrXccA2* (XCC2396), is important to host-plant interactions (Table 5). Other genes include a putative malic acid transporter (XCC3392), a gene coding for a protein related to nisin (bacteriocin) resistance (XCC3409), and a virulence associated protein (XCC4197).

Xf-9a5c–specific genes. Xf-9a5c has five large chromosomal unique regions with more than eight genes and at least 95% of hypothetical or conserved hypothetical genes, always flanked by phage proteins. Xf-9a5c has 20 genes that code for conjugal transfer proteins located in two unique regions. Eleven are in the pXF51 plasmid and nine are in the chromosome (XF2048-XF2079).

The daunorubicin C-13 ketoreductase is a unique Xf-9a5c gene involved in antibiotic production in *Streptomyces peucetius*. There is also a nickase protein similar to a *virD2* protein from an *Agrobacterium tumefaciens* plasmid, which is a site-specific endonuclease, and a plasmid maintenance protein responsible for stable maintenance of the plasmid during cell division, both located at the pXF51 plasmid. Finally, a type IV pilin involved in competence, an adherence associated pilin and a type I restriction-modification system specificity determinant protein, involved in restriction of exogenous DNA sequences, are also genes unique to Xf-9a5c (Fig. 2, squares 4F and 5F).

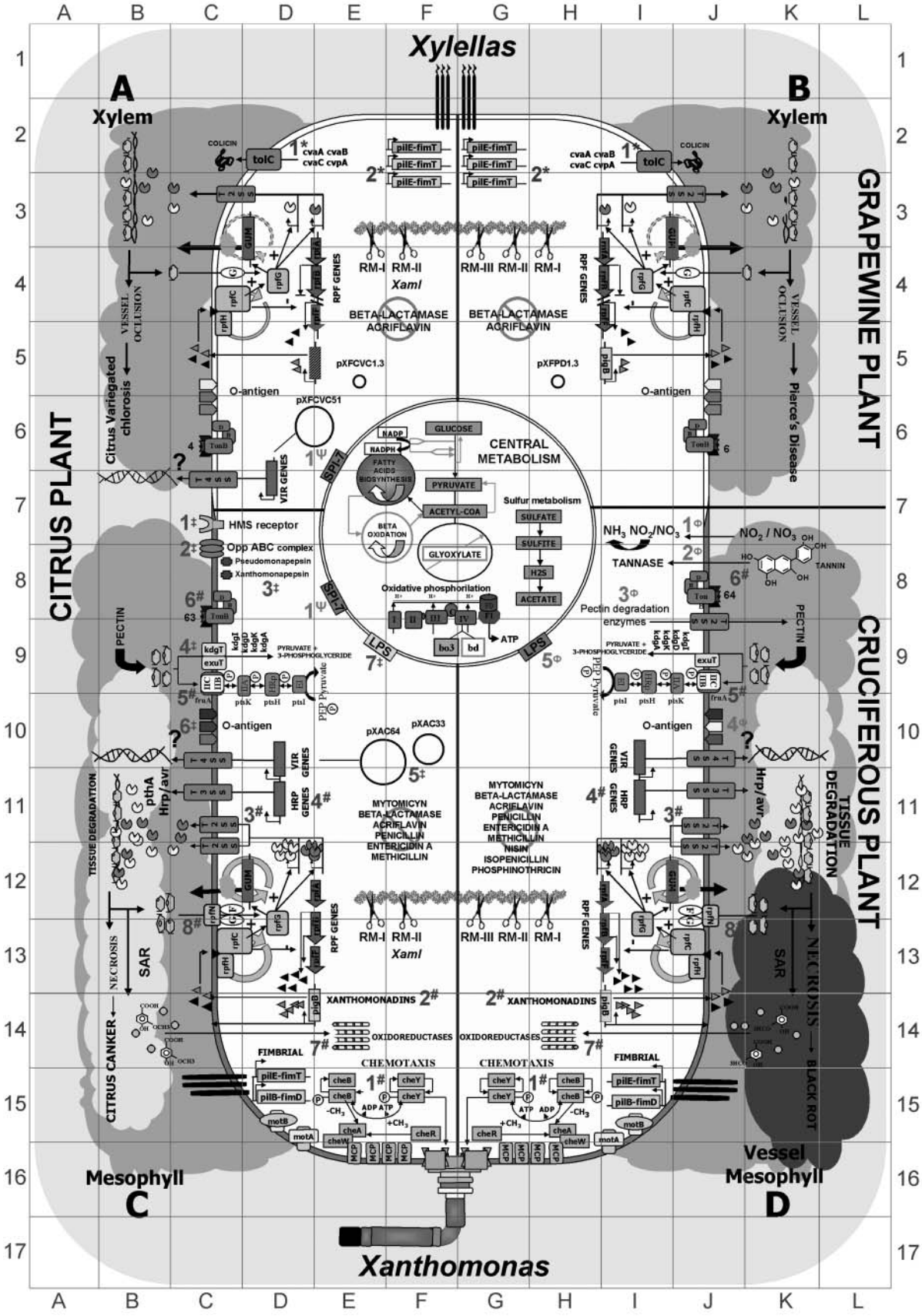
Xf-temecula–specific genes. Among the genomes compared, Xf-temecula presents the smallest number of unique genes (Table 4). Xf-temecula has four unique regions that deserve attention. Three genes, including a site-specific DNA-methyltransferase involved in restriction and modification of DNA sequences, compose the first region (Fig. 2, squares 4G and 5G). The second region (PD0906–PD0951), is the largest region unique to Xf-temecula with 30 unique CDSs, including 16 phage-related proteins and 14 hypothetical proteins. The third region comprises four genes, a proteic killer suppression protein, involved in regulation of toxin activity mediated by specific plasmids, a virulence-associated protein (*vapI*), related to integrative plasmids, and a *HicA*-related protein that in *Haemophilus influenzae* is inserted into the *Hif*-contiguous pilus cluster. The fourth region has a phage-related endolysin, a type II restriction enzyme (*nspV*), and its respective methylase (PD1667-PD1668), which are involved in site-specific restriction and modification (Fig. 2, squares 4G and 5G).

Xf-temecula has two unique genes that are involved in cell wall formation: a UDP-*N*-acetylmuramate-L-alanyl-gamma-D-glutamyl-meso-diaminopimelate ligase, which catalyzes the last step in the murein tripeptide recycling pathway, and a UDP-*N*-acetylglucosamine-*N*-acetylmuramyl-(pentapeptide) pyrophosphoryl-undecaprenol *N*-acetylglucosamine transferase, that catalyzes the last step of peptidoglycan biosynthesis.

Genes specific for several genomes

Genes unique to Xanthomonas *with respect to* Xylella. Among the *Xanthomonas* genes arranged in clusters, we have found genes coding for chemotaxis proteins, flagellar structural units, copies of type II and III secretory systems and iron receptors. Both *Xanthomonas* possess a complete set of genes involved in synthesis and regulation of flagella (Fig. 2, squares 15E to 15H, and 16E to 16H) (da Silva et al., 2002; Moreira et al., 2004). Ten XAC and nine XCC copies of a methyl chemotactic protein gene (*mcp*), called *tsr*, are located in a tandem duplication, comprising about 15 Kb. The *tsr* gene is involved in capture and transference of chemotactic signals from the environment to the bacterial cytoplasm, and responds to presence of the amino acids serine, alanine or glycine (Stock and Surette, 1994; Stock et al., 1994; Stock et al., 1992). Phylogenetic analysis of *mcp* and *tsr* genes from XAC, XCC and some other bacteria have shown that, although all copies of *tsr* in this region belong to a single clade, some *mcp* genes from outside this region also cluster together with high bootstrap support values in both genomes. Therefore, these genes must have been duplicated before divergence of XAC and XCC lineages (Fig. 4) (Moreira et al., 2004).

F4


Xylellas
A
Xylem
B
Xylem
GRAPEWINE PLANT
CRUCIFEROUS PLANT
CITRUS PLANT
Xanthomonas
C
Mesophyll
D
Vessel
Mesophyll
COLICIN
tolC
cvaA cvaB
cvaC cvpA
pilE-fimT
pilB-fimD
RM-I
RM-II
RM-III
XamI
RPF GENES
BETA-LACTAMASE
ACRIFLAVIN
VESSEL OCCLUSION
Citrus Variegated chlorosis
Pierce's Disease
pXFCVC1.3
pXFPD1.3
pXFCVC51
O-antigen
VIR GENES
CENTRAL METABOLISM
GLUCOSE
NADP
NADPH
FATTY ACIDS BIOSYNTHESIS
PYRUVATE
ACETYL-COA
BETA OXIDATION
GLYOXYLATE
Sulfur metabolism
SULFATE
SULFITE
H2S
ACETATE
Oxidative phosphorilation
ATP
bo3
bd
SPI-1
SPI-2
LPS
HMS receptor
Opp ABC complex
Pseudomonapepsin
Xanthomonapepsin
PECTIN
PYRUVATE + 3-PHOSPHOGLYCERIDE
PEP Pyruvate
pXAC64
pXAC33
HRP GENES
Hrp/avr
pthA
TISSUE DEGRADATION
NECROSIS
SAR
CITRUS CANKER
MYTOMICYN
BETA-LACTAMASE
ACRIFLAVIN
PENICILLIN
ENTERICIDIN A
METHICILLIN
NISIN
ISOPENICILLIN
PHOSPHINOTHRICIN
XANTHOMONADINS
OXIDOREDUCTASES
FIMBRIAL
CHEMOTAXIS
motB
motA
cheA
cheB
cheR
cheW
cheY
MCP
NH3 NO2/NO3
NO2 / NO3
TANNASE
TANNIN
Pectin degradation enzymes
BLACK ROT

The type II secretory system (T2SS) is responsible for the export of exoenzymes essential to host colonization in plant pathogens (Housby et al., 1998; Lee et al., 2004; Lee et al., 2000; Thomas et al., 1997). In *Xanthomonas*, two copies of this system were described (Fig. 2, square 11C and 11J), while for *Xylella* strains just the copy known as xps was found (Fig. 2, square 3C and 3J) (Moreira et al., 2004). Both T2SS clusters of *Xanthomonas* genomes are located in the largest genomic rearrangements between XAC and XCC genomes (da Silva et al., 2002), near to genes coding for XerC and XerD recombinases (Fig. 1). A phylogeny for the two copies of T2SS (Fig. 5) suggests that divergence between these two operons is ancient, dating at least to the ancestor of *Xanthomonadaceae* (Moreira et al., 2004).

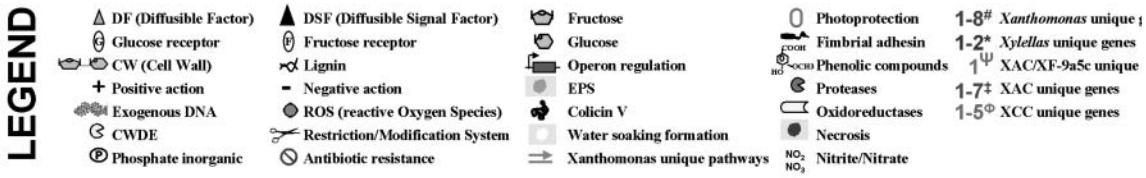
F5

The type III secretory system (T3SS) is a molecular device of gram-negative bacteria specialized in delivery of effectors proteins across the membrane barrier of compatible hosts (Collmer et al., 2000). Both *Xanthomonas* strains have genes that code for T3SS, known as *hrp* genes (Fig. 2, square 11C and 11J) (da Silva et al., 2002). There are nine XAC and thirteen XCC *hrp* genes, nine XAC/XCC hrp-conserved genes (*hrc*), six XAC and five XCC hrp-associated genes (*hpa*) and one unknown gene unique to XAC. Copies of T3SS in XAC and XCC are located at non-equivalent positions, that is, they are flanked by non-homologous regions at clearly different genome loci (Fig. 1). *Xylella* strains do not have a copy of this system (Simpson et al., 2000; Van Sluys et al., 2003) (Fig. 2); they also lack a cluster that codes for poly- and oligosaccharide (DePo) degradation, which flanks the T3SS cluster on both *Xanthomonas* (Fig. 1) (Moreira et al., 2004).

Besides the complexes and genes cited above, the two *Xanthomonas* present some other unique gene clusters that may be related to adaptation and pathogenesis. A unique cluster related to uptake and metabolism of carbon sources codes for a C4-dicarboxylate transport system and two tandem copies of 4-hydroxy-2-oxoglutarate-aldolase/2-deydro-3-deoxyphosphogluconate aldolase, related to degradation of D-galactonate, might provide an alternative route for carbon uptake. A complete PTS fructose system, related to uptake and metabolism of fructose, is present, including the subunits IIC and IIB, which are absent in *Xylella*, and may easy entrance of sugars into the cell (da Silva et al., 2002) (Fig. 2, squares 9C, 9D, 9I and 9J).

Two systems related to cytochrome biosynthesis are also unique to *Xanthomonas*: a cluster related to cobalamin and pyrroloquinoline quinone (PQQ) biosynthesis, and three clusters coding for cytochrome D ubiquinol oxidase I (*cydA*) and II (*cydB*) subunits. Cytochrome D ubiquinol oxidase was implicated in maintaining low intracellular oxygen concentrations as a requirement for eventual nitrogen fixation in nitrogen-fixing bacteria (Kelly et al., 1990). The *Xanthomonas* species also have two unique clusters possibly involved in osmotic stress: a potassium-transporting ATPase cluster, a cluster involved in putrescine binding, and a transport protein required for the periplasmic transport of putrescine. There are also three copies of microcystin dependent proteins, in tandem, which are related to inhibition of protein phosphatases (Sivonen et al., 1992). Microcystin is a nonribosomally produced cyclic heptapeptide found in toxic strains of *Microcystis*, *Anabaena*, *Nostoc*, and *Oscillatoria* (Sivonen et al., 1992).

FIG. 2. Comparative view of the Xanthomonadaceae biological processes. Comparative view of the biological processes involved in the lifestyle of Xf-9a5c (**A**), Xf-temecula (**B**), XAC (**C**), and XCC (**D**). Horizontal and vertical black lines separate each organism's biological processes according to the environments in which they live. Using the horizontal axis, it is possible to compare same genus organisms. The central circle represents the main chromosome and within this circle is shown the central metabolism found in these bacteria. The central circle border alludes to the main chromosome DNA molecule, and some unique systems were plotted on it, such as the SPI-7 island (Fig. 7) and LPS synthesis genes (Fig. 3).



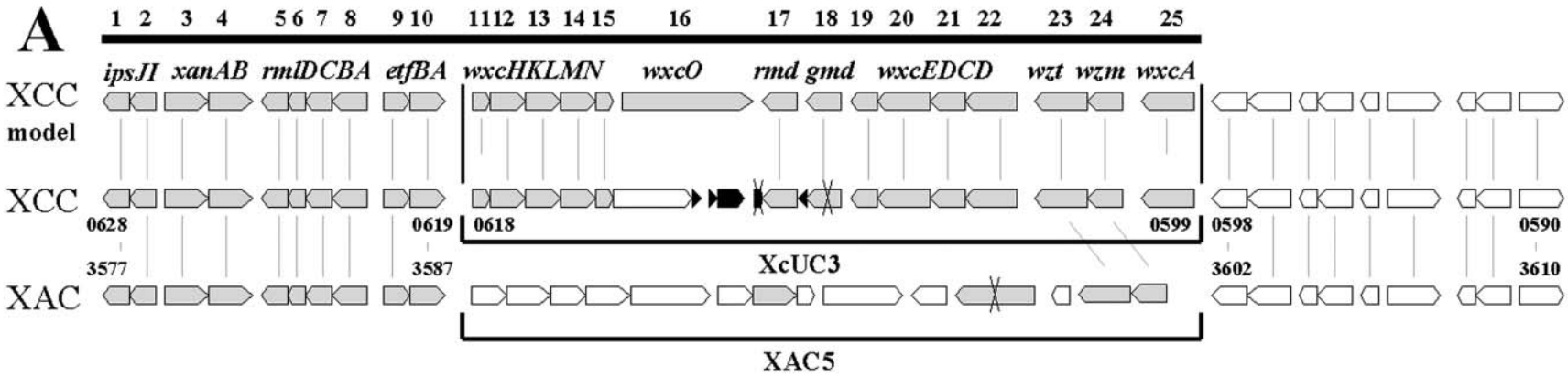


FIG. 3. (**A**) Analysis of the genomic region that codes for genes related to LPS biosynthesis in XAC and XCC. The first line describes the structural composition previously given by Vorholter et al. (2001) for XCC, using their gene IDs. The horizontal bar above XCC was created to facilitate reference in B. Each arrow represents one gene that may (gray) or may not (white) have a function closely associated with LPS biosynthesis. The second and third lines represent the information described previously by da Silva et al. (2002). XcUC3 is one of the regions unique to the XCC genome (Table 3), whereas XAC5 is a unique region described by Lima et al. (2005). (**B**) Model of LPS biosynthesis. Model of LPS biosynthesis adapted from Vorholter et al. (2001). The numbering 1-17 represents the genes given in A. The genes that do not have a number are named with their respective gene names. All boxed reactions are absent in the *Xylella* genomes. The reactions boxed by continuous lines are unique to XCC, while the dashed box is unique to both XAC and XCC with respect to *Xylella*.

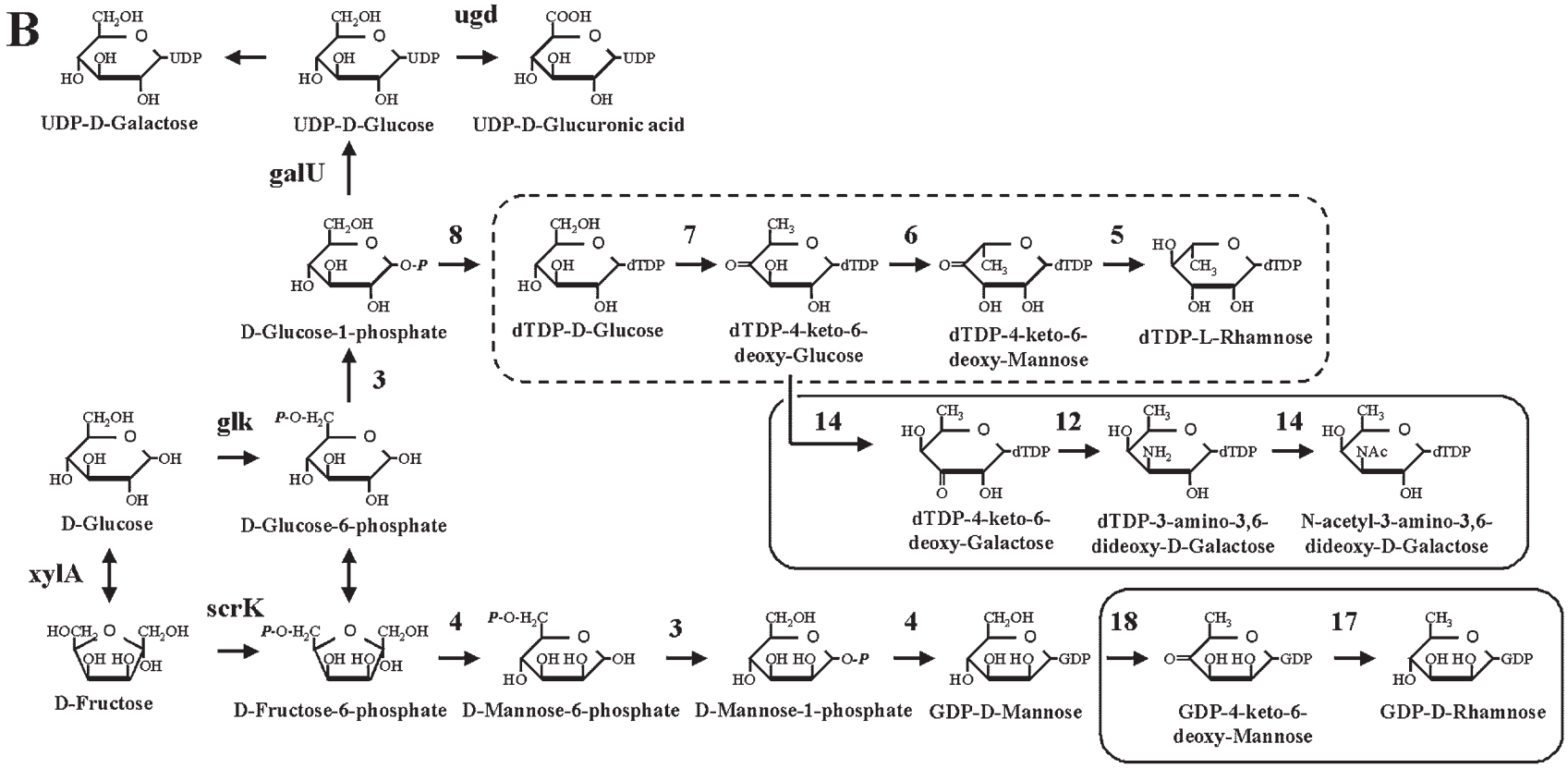
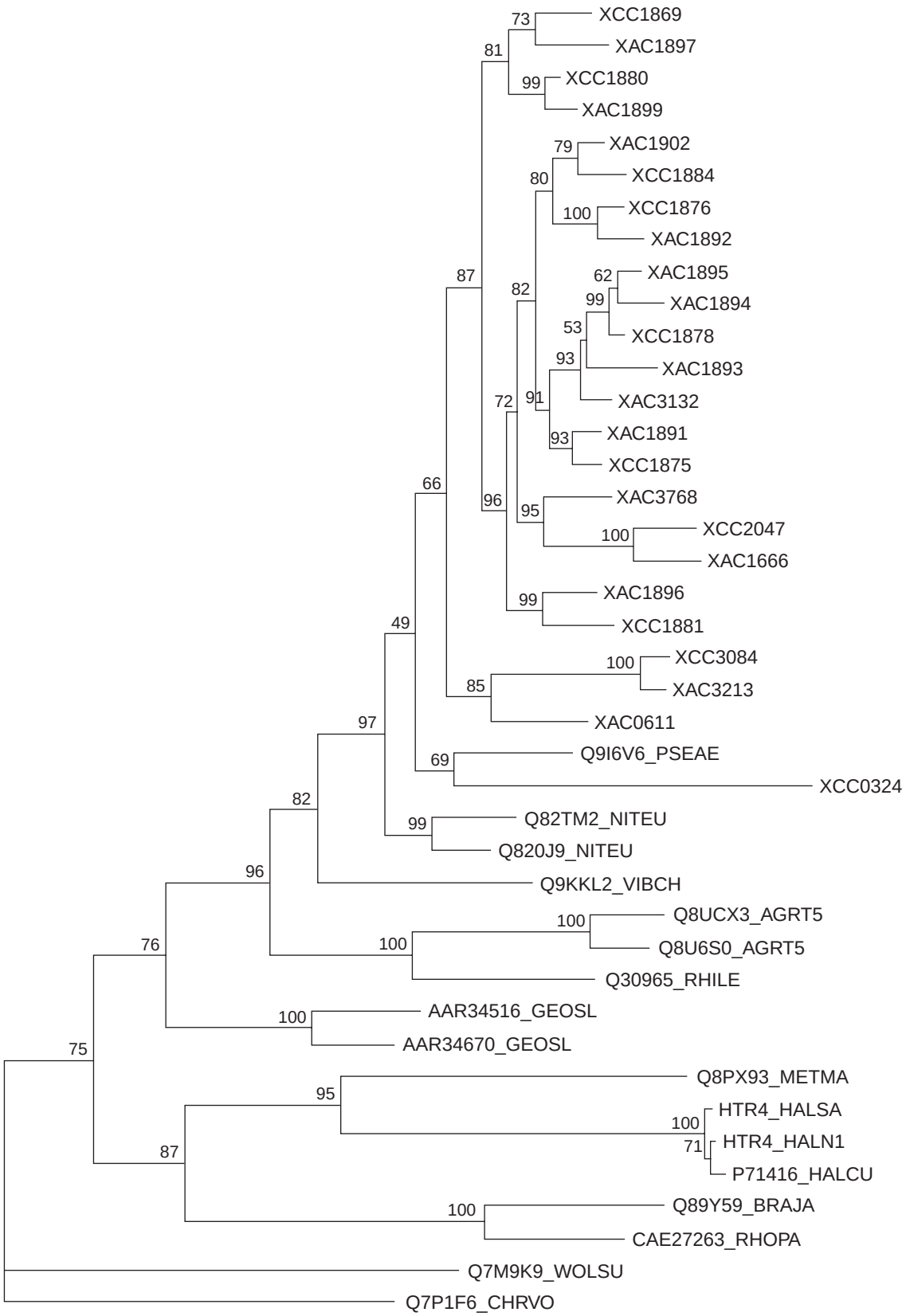


FIG. 3. (*Continued*)


XCC1869
XAC1897
XCC1880
XAC1899
XAC1902
XCC1884
XCC1876
XAC1892
XAC1895
XAC1894
XCC1878
XAC1893
XAC3132
XAC1891
XCC1875
XAC3768
XCC2047
XAC1666
XAC1896
XCC1881
XCC3084
XAC3213
XAC0611
Q9I6V6_PSEAE
XCC0324
Q82TM2_NITEU
Q820J9_NITEU
Q9KKL2_VIBCH
Q8UCX3_AGRT5
Q8U6S0_AGRT5
Q30965_RHILE
AAR34516_GEOSL
AAR34670_GEOSL
Q8PX93_METMA
HTR4_HALSA
HTR4_HALN1
P71416_HALCU
Q89Y59_BRAJA
CAE27263_RHOPA
Q7M9K9_WOLSU
Q7P1F6_CHRVO

Unique *Xanthomonas* genes were also found located scattered in both genomes. Most of these genes have a regulatory function and may be directly involved in regulation of pathogenicity and adaptation. Among them, we highlight a large number of genes involved directly or indirectly with sugar metabolism: a copy of fructose-1,6-bisphosphatase, an important enzyme in the gluconeogenesis pathway; *kdgA* gene (KDPG and KHG aldolase), which codes a key enzyme in the Entner-Doudoroff pathway (Fig. 2, squares 9D and 9J); two xylose repressor genes involved in regulation of the *xylBAFGHR* operon; a *tldD* gene that suppresses the inhibitory activity of the carbon storage regulator (*csrA*), possibly affecting glycogen biosynthesis, gluconeogenesis, cell size and surface properties; two copies in XAC and a copy in XCC of 2-keto-3-deoxygluconate kinase, that catalyzes phosphorylation of the first common intermediate in the D-glucuronate and D-galacturonate catabolic pathways. Besides genes for cell-wall degradation, sugar uptake and metabolism, the two *Xanthomonas* genomes also have two subunits of gum genes that are absent in *Xylella* (da Silva et al., 2001).

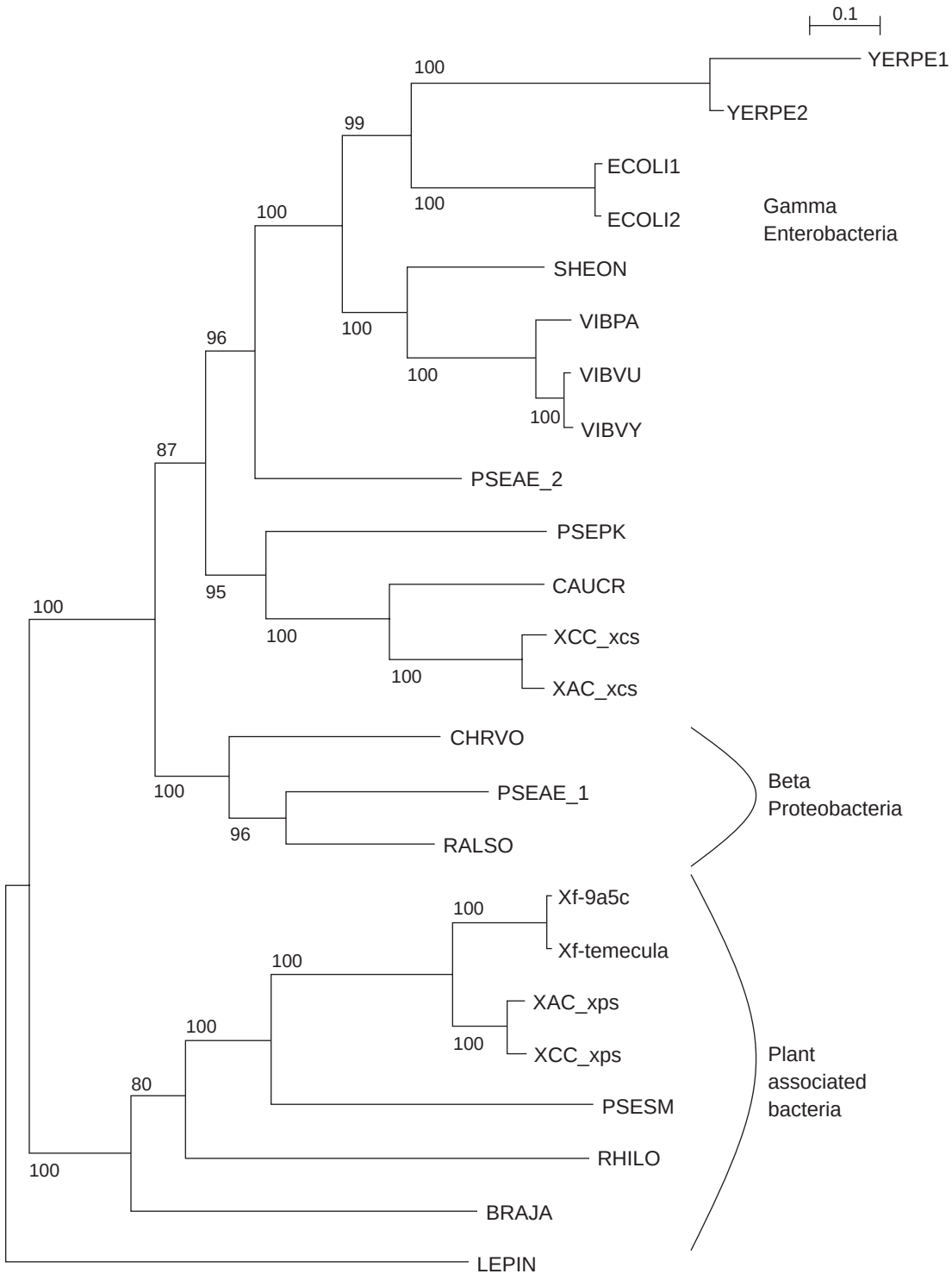
The two *Xanthomonas* genomes have unique genes related to the uptake of nitrogen, such as an ABC transporter amino acid permease, which eases uptake and processing of nitrogen, associated to a nitrogen regulatory IIA protein and a fumarate and nitrate reduction regulatory protein, which may have a role in linking carbon and nitrogen assimilation. Both genomes also have two copies of a sulfate permease that may provide alternative routes of energy production based on sulfur metabolism.

Both *Xanthomonas* present a large number of genes related to adaptation after the plant response to infection. Eight and six copies of chloroacetaldehyde dehydrogenase were found on XAC and XCC respectively. This gene is involved in detoxification of chloroacetaldehyde to chloroacetic acid, and therefore is important to adaptation and colonization of mesophyll. Genes related to vanilate, protocatechuate, hydroxybenzoate and benzoate degradation, all phenolic compounds, also exist and are directly involved in defense against plant chemical defenses produced during systemic acquired resistance (SAR) (Fig. 2, squares 14B to 14E and 14H to 14K).

The two *Xanthomonas* genomes studied here have a large number of genes coding for efflux and influx pumps, and genes related to detoxification and antibiotic resistance. Among the unique genes that mediate the interchange of compounds between the cytosol and the environment, we highlight nine copies in XAC and six copies in XCC of outer membrane proteins homologous to an efflux system protein called *nodT*, a component of multidrug efflux pump; four copies of a multidrug resistance efflux pump (*fusE*), responsible for export of toxins; a gene that codes for a toxin secretion ABC transporter ATP-binding protein; genes associated with iron storage and detoxification (bacteriferritin and bacterioferritin-associated ferredoxin); and genes related to the synthesis of choline, betaine and taurine, compounds involved in osmotic control.

Both *Xanthomonas* present a large number of genes that code for different iron receptors, with 63 genes in XAC and 64 genes in XCC (Fig. 2, squares 8C and 8J). Among these genes, there are receptors for specific siderophores, such as: pyoverdine, enterobactin and pseudobactin; and for other compounds that also chelate iron, such as hemin, cobalamin and citrate (Braun and Braun, 2002; Buyer and Leong, 1986; Cor-

FIG. 4. Unrooted phylogenetic tree inferred for *tsr/mcp* genes from XAC, XCC, and other bacteria. *Xanthomonas* genes are identified by their ordered locus names (XAC#### or XCC####), as available in the GN field from UniProt. These genes form a monophyletic group at the top, and most *tsr* genes in this group are present in orthologous pairs and arranged in tandem in both *Xanthomonas*, thus implying a lineage specific expansion of *tsr* genes before the divergence of XAC and XCC. Selection of homologs was made requiring all sequences to be BLAST hits of XCC1869 with at least 60% coverage and e-value of $\leq 10^{-10}$. Phylogeny inference was performed using maximum likelihood and assuming a Poisson model of evolution. Numbers close to internal branches indicate bootstrap support values. Sequences from other bacteria are identified by their SwissProt accession numbers (four alphanumeric + underscore + five letters) or by a TrEMBL ID and SwissProt species code. Species codes are *Agrobacterium tumefaciens* (strain C58/ATCC 33970) (**AGRT5**); *Bradyrhizobium japonicum* (**BRAJA**); *Chromobacterium violaceum* (**CHRVO**); *Geobacter sulfurreducens* (**GEOSL**); *Halobacterium cutirubrum* (**HALCU**); *Halobacterium sp*. (strain NRC-1/ATCC 700922 / JCM 11081) (**HALN1**); *Halobacterium salinarium* (**HALSA**); *Methanosarcina mazei* (**METMA**); *Nitrosomonas europaea* (**NITEU**); *Pseudomonas aeruginosa* (**PSEAE**); *Rhizobium leguminosarum* (**RHILE**); *Rhodopseudomonas palustris* (**RHOPA**); *Vibrio cholerae* (**VIBCH**) and *Wolinella succinogenes* (**WOLSU**).


0.1
YERPE1
YERPE2
ECOLI1
ECOLI2
Gamma
Enterobacteria
SHEON
VIBPA
VIBVU
VIBVY
PSEAE_2
PSEPK
CAUCR
XCC_xcs
XAC_xcs
CHRVO
PSEAE_1
RALSO
Beta
Proteobacteria
Xf-9a5c
Xf-temecula
XAC_xps
XCC_xps
PSESM
RHILO
BRAJA
LEPIN
Plant
associated
bacteria
100
99
100
100
96
100
100
87
95
100
100
100
100
96
100
100
100
80
100

nelis and Matthijs, 2002; Khalil-Rizvi et al., 1997; Koster et al., 1995). The *Xylella* genomes have significantly fewer of these receptors. Six genes for iron receptors are found in Xf-temecula and four genes were described in Xf-9a5c (Fig. 2, squares 6C and 6J).

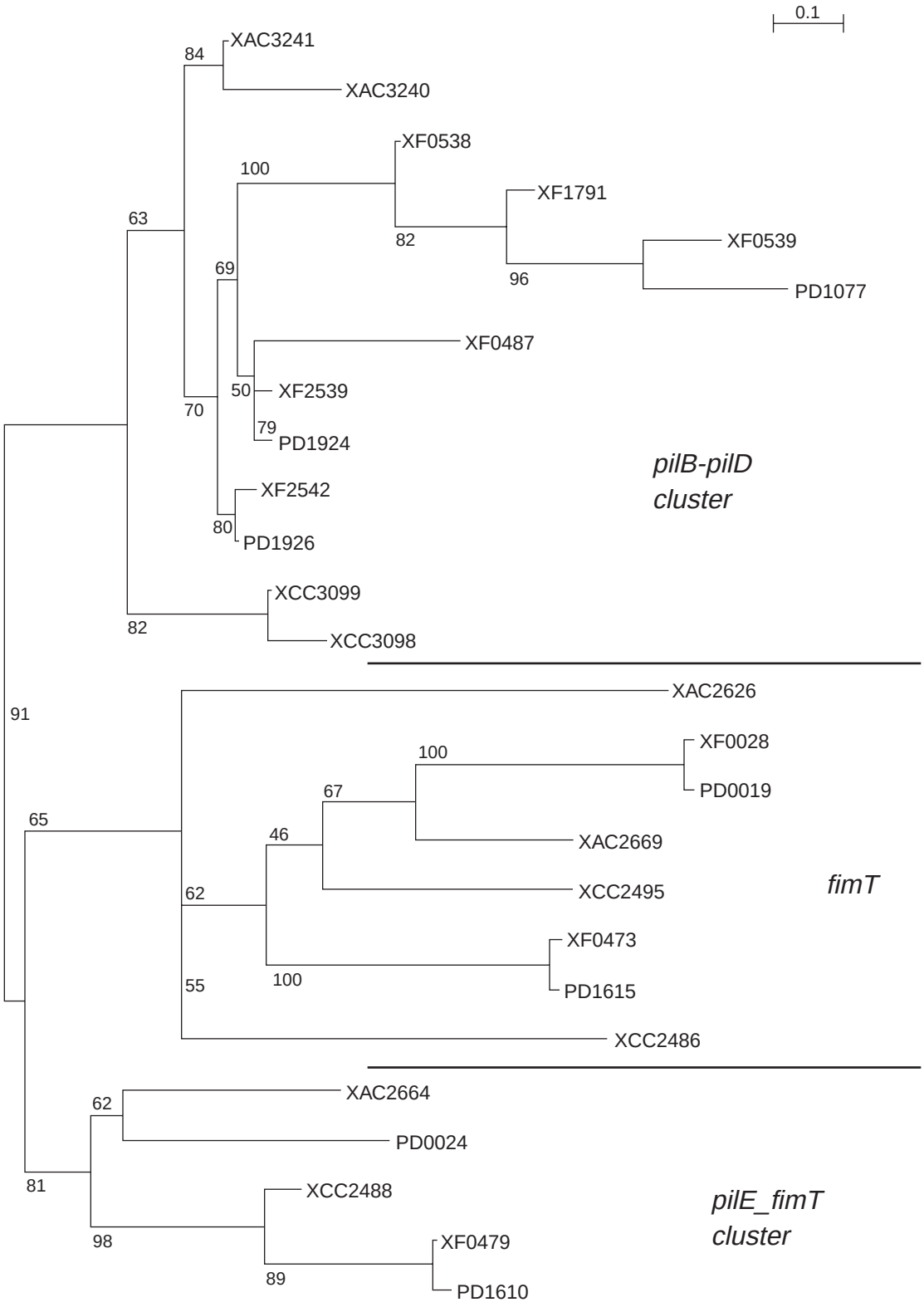
Genes unique to Xylella with respect to Xanthomonas. Although *Xylella* has fewer unique genes than *Xanthomonas* (Table 4), some important systems are found among them. Type IV pili are filamentous structures placed at bacterial polar regions and are responsible for movements called twitching motility on adhesion surfaces (Alm and Mattick, 1997; Strom and Lory, 1993). Both *Xylella* and *Xanthomonas* have genes that code for components of this structure, but *Xylella* shows a higher number of copies of such genes, including several copies of the structural subunit gene *pilE* and, most importantly, two copies of the cluster *pilE-fimT* (Moreira et al., 2004) (Fig. 2, squares 2F and 2G). Phylogenetic analysis suggests the presence and divergence of two copies of the *pilE-fimT* cluster in the ancestor of *Xanthomonadaceae*, both copies being preserved in *Xylella* (Fig. 6), while one copy was lost in XAC and the other in XCC, thus implying independent losses after divergence of the two *Xanthomonas* lineages (Fig. 2, squares 15D and 15I). The *pilE-fimT* clusters in XAC (XAC0974-XAC0984) and XCC (XCC2488-XCC2495) have low %GC (54%) when compared to the genome average (65%) and to GC content in its flanking regions, which are rich in phage and transposase insertions.

F6

Among the several unique genes found scattered in *Xylella* genomes, those involved in biosynthesis and secretion of colicin and the gene that codes for arginine deaminase deserve special attention. Colicins are small proteins that promote bacterial antagonism during nutritional limitation (Lazdunski et al., 1998). *Xylella* strains have copies of all genes necessary for colicin biosynthesis, but both *Xanthomonas* strains have no copy of the *cvaC* and *cvaA* genes and the *cvaB* gene is present only in XAC (Fig. 2, squares 2C, 2D, 2I, 2J). The absence of *cvaA* gene in XAC is paralleled by the insertion of two transposases that might have mediated rupture and depletion of the *cvaAB* cluster. In XCC, both colicin genes are absent and there are no transposases. The genes *cvaABC* are regulated by the *cvpA-purF* gene cluster, which was first identified in *Escherichia coli* plasmid pColV-K30 (Fath et al., 1989; Waters and Crosa, 1991). All four phytopathogens have copies of the *cvpA-purF* cluster, showing a high degree of similarity to *E. coli* genes, but located in the chromosome (Fath et al., 1989; Lazdunski et al., 1998; Waters and Crosa, 1991). A recent *in vitro* approach revealed that the colicin genes are functional and are expressed under high sugar concentration (Pashalidis et al., 2005).

Arginine deaminase (*rocF*) was formerly described as a gene that may define the pathogenic or non-pathogenic phenotype for strains of *Xylella fastidiosa* (Koide et al., 2004). Curiously, this gene is not found in *Xanthomonas* and is located inside a region that presents other seven unique genes (XF1232-XF1255 and PD0508-PD0521), most of them hypothetical or conserved hypothetical proteins. This gene may be important against NO produced by the plant.

FIG. 5. Unrooted phylogenetic tree for the concatenated alignment of subunits D, E, F, G, H, and K of the T2SS gene cluster. The *xps* clusters found in **XAC** and **XCC** are homologous to the T2SS from *Xylella* (**Xf-9a5c** and **Xf-temecula**) and belong to a well-supported clade together with other plant-associated bacteria, while *xcs* is more related to secrectory systems found in enterobacteria. *Leptospira* (**LEPIN**) was used to orient the tree in order to separate the systems found in entero- and plant-associated bacteria. Selection of homologous sequences was made based on KEGG and COG orthologous groups. Phylogenies were inferred using maximum likelihood and a Poisson model for aminoacid substitution. Numbers close to internal branches indicate bootstrap support values. Codes for species are **XAC**—*Xanthomonas axonopodis* (pv. *citri*); **XCC**—*Xanthomonas campestris* (*pv. campestris*); **XF-9a5c**—*Xylella fastidiosa*; **XF-temecula**—*Xylella fastidiosa* (strain Temecula1/ATCC 700964); **BRAJA**—*Bradyrhizobium japonicum*; **CAUCR**—*Caulobacter crescentus*; **CHRVO**—*Chromobacterium violaceum*; **ECOLI1**—*Escherichia coli* CFT073; **ECOLI2**—*Escherichia coli* K-12 MG1655; **LEPIN**—*Leptospira interrogans*; **RHILO**—*Mesorhizobium loti*; **PSEAE**—*Pseudomonas aeruginosa*; **PSEPK**—*Pseudomonas putida* (strain KT2440); **PSESM**—*Pseudomonas syringae* (pv. *tomato*); **RALSO**—*Ralstonia solanacearum*; **SHEON**—*Shewanella oneidensis*; **VIBCH**—*Vibrio cholerae*; **VIBPA**—*Vibrio parahaemolyticus*; **VIBVU**—*Vibrio vulnificus*; **VIBVY**—*Vibrio vulnificus* (strain YJ016); **YERPE1**—*Yersinia pestis* CO92; **YERPE2**—*Yersinia pestis* KIM. Ending numbers or letters after subscripts highlight different copies of the system in a single organism.


0.1
XAC3241
84
XAC3240
XF0538
100
XF1791
63
82
XF0539
69
96
PD1077
XF0487
50
XF2539
70
79
PD1924
pilB-pilD
cluster
XF2542
80
PD1926
XCC3099
82
XCC3098
XAC2626
91
XF0028
100
PD0019
67
65
46
XAC2669
fimT
62
XCC2495
XF0473
55
100
PD1615
XCC2486
XAC2664
62
PD0024
81
XCC2488
pilE_fimT
cluster
98
XF0479
89
PD1610

Unique genes in XAC and Xf-9a5c/Xf-temecula with respect to XCC. Eighteen of these genes are located inside PinDel-8 of Xf-9a5c and PinDel-4 of XAC, comprising a region that is homologous to the SPI-7 pathogenicity island from *Salmonella thiphymurium* (Fig. 2, squares 8E and 6E) (Moreira et al., 2004; Pickard et al., 2003). Detailed analysis of the SPI-7 region in both genomes suggests probable lateral transfer events in each lineage inside this region: insertion of 37 genes comprising PinDel-7 in Xf-9a5c (XF1718-XF1754), which is located upstream of SPI-7 in Xf-9a5c and is flanked by a tRNA-G, and an added cluster of fourteen genes, including an integrase (XAC2220), inside the SPI-7 region in XAC (Fig. 7). Our analysis has shown that, in Xf-9a5c, 13 genes in PinDel-7 and 14 genes in SPI-7 show atypically high GC content (68%) when compared to the Xf-9a5c genome average (53%). Their homologs in the region SPI-7 from XAC have GC content very close to the average (65%) for XAC genes. In spite of this, and given the absence of SPI-7 both in XCC and Xf-temecula, plus the other evidence presented above, we speculate that the presence of SPI-7 in XAC and XF-9a5c resulted from two separate lateral transfer events; this is consistent with Pickard et al.'s hypothesis (Pickard et al., 2003) that SPI-7 may be a mobile element.

F7

Other CDSs in the intersection of XAC and Xf-9a5c gene sets include three conserved hypothetical proteins: a transcriptional regulator, a plasmid stabilization protein, and a cytochrome B561-like gene. A single gene was classified as unique to the intersection between the XAC and Xf-temecula gene sets: CDP-diacylglycerol-glycerol-3-phosphate3-phosphatidyltransferase, which catalyzes a committed step to the synthesis of the acidic phospholipids, and is located near transposases and phage-related genes in Xf-temecula.

Genes unique to XCC and Xf-9a5c/Xf-temecula with respect to XAC. Four genes are unique to XCC and Xf-9a5c gene sets: three conserved hypothetical proteins and a phage-related protein. Only three genes are unique to XCC and Xf-temecula: two phage-related proteins, which are duplicated in XCC, and one hypothetical protein.

Xanthomonadaceae unique genes. These were selected based on published results. Among the most important of the genes unique to the *Xanthomonadaceae* are those responsible for the synthesis of DFS (Diffusible Signal Factor) signaling molecules, the *rpf* cluster (regulation of pathogenicity factors) (Barber et al., 1997; Tang et al., 1991). Inside the *rpf* operon, XAC differs from XCC because it does not have the *rpfH* and *rpfI* genes. The insertion of two transposases between the genes *recJ* and *prfB* may have replaced *rpfI* (Dow et al., 2000), accompanied by the insertion of a gene coding for a wall-associated protein (*wapA*) of the family *rhs*, composed by nine copies of the *rhs* repeat (Apweiler et al., 2000). The *rhs* domain can act as a recombination hot spot (Minet and Chiquet-Ehrismann, 2000) and was related to duplications and rearrangements in *E. coli* (Lin et al., 1984). *Rpf* clusters in both *Xylella* strains lack copies of the *rpfI*, *rpfH* and *rpfD* genes, in addition showing a rupture of this cluster between genes *rpfF* and *rpfB* (da Silva et al., 2001; Moreira et al., 2004). *Rpf* genes also regulate expression of gum genes, responsible for the synthesis and exporting of the xanthan gum, an extracellular polysaccharide important to pathogenesis (Vojnov et al., 2001). Gum genes from *Xanthomonas* strains show a high degree of similarity, with 98% of identity for pairs of orthologs, while in *Xylella*, identities between genes from different pathovars decrease to values between 65 and 83%, with absence of the genes *gumGIL*, responsible for joining and modification of mannoses residues (da Silva et al., 2001).

FIG. 6. (**A**) Unrooted phylogenetic tree for the genes *pilE-fimT* cluster. (**B**) Unrooted phylogenetic tree for the copies of *pilE* unique to Xf-9a5c. Unrooted phylogenetic tree for the genes *pilE* and *fimT* found in XAC, XCC, Xf-9a5c, and Sf-temecula, based on their encoded aminoacid sequences. Horizontal lines separate copies of *pilE* located at the clusters *pilBD* or *pilE_fimT* and the copies of *fimT*. All unique copies of *pilE* found in Xf-9a5c and Xf-temecula are members of a cluster that includes all copies of *pilE* from the *pilB-pilD* cluster in the four bacteria. Note that *pilE* is also duplicated inside the cluster *pilBD* in all lineages. This phylogeny was built using maximum likelihood and a Poisson model of amino acid substitution. ORFs were selected based on the presence of a conserved amino-terminal pilin signature and some truncated sequences were later discarded. Although the resulting alignment, after selection of conserved regions, was too short (40 aa), we inspected phylogenies based on different selection of homologs and observed that this phylogeny reproduces all important features obtained when using longer alignments.

FIG. 7. Analysis of the region comprising PinDel-6 of XAC and PinDel-5 of XF-9a5c. Analysis of the region comprising PinDel-6 of XAC and PinDel-5 of XF-9a5c, which corresponds to island SPI-7 described for *Salmonella typhimurium* (Pickard et al., 2003). The region between genes XF1718 and XF1754 corresponds to PinDel-4 in XF-9a5c. Gray rectangles, genes homologous to genes in the SPI-7 island; white rectangles, genes with function not related to this analysis; black rectangles, phage integrases; black arrows, tRNA genes. Genes with atypically high GC content in Xf-9a5c.

The xanthomonadins cluster is another group of genes specific to the *Xanthomonadacea* (Starr et al., 1977). These genes code for proteins involved in the synthesis of pigments useful for protection against photobiological damage (Jenkins and Starr, 1985; Poplawsky and Chun, 1997; Poplawsky et al., 2000; Rajagopal et al., 1997). XCC and XAC xanthomonadin clusters have the same structure described for *Xanthomonas oryzae pv. oryzae* (XOO) (Goel et al., 2002) (Fig. 2, squares 14E and 14I), except for the first two genes, located away from the xanthomonadin operon in both genomes. Upstream to XAC and XCC genes homologous to genes 3-12 in XOO (referred to as *large cluster* in Fig. 8), a new cluster of genes, which we call the *small cluster*, may be associated with xanthomonadins biosynthesis, because genes in this relatively conserved cluster are related to fatty acids and phosphatidic acid biosynthesis. An interesting feature shown in Figure 8 is the extensive modifications of the region containing the two clusters in XOO. The entire small cluster was either lost or translocated in XOO, and the same happened to isolated genes around and inside the large cluster. In particular, the genes XAC4094, XAC4095, and XCC4006 are truncated copies of the N-terminal region of the homologous XAC4093, XCC4005 and the CDS 4 from XOO. The gene flanking the xanthomonadins cluster in XOO, CDS14, is homologous to a gene located in the XcUC1 region, close to the tandem duplication of xanthomonadins-related genes inside this region.

F8

In both *Xylella* genomes, the small cluster is located far downstream from the large cluster. The small cluster is flanked by phage XFP4 in Xf-9a5c and by PinDel-6 in Xf-temecula. The breakpoint for the separation of the two clusters in both *Xylella* genomes corresponds to the location of the repetitive non-coding DNA sequences in XOO. Three CDSs in both Xf-9a5c clusters have frameshifts and therefore are considered to be pseudogenes, suggesting degeneration of this system (Moreira et al., 2004). Homologs of these pseudogenes in Xf-temecula show no frameshifts.

Finally, although genes *tonB*, *exbB*, and *exbD*, responsible for iron uptake, form a cluster in many bacteria and a second copy of *exbD* (*exbD2*) is common, the *Xanthomonadaceae* are unusual for having all these genes in a single cluster, with *exbD2* located downstream to *tonB-exbB-exbD*. Knock-out of exbD2 in XCC induces hypersensitive response from the host (Wiggerich and Puhler, 2000), but differently from the other genes in this cluster, it is not essential for penetration of phage ΦL7 and shows no impact on iron uptake (Hung et al., 2003; Wiggerich et al., 1997).

DISCUSSION

Differences in gene content between *Xanthomonas* and *Xylella* complete genome sequences have been characterized previously (Van Sluys, M.A., et al., 2003; da Silva, et al., 2002; Moreira, LM, et al., 2004). The data presented here complements such analyses with an extensive description of important genome rearrangements and a comparison of the putative functions of all genes leading to variations in gene content among the four strains analyzed. As presented above and explored in previous works, all four bacteria studied here share some characteristic systems that are commonly used to identify members of their family, and were shown to have important roles in their survival and adaptation. For instance, the gum operon, the extended xanthomonadin cluster, and the *rpf* genes, although present in all four bacteria, show important variations in gene order and gene content among these organisms, and these variations are likely to be a factor in each strain's phenotype (da Silva et al., 2002).

Following a trend seen in other shared systems, copies of the xanthomonadin cluster in *Xylella* strains are highly fragmented when compared to the structure seen in *Xanthomonas*. In Xf-9a5c, the fragmentation of the cluster is paralleled by the presence of three pseudogenes inside the large and small clusters. Although Xf-temecula has intact copies of the Xf-9a5c pseudogenes, experimental data has shown that both *Xylella* strains do not synthesize the xanthomonadin pigment (Almeida et al., 2004; Almeida and Purcell, 2003; Chang and Donaldson, 2000; Leite et al., 2004). Considering the changes in the xanthomonadin clusters from XAC and XCC to Xf-temecula and Xf-9a5c, the trend represented in Figure 8 that we wish to highlight is an apparent progressive depletion of the xanthomonadin cluster along the evolution of *Xylella*.

A similar fragmentation is observed for the *rpf* gene cluster in *Xylella* (Moreira et. al., 2004). It was suggested that the fragmented structure of the *rpf* cluster in *Xylella* might induce a reduction of DSF synthesis, thus leading to reduced quorum sensing signaling and no feedback for regulation of expression of the

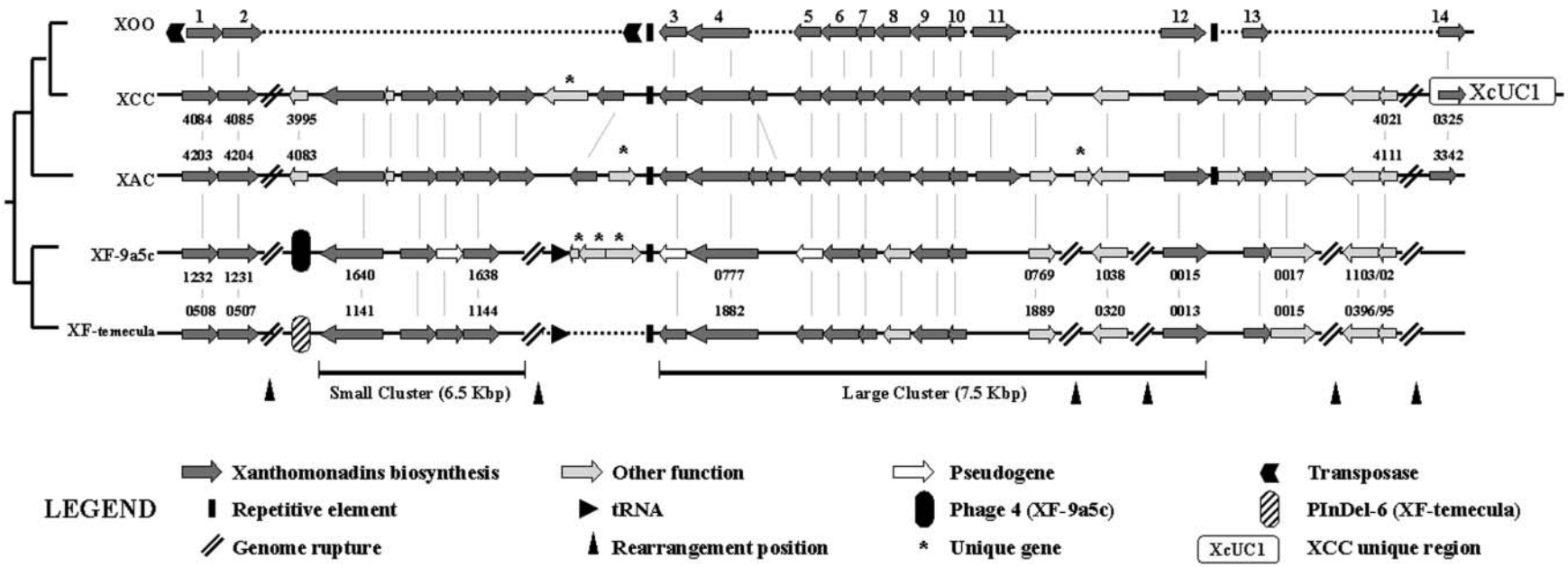


FIG. 8. Analysis of genes predicted to be involved with Xanthomonadins synthesis. Analysis of genes predicted to be involved with Xanthomonadins synthesis in XCC, XAC, XF-9a5c, and XF-temecula, using *Xanthomonas oryzae* pv *oryzae* (XOO) as model, following Goel et al. (2002). The dendrogram on the left shows the 16S phylogeny. Although XOO and XCC are closer phylogenetically, the gene organization of XCC is more similar to that of XAC. This suggests a rearrangement or deletion of the "small cluster" in XOO. A transposase and a repetitive element present upstream of the large cluster in XOO are evidence in favor of this hypothesis. It can be seen in this comparison that the *Xylella* genomes appear to have been subject to many rearrangements in this region, consistent with data shown in Figure 1. This analysis suggests that Xf-9a5c phage 4 is related to PinDel-6 in Xf-temecula; note also that XOO gene 14 has an ortholog in XcUC1.

rpf genes themselves (Moreira et al., 2004). We note that the hypothesis of Moreira et al., (2004) for the consequences of *rpf* fragmentation is compatible with the relative concentration of DSF in *Xylella* and *Xanthomonas* cultures (Scarpari et al., 2003). Such loss of feedback control over *rpf* genes probably reduces *Xylella*'s virulence, because it lowers the level of chemical signals that activate gum synthesis. With a reduced DSF-mediated quorum sensing signaling to *rpf* and living in a low sugar milieu (the xylem), it is likely that *Xylella* either posses another system for gum synthesis regulation or express this system constitutively, differently from *Xanthomonas*, which are regulated by sugar concentration (Vojnov et al., 2001). At any rate, such structural differences in the *rpf* operon point to the relevance that changes in gene order might have over the phenotype of bacteria. The presence of a complete PTS system only in *Xanthomonas*, associated to the presence of clusters involved in degradation of plant cell-wall and sugar storage, suggests that *Xanthomonas* might use degraded cell walls as carbon sources, another function probably absent from *Xylella* cells, due to the degradation of the PTS system and a reduced number of cell-wall degrading enzymes (CWDEs) in *Xylella*. The presence of genes encoding the enzyme 2-keto-3-deoxygluconate kinase, along with genes for CWDEs, suggests that *Xanthomonas* may use the CWDEs to obtain energy from host cell-wall degradation.

Some systems were found to be complete in *Xylella* and reduced in *Xanthomonas*, probably due to the existence of negative selection against the loss of such functions, which are probably necessary for adaptation into the xylem. This is likely to be the case for the presence of a complete operon for colicin production in *Xylella*, which is not found in *Xanthomonas*. The lack of both *cvaA* and *cvaB* genes in XCC is probably a derived state, XAC representing an intermediate depletion state, while *Xylella* probably corresponds to the ancestral set of colicin genes, all being preserved because synthesis of these compounds is essential to avoid competition with other endophytic bacteria of the plant or resident at the insect foregut, as proposed by Pashalidis et al. (2005). In fact, it has been demonstrated that Xf-9a5c lives in an environment in citrus plants that is rich in endophytic bacteria (Araujo et al., 2002; Lacava et al., 2004), and therefore the presence of a complete and functional colicin V biosynthesis and secretion system may enhance its chances of survival in the xylem.

Two other important examples of differences in gene content between *Xanthomonas* and *Xylella* are the presence in *Xanthomonas* of a second copy of the type II secretion systems (T2SS), known as xcs, and of a type III secretion system (T3SS). These systems differ from the systems discussed above for being entirely absent from *Xylella* genomes, instead of partially depleted. Therefore, although one can argue, for the systems discussed above, that a partially degraded copy of a cluster of genes is enough evidence to support a hypothesis of vertical heritage followed by gene loss, the full absence of both secretion systems from *Xylella* might well be explained by loss of the cluster in the *Xylella* lineage or acquisition by *Xanthomonas* after divergence, through lateral gene transfer. In the case of *xcs*, we believe this system was lost from a common ancestor of Xf-temecula and Xf-9a5c. Two pieces of evidence support this hypothesis: (i) the regions flanking this cluster in *Xanthomonas* are homologous to neighboring regions in *Xylella* (Fig. 1, regions A and B) (Moreira et al., 2004); (ii) the phylogeny shown in Figure 5 indicates that copies of the T2SS distributed among several proteobacteria radiate in a pattern that can be easily reconciled with the species tree, by postulating just a few events of gene duplication and loss (Page and Charleston, 1997). In particular, the copy of xcs present in XAC and XCC is placed at the base of the branch that leads to copies of T2SS present in *Pseudomonas* and several enterobacteria, thus reproducing the location of *Xanthomonas* in a species tree (Lerat et al., 2003). On the other hand, the *Caulobacter* T2SS copy is placed in an unexpected location, well inside a γ-proteobacteria group, although this organism belongs to the alpha subdivision of Proteobacteria. Even more suggestive is the grouping of *xps*, the other T2SS cluster found in *Xanthomonas* and *Xylella*, with plant-associated bacteria of the a-proteobacteria subdivision (*Bradyrhizobium japonicum* and *Mesorhizobium loti* in Fig. 5), which suggests that this cluster, instead of *xcs*, has originated from lateral transfer from plant-associated bacteria. We note however that branch lengths in the *xps* subtree leading to *B. japonicum* and *M. loti* are long, thus implicating early divergence of these genes, maybe dating to the divergence among these organisms' lineages. Whatever the origins of *xps* and *xcs*, their phylogeny and sequence alignment show that their divergence is very old, probably dating to the ancestor of all proteobacteria. This suggests that these systems have specialized to perform different physiological roles, and therefore *Xylella* may lack some functionality associated with these systems.

For T3SS, a single BLAST run is able to show how easy it is to find such a system among members of the *Xanthomonas* genus. Sequence and phylogenetic analyses have shown T3SS clusters from members of the β-proteobacteria subdivision to be the closest relatives of the T3SS from XAC and XCC (data not shown). Such a relationship is in agreement with the species tree, but contradicts the fact that the XCC copy of T3SS possesses much lower GC composition and high codon bias when compared to the rest of XCC genes, a feature common to laterally transferred genes, although not unique to them. Loci homologous to regions C and D in *Xylella* (Fig. 1), which flank *Xanthomonas* T3SS, are located near the replication terminus in both *Xylella* strains, a pattern that suggests deletion of this system from *Xylella*. So we cannot, with known evidence, draw conclusions on what is the origin of T3SS found *Xanthomonas*.

Lineage specific expansion and diversification is another evolutionary process involved in variations in gene content between the two genera. This is the mechanism behind the much higher number of chemotactic genes in *Xanthomonas*, and was especially important in the evolution of *tsr* genes. This family of genes has undergone fast expansion in the *Xanthomonas* lineage after separation of the *Xylella* branch (Moreira et al., 2004) (Fig. 4). It is possible that a mechanism of dosage compensation has prompted the very first duplication events of the *tsr* gene in the ancestor of XAC and XCC, as proposed by Heidelberg et al. (2000) in relation to the genes involved in chemotaxis in *Vibrio cholerae*.

The presence of several copies of the structural subunit of type IV pili in *Xylella* may also have involved a similar mechanism of dosage compensation, followed by divergence of some copies (Fig. 4). The presence of two distinct clusters of pili type IV biosynthesis in *Xylella*, as described by Moreira et al. (2004), associated with a complete cluster of chemotaxis genes related to twitching motility, may have a relation to the lifestyle of *Xylella*. Each pili cluster may have a specific function such as, for example, fixation into the xylem vessel and fixation into the insect foregut. In addition, the presence of additional copies of genes for fimbrial adhesins may augment the expression of proteins needed for colonization and cell agglomeration into the plant. A recent analysis of genome composition of pathogenic (CVC) and nonpathogenic *Xylella fastidiosa* isolates showed that the absence of fimbrial adhesins can determine the pathogenic phenotype of an isolate (Koide et al., 2004).

Another mechanism involved in determining variations in gene content among the four bacteria is the process of lateral gene transfer, already referred to when discussing the secretory systems in *Xanthomonas*. Several phage- and plasmid-related genes are present in all four genomes and most certainly were acquired directly from other organisms and phages through conjugation or transfection followed by recombination. We note also that many of the unique genes described for the four genomes that had assigned functions were restriction enzymes, known to be transmitted horizontally (Gelfand and Koonin, 1997; Sekizaki et al., 2001; Sharp et al., 1992). The many transposases seen in *Xanthomonas* might be related to plasmids and therefore be related to the transposition of genes from plasmids into the genome. An important example of a gene likely to have been acquired laterally is the copy of *kdgT*, found only in XAC, and placed near two ISxac3 transposases. This gene is related to pathogenicity, since it is involved in the transport of degraded pectin products in the bacterial cell, where these products may be used as carbon and energy sources.

As a final comment on our detailed characterization of unique gene functions among the two *Xanthomonadaceae* genera analyzed here, some unique genes related to differences in nitrogen metabolism for XCC and XAC and NO resistance in *Xylella* are worth mentioning. The presence of a unique cluster that codes for nitrate assimilation in XCC may have a horizontal origin and may be coupled with the *Xanthomonas* unique cytochrome genes *cydAB*, which have been implicated in nitrogen fixation (da Silva et al., 2001), thus characterizing a unique form of nitrogen uptake among the bacteria analyzed here. In association with the *opp* genes located in the XaUC1 region, the pseudomonapepsin and xanthomonapepsin carboxypeptidases may be part of an oligopeptide absorption and degradation pathway that may be linked to XAC's ability to grow in different environments (da Silva et al., 2002). Since *opp* genes code for components of an ABC permease complex that is associated to an oligopeptide binding protein (*oppA*), it is possible that these genes may help entry of small oligopeptide products for use as a nitrogen source.

In *Xylella* strains, the unique *rocF* gene is homologous to *Helicobacter pylori*'s arginase *rocF*. In *Helicobacter pylori*, this gene inhibits NO production by macrophages at physiological concentrations of arginine. *RocF* expression is aborted at normal levels of NO, indicating that this gene down-regulates NO production, thus acting as a survival mechanism and contributing to the success of infection (Gobert et al.,

2001). Gobert et al. (2002) showed that macrophage apoptosis is induced by activation of the arginase II gene of *H. pylori* (Gobert et al., 2002). Based on such observations, Koide et al. (2004) proposed that absence of this gene in *Xylella fastidiosa* J1a12, a non-pathogenic strain, is linked to reduced growth in the plant and incapacity to colonize the xylem vessels, due to inability to inhibit NO production by the plant host. On the other hand, the hypothesis of Koide et al. (2004) for *Xylella fastidiosa* J1a12 cannot be extended to *Xanthomonas*, since these bacteria lack *rocF* but grow very well inside the plant and, in the case of *Xanthomonas campestris*, are able to colonize vascular tissues. This suggests that either *Xanthomonas* has other NO detoxification systems or that it has alternative routes to use arginine produced by the plant.

Pindels and phages. We have discussed above evolutionary events that may have lead to the observed variations in gene content among XAC, XCC, Xf-9a5c, and Xf-temecula. Two other important processes were characterized regarding the evolution of these genomes: the rearrangements observed between species in each genus and the distribution of putative insertion/deletion islands, referred to as PinDels.

All four genomes present important rearrangements. The *Xanthomonas* genomes have three important rearrangements with respect to each other: an inversion at the terminus of replication and a reciprocal translocation with inversion of the translocated regions (da Silva et al., 2002). The *Xylella* genomes, on the other hand, exhibit a more varied set of rearrangements with respect to each other, including a translocation of the replication terminus (Van Sluys et al., 2003) (Fig. 1). Probably the most important aspect of the *Xylella* genome rearrangements is the correlation between the borders of shuffled regions and the presence of phages or PinDels near such borders, which indicate a possible involvement of phage integrases in the rearrangements observed in these two lineages, as previously suggested by Van Sluys et al. (2003); no similar connection can be made between phage integrase distribution and rearrangements in *Xanthomonas*.

All strains analyzed show a high number of phage-related genes, many of them located inside unique gene islands rich in hypothetical genes. Note however that *Xylella* strains possess much more phage-related and complete integrated phages than *Xanthomonas*, and that *Xanthomonas* integrases appear to have less activity than in *Xylella*, since the higher number of rearrangements in this lineage seems related to phage integrases. Also, most phage-related genes in *Xanthomonas* are located inside unique regions that seem to be remains of degraded phages. Which mechanism is behind the abundance of complete phages integrated into the genomes of the two *Xylella*s? Multiple alignments of all integrases found in both genomes lead to classification of these integrases into groups based on sequence similarity and integration site class, as described in Materials and Methods. In Xf-9a5c, integrases of group ***A*** may be related to insertion on the 3′ end of tRNAs (Campbell, 2003; Williams, 2002), whereas the integrases of group ***B*** may determine prophage termination. Although no experimental work has been done to verify the relation between the pili machinery in Xf-9a5c and this organism's capacity to acquire phage insertions, pili biosynthesis may be an important mediator of the entrance of phages into *Xylella*, as suggested by Bradley (1972, 1974) for *Pseudomonas aeruginosa*. Also, besides mediating the uptake of iron and other compounds, such as cobalamin and heme groups, the tonB and tolA complexes may be involved in interactions with lytic bacteriophages and the bacteria cell host (Braun et al., 1976), and facilitate entrance of phages into the genome of *Xylella*, thus providing another mechanism through which phage invasion may be facilitated in *Xylella*. The abundance of phage-related genes in *Xylella*, as well as the greater abundance of transposases in *Xanthomonas*, may be related to distinct restriction and modification systems in each bacterium. Interestingly, although our initial approach to identify probable insertions was based on phage integrases and site integration, this approach identified nearly the same regions obtained from an analysis based on GC content and codon bias, as shown in Figure 1.

Although molecular mechanisms that promote a higher rate of phage integration into the genome of *Xylella* might be involved in the presence of a higher number of complete phages and phage-related genes, a population genetics explanation could as well account for such differences. With nearly half the genome size of *Xanthomonas*, it is possible that the *Xylella* genome is evolving under a regime of accumulation of deleterious mutations (Crow and Kimura, 1970), both in the form of degraded gene systems and pseudogenes as well as integrated phages. Together with its limited environmental range, restricted to the xylem, slow growth and fastidious phenotype, *Xylella* presents limitations to its propagation derived from the dependence on an insect vector, while *Xanthomonas* can and usually does spread easily from contaminated

to noncontaminated hosts, being able to survive under adverse conditions over leaf surfaces and on the ground (Lopez et al., 1999). As *Xanthomonas* also has a much smaller doubling time both in plants and the laboratory, we believe that the effective population size for *Xanthomonas* strains is much greater than those for *Xylella* strains. If real, the reduction of effective population size for *Xylella* might be regarded as a mechanism that enables fixation of both gene deletion events, leading to a higher number of depleted systems and pseudogenes, and the integration of phages into its genome. Both gene deletion and phage insertion might be considered deleterious events to bacteria and are expected to be removed from sufficiently large populations by means of natural selection. On smaller populations, though, these mutations might become fixed by random genetic drift (Woolfit and Bromham, 2003).

As a final note, a link between two of the main threads of this paper, namely variation in gene content among phylogenetically close genomes and the observed correlation between phage-related genes and regions rich in unique genes, is provided by recent analyses that show that phage-bacteria interactions are important mechanisms for new gene creation in bacterial hosts (Daubin and Ochman, 2004).

ACKNOWLEDGMENTS

This work was supported in part by Fundação de Amparo à Pesquisa do Estado de São Paulo, FAPESP (Ph.D. fellowship to L.M.M.), by Conselho Nacional de Desenvolvimento CientÌfico e Tecnológico, CNPq (Ph.D. fellowship to R.F.S.), and by Coordenação para Aperfeiçoamento de Pessoal de Ensino Superior, CAPES (Ph.D. fellowship to L.A.D.), all of them Brazilian funding agencies.

REFERENCES

ADACHI, J., and HASEGAWA, M. (1996). *MOLPHY Version 2.3: Programs for Molecular Phylogenetics Based on Maximum Likelihood*. J. Adachi, ed. (Institute of Statistical Mathematics, Tokyo).

ALMRA, and MATTICK, J.S. (1997). Genes involved in the biogenesis and function of type-4 fimbriae in *Pseudomonas aeruginosa*. Gene **192,** 89–98.

ALMEIDA, R.P., MANN, R., and PURCELL, A.H. (2004). *Xylella fastidiosa* cultivation on a minimal solid defined medium. Curr Microbiol **48,** 368–372.

ALMEIDA, R.P., and PURCELL, A.H. (2003). Biological traits of *Xylella fastidiosa* strains from grapes and almonds. Appl Environ Microbiol **69,** 7447–7452.

ALTSCHUL, S.F., MADDEN, T.L., SCHAFFER, A.A., et al. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res **25,** 3389–3402.

APWEILER, R., ATTWOOD, T.K., BAIROCH, A., et al. (2000). InterPro: an integrated documentation resource for protein families, domains and functional sites. Bioinformatics **16,** 1145–1150.

APWEILER, R., BAIROCH, A., WU, C.H., et al. (2004). UniProt: the Universal Protein knowledgebase. Nucleic Acids Res. **32,** D115–D119.

ARAUJO, W.L., MARCON, J., MACCHERONI, W., JR., et al. (2002). Diversity of endophytic bacterial populations and their interaction with *Xylella fastidiosa* in citrus plants. Appl Environ Microbiol **68,** 4906–4914.

BARBER, C.E., TANG, J.L., FENG, J.X., et al. (1997). A novel regulatory system required for pathogenicity of *Xanthomonas campestris* is mediated by a small diffusible signal molecule. Mol Microbiol **24,** 555–566.

BHATTACHARYYA, A., STILWAGEN, S., IVANOVA, N., et al. (2002a). Whole-genome comparative analysis of three phytopathogenic *Xylella fastidiosa* strains. Proc Natl Acad Sci USA **99,** 12403–12408.

BHATTACHARYYA, A., STILWAGEN, S., REZNIK, G., et al. (2002b). Draft sequencing and comparative genomics of *Xylella fastidiosa* strains reveal novel biological insights. Genome Res **12,** 1556–1563.

BRADLEY, D.E. (1972). Shortening of *Pseudomonas aeruginosa* pili after RNA-phage adsorption. J Gen Microbiol **72,** 303–319.

BRADLEY, D.E. (1974). The adsorption of *Pseudomonas aeruginosa* pilus-dependent bacteriophages to a host mutant with nonretractile pili. Virology **58,** 149–163.

BRAUN, V., and BRAUN, M. (2002). Iron transport and signaling in *Escherichia coli*. FEBS Lett **529,** 78–85.

BRAUN, V., HANCOCK, R.E., HANTKE, K., et al. (1976). Functional organization of the outer membrane of *Escherichia coli*: phage and colicin receptors as components of iron uptake systems. J Supramol Struct **5,** 37–58.

BUYER, J.S., and LEONG, J. (1986). Iron transport–mediated antagonism between plant growth-promoting and plant-deleterious *Pseudomonas strains*. J Biol Chem **261,** 791–794.

CAMPBELL, A. (2003). Prophage insertion sites. Res Microbiol **154,** 277–282.

CASTRESANA, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol Biol Evol **17,** 540–552.

CHANG, C.J., and DONALDSON, R.C. (2000). Nutritional requirements of *Xylella fastidiosa*, which causes Pierce's disease in grapes. Can J Microbiol **46,** 291–293.

CHANG, C.J., GARNER, M., ZREIK, L., et al. (1993). Culture and serological detection of the xylem-limited bacterium causing citrus variegated chlorosis and identification as a strain of *Xylella fastidiosa*. Curr Microbiol **27,** 137–142.

COLLMER, A., BADEL, J.L., CHARKOWSKI, A.O., et al. (2000). *Pseudomonas syringae* Hrp type III secretion system and effector proteins. Proc Natl Acad Sci USA **97,** 8770–8777.

CORNELIS, P., and MATTHIJS, S. (2002). Diversity of siderophore-mediated iron uptake systems in fluorescent pseudomonads: not only pyoverdines. Environ Microbiol **4,** 787–798.

CROW, J.F., and KIMURA, M. (1970). *An Introduction to Population Genetics Theory* (Harper and Row, New York).

DA SILVA, A.C., FERRO, J.A., REINACH, F.C., et al. (2002). Comparison of the genomes of two *Xanthomonas* pathogens with differing host specificities. Nature **417,** 459–463.

DA SILVA, F.R., VETTORE, A.L., KEMPER, E.L., et al. (2001). Fastidian gum: the *Xylella fastidiosa* exopolysaccharide possibly involved in bacterial pathogenicity. FEMS Microbiol Lett **203,** 165–171.

DAUBIN, V., and OCHMAN, H. (2004). Start-up entities in the origin of new genes. Curr Opin Genet Dev **14,** 616–619.

DELVER, E.P., KOTOVA, V.U., ZAVILGELSKY, G.B., et al. (1991). Nucleotide sequence of the gene (ard) encoding the antirestriction protein of plasmid colIb-P9. J Bacteriol **173,** 5887–5892.

DIGIAMPIETRI, L.A., MEDEIROS, C.B., and SETUBAL, J.C. (2003). A data model for comparative genomics. Rev Tecnol Inform **3,** 35–40.

DOW, J.M., FENG, J.X., BARBER, C.E., et al. (2000). Novel genes involved in the regulation of pathogenicity factor production within the *rpf* gene cluster of *Xanthomonas campestris*. Microbiology **146,** 885–891.

FATH, M.J., MAHANTY, H.K., and KOLTER, R. (1989). Characterization of a purF operon mutation which affects colicin V production. J Bacteriol **171,** 3158–3161.

GELFAND, M.S., and KOONIN, E.V. (1997). Avoidance of palindromic words in bacterial and archaeal genomes: a close connection with restriction enzymes. Nucleic Acids Res **25,** 2430–2439.

GOBERT, A.P., CHENG, Y., WANG, J.Y., et al. (2002). *Helicobacter pylori* induces macrophage apoptosis by activation of arginase II. J Immunol **168,** 4692–4700.

GOBERT, A.P., McGEE, D.J., AKHTAR, M., et al. (2001). *Helicobacter pylori* arginase inhibits nitric oxide production by eukaryotic cells: a strategy for bacterial survival. Proc Natl Acad Sci USA **98,** 13844–13849.

GOEL, A.K., RAJAGOPAL, L., NAGESH, N., et al. (2002). Genetic locus encoding functions involved in biosynthesis and outer membrane localization of xanthomonadin in *Xanthomonas oryzae* pv. *oryzae*. J Bacteriol **184,** 3539–3548.

GOMEZ, P., RIBAS-APARICIO, R.M., PELAEZ, A.I., et al. (1997). Isolation and nucleotide sequence of the gene encoding the XamI DNA methyltransferase of *Xanthomonas campestris* pv. *amaranthicola*. Biochim Biophys Acta **1351,** 261–266.

GOTTWALD, T., and GRAHAM, J. (2000). Bacteria diseases. In *Compendium of Citrus Diseases*, L.W. Timmer, S.M. Garnsey, and J.H. Graham, eds. (APS Press, St. Paul).

GRAHAM, J.H., GOTTWALD, T.R., CUBERO, J., et al. (2004). *Xanthomonas axonopodis* pv. *citri*: factors affecting successful eradiction of citrus canker. Mol Plant Pathol **5,** 1–15.

HEIDELBERG, J.F., EISEN, J.A., NELSON, W.C., et al. (2000). DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*. Nature **406,** 477–483.

HENDRICKS, E.C., SZERLONG, H., HILL, T., et al. (2000). Cell division, guillotining of dimer chromosomes and SOS induction in resolution mutants (dif, xerC and xerD) of *Escherichia coli*. Mol Microbiol **36,** 973–981.

HENDSON, M., PURCELL, A.H., CHEN, D., et al. (2001). Genetic diversity of Pierce's disease strains and other pathotypes of *Xylella fastidiosa*. Appl Environ Microbiol **67,** 895–903.

HOLT, J.G. (1994). *Xylella*. *Bergey's Manual of Determinative Bacteriology*, 9th ed. J.G. Holt, ed. (Williams and Wilkins, London), pp. 100–115.

HOPKINS, D.L. (1989). *Xylella fastidiosa*: xylem-limited bacaterial pathogen of plants. Annu Rev Phytopathol **27,** 271–290.

HOUSBY, J.N., THOMAS, J.D., WHARAM, S.D., et al. (1998). Conditional mutations in OutE and OutL block exoenzyme secretion across the *Erwinia carotovora* outer membrane. FEMS Microbiol Lett **165,** 91–102.

HUNG, C.H., YANG, C.F., YANG, C.Y., et al. (2003). Involvement of tonB-exbBD1D2 operon in infection of *Xanthomonas campestris* phage phi L7. Biochem and Biophys Res Commun **302,** 878–884.

JENKINS, C.L., and STARR, M.P. (1985). Formation of halogenated aryl-polyene (xanthomonadin) pigments by the type and other yellow-pigmented strains of *Xanthomonas maltophilia*. Ann l'Institut Pasteur Microbiol **136B,** 257–264.

KANEHISA, M., GOTO, S., KAWASHIMA, S., et al. (2004). The KEGG resource for deciphering the genome. Nucleic Acids Res **32,** D277–D280.

KARLIN, S. (2001). Detecting anomalous gene clusters and pathogenicity islands in diverse bacterial genomes. Trends Microbiol **9,** 335–343.

KELLY, M.J., POOLE, R.K., YATES, M.G., et al. (1990). Cloning and mutagenesis of genes encoding the cytochrome bd terminal oxidase complex in *Azotobacter vinelandii*: mutants deficient in the cytochrome d complex are unable to fix nitrogen in air. J Bacteriol **172,** 6010–6019.

KHALIL-RIZVI, S., TOTH, S.I., VAN DER HELM, D., et al. (1997). Structures and characteristics of novel siderophores from plant deleterious *Pseudomonas fluorescens* A225 and *Pseudomonas putida* ATCC 39167. Biochemistry **36,** 4163–4171.

KOIDE, T., ZAINI, P.A., MOREIRA, L.M., et al. (2004). DNA microarray-based genome comparison of a pathogenic and a nonpathogenic strain of *Xylella fastidiosa* delineates genes important for bacterial virulence. J Bacteriol **186,** 5442–5449.

KOSTER, M., OVAA, W., BITTER, W., et al. (1995). Multiple outer membrane receptors for uptake of ferric pseudobactins in *Pseudomonas putida* WCS358. Mol Gen Genet **248,** 735–743.

LACAVA, P.T., ARAUJO, W.L., MARCON, J., et al. (2004). Interaction between endophytic bacteria from citrus plants and the phytopathogenic bacteria *Xylella fastidiosa*, causal agent of citrus-variegated chlorosis. Lett Appl Microbiol **39,** 55–59.

LAZDUNSKI, C.J., BOUVERET, E., RIGAL, A., et al. (1998). Colicin import into *Escherichia coli* cells. J Bacteriol **180,** 4993–5002.

LEE, H.M., CHEN, J.R., LEE, H.L., et al. (2004). Functional dissection of the XpsN (GspC) protein of the *Xanthomonas campestris* pv. *campestris* type II secretion machinery. J Bacteriol **186,** 2946–2955.

LEE, H.M., WANG, K.C., LIU, Y.L., et al. (2000). Association of the cytoplasmic membrane protein XpsN with the outer membrane protein XpsD in the type II protein secretion apparatus of *Xanthomonas campestris* pv. *campestris*. J Bacteriol **182,** 1549–1557.

LEE, R.F.E.A., BERETTA, M.J.G., HARTUNG, J.H., et al. (1993). Citrus variegated chlorosis: confirmation of a *Xylella fastidiosa* as the causal agent. Summa Phytopathol **19.2,** 123–125.

LEITE, B., ANDERSEN, P.C., and ISHIDA, M.L. (2004). Colony aggregation and biofilm formation in xylem chemistry-based media for *Xylella fastidiosa*. FEMS Microbiol Lett **230,** 283–290.

LERAT, E., DAUBIN, V., and MORAN, N.A. (2003). From gene trees to organismal phylogeny in prokaryotes: the case of the gamma-Proteobacteria. PLoS Biol **1,** E19.

LIMA, W.C., PAQUOLA, A.C.M., VAN SLUYS, M.-A., et al. (2005). Non-gamma proteobacteria gene islands contribution to *Xanthomonas* genome. OMICS (this issue).

LIN, R.J., CAPAGE, M., and HILL, C.W. (1984). A repetitive DNA sequence, rhs, responsible for duplications within the *Escherichia coli* K-12 chromosome. J Mol Biol **177,** 1–18.

LOBRY, J.R. (1996). Asymmetric substitution patterns in the two DNA strands of bacteria. Mol Biol Evol **13,** 660–665.

LONG, S.R., and STASKAWICZ, B.J. (1993). Prokaryotic plant parasites. Cell **73,** 921–935.

LOPES, J.R.S., BERETTA, M.J.G., HARAKAVA, R., et al. (1996). Confirmação da transmissão por cigarrinhas do agente causal da clorose variegada dos citros, *Xylella fastidiosa*. Fitopatol Brasil **21**(supplement), 343.

LOPEZ, N.I., HAEDO, A.S., and MENDEZ, B.S. (1999). Evaluation of *Xanthomonas campestris* survival in a soil microcosm system. Int Microbiol **2**, 111–114.

MINET, A.D., and CHIQUET-EHRISMANN, R. (2000). Phylogenetic analysis of teneurin genes and comparison to the rearrangement hot spot elements of *E. coli.* Gene **257,** 87–97.

MOREIRA, L.M., DE SOUZA, R.F., ALMEIDA, N.F., JR., et al. (2004). Comparative genomics analyses of citrus-associated bacteria. Annu Rev Phytopathol **42,** 163–184.

NOEL, L., THIEME, F., GABLER, J., et al. (2003). XopC and XopJ, two novel type III effector proteins from *Xanthomonas campestris* pv. *vesicatoria*. J Bacteriol **185,** 7092–7102.

NUNES, L.R., ROSATO, Y.B., MUTO, N.H., et al. (2003). Microarray analyses of *Xylella fastidiosa* provide evidence of coordinated transcription control of laterally transferred elements. Genome Res **13,** 570–578.

PAGE, R.D., and CHARLESTON, M.A. (1997). From gene to organismal phylogeny: reconciled trees and the gene tree/species tree problem. Mol Phylogenet Evol **7,** 231–240.

PASHALIDIS, S., MOREIRA, L.M., ZAINI, P.A., et al. (2005). Whole-genome expression profiling of *Xylella fastidiosa* in response to growth on glucose. OMICS (this issue).

PENDRAK, M.L., and PERRY, R.D. (1991). Characterization of a hemin-storage locus of *Yersinia pestis*. Biol Met **4,** 41–47.

PICKARD, D., WAIN, J., BAKER, S., et al. (2003). Composition, acquisition, and distribution of the Vi exopolysaccharide-encoding *Salmonella enterica* pathogenicity island SPI-7. J Bacteriol **185,** 5055–5065.

POPLAWSKY, A.R., and CHUN, W. (1997). pigB determines a diffusible factor needed for extracellular polysaccharide slime and xanthomonadin production in *Xanthomonas campestris* pv. *campestris*. J Bacteriol **179,** 439–444.

POPLAWSKY, A.R., URBAN, S.C., and CHUN, W. (2000). Biological role of xanthomonadin pigments in *Xanthomonas campestris* pv. *campestris*. Appl Environ Microbiol **66,** 5123–5127.

PURCELL, A.H. (1980). Almond leaf scorch: leafhooper and spittle bug vectors. J Econ Entomol **73,** 834–838.

RAJAGOPAL, L., SUNDARI, C.S., BALASUBRAMANIAN, D., et al. (1997). The bacterial pigment xanthomonadin offers protection against photodamage. FEBS Lett **415,** 125–128.

ROSSETTI, V., GARNIER, M., BOVÉ, J.M., et al. (1990). Presence de bacteries dans le xyleme d'oranges atteints de chlorose variegee, une nouvelle maladie des agrumes au Bresil. CRAcad Paris **310,** 345–349.

SCARPARI, L.M., LAMBAIS, M.R., SILVA, D.S., et al. (2003). Expression of putative pathogenicity-related genes in *Xylella fastidiosa* grown at low and high cell density conditions *in vitro*. FEMS Microbiol Lett **222,** 83–92.

SCHAAD, N.W., and ALVAREZ, A. (1993). *Xanthomonas campestris* pv. *campestris*: cause of black rot of crucifers. In *Xanthomonas*, 1st ed. J.G. Swings and E.L. Civerolo, eds. (London.), pp. 51–53.

SCHAAD, N.W., POSTNIKOVA, E., LACY, G., et al. (2004). *Xylella fastidiosa* subspecies: *X. fastidiosa* subsp piercei, subsp. nov., *X. fastidiosa* subsp. multiplex subsp. nov., and *X. fastidiosa* subsp. pauca subsp. nov. Syst Appl Microbiol **27,** 290–300.

SCHUBERT, S., RAKIN, A., KARCH, H., et al. (1998). Prevalence of the "high-pathogenicity island" of *Yersinia* species among *Escherichia coli* strains that are pathogenic to humans. Infect Immun **66,** 480–485.

SEKIZAKI, T., OTANI, Y., OSAKI, M., et al. (2001). Evidence for horizontal transfer of SsuDAT1I restriction-modification genes to the *Streptococcus suis* genome. J Bacteriol **183,** 500–511.

SHARP, P.M., KELLEHER, J.E., DANIEL, A.S., et al. (1992). Roles of selection and recombination in the evolution of type I restriction-modification systems in enterobacteria. Proc Natl Acad Sci USA **89,** 9836–9840.

SIMPSON, A.J., REINACH, F.C., ARRUDA, P., et al. (2000). The genome sequence of the plant pathogen *Xylella fastidiosa*. The *Xylella fastidiosa* Consortium of the Organization for Nucleotide Sequencing and Analysis. Nature **406,** 151–157.

SIVONEN, K., NAMIKOSHI, M., EVANS, W.R., et al. (1992). Three new microcystins, cyclic heptapeptide hepatotoxins, from *Nostoc* sp. strain 152. Chem Res Toxicol **5,** 464–469.

STARR, M.P., JENKINS, C.L., Bussey, L.B., et al. (1977). Chemotaxonomic significance of the xanthomonadins, novel brominated aryl-polyene pigments produced by bacteria of the genus *Xanthomonas*. Arch Microbiol **113,** 1–9.

STEINMANN, D., KOPLIN, R., PUHLER, A., et al. (1997). *Xanthomonas campestris* pv. *campestris* lpsI and lpsJ genes encoding putative proteins with sequence similarity to the alpha- and beta-subunits of 3–oxoacid CoA-transferases are involved in LPS biosynthesis. Arch Microbiol **168,** 441–447.

STOCK, J., and SURETTE, M. (1994). Bacterial chemotaxis. The motor connection. Curr Biol **4,** 143–144.

STOCK, J., SURETTE, M., and PARK, P. (1994). Chemosensing and signal transduction in bacteria. Curr Opin Neurobiol **4,** 474–480.

STOCK, J.B., SURETTE, M.G., McCLEARY, W.R., et al. (1992). Signal transduction in bacterial chemotaxis. J Biol Chem **267,** 19753–19756.

STROM, M.S., and LORY, S. (1993). Structure-function and biogenesis of the type IV pili. Annu Rev Microbiol **47,** 565–596.

SWINGS, J.G., and CIVEROLO, E.L. (1993). *Xanthomonas*, 1st ed. (Chapman & Hall, London).

TANG, J.L., LIU, Y.N., BARBER, C.E., et al. (1991). Genetic and molecular analysis of a cluster of *rpf* genes involved in positive regulation of synthesis of extracellular enzymes and polysaccharide in *Xanthomonas campestris* pathovar *campestris*. Mol Gen Genet **226,** 409–417.

TATUSOV, R.L., NATALE, D.A., GARKAVTSEV, I.V., et al. (2001). The COG database: new developments in phylogenetic classification of proteins from complete genomes. Nucleic Acids Res **29,** 22–28.

THOMAS, J.D., REEVES, P.J., and SALMOND, G.P. (1997). The general secretion pathway of *Erwinia carotovora* subsp. *carotovora*: analysis of the membrane topology of OutC and OutF. Microbiology **143**, 713–720.

THOMPSON, J.D., HIGGINS, D.G., and GIBSON, T.J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res **22,** 4673–4680.

VAN SLUYS, M.A., DE OLIVEIRA, M.C., MONTEIRO-VITORELLO, C.B., et al. (2003). Comparative analyses of the complete genome sequences of Pierce's disease and citrus variegated chlorosis strains of *Xylella fastidiosa*. J Bacteriol **185,** 1018–1026.

VAN SLUYS, M.A., MONTEIRO-VITORELLO, C.B., CAMARGO, L.E., et al. (2002). Comparative genomic analysis of plant-associated bacteria. Annu Rev Phytopathol **40,** 169–189.

AU1

VOJNOV, A.A., SLATER, H., NEWMAN, M.A., et al. (2001). Regulation of the synthesis of cyclic glucan in *Xanthomonas campestris* by a diffusible signal molecule. Arch Microbiol **176,** 415–420.

VORHOLTER, F.J., NIEHAUS, K., and PUHLER, A. (2001). Lipopolysaccharide biosynthesis in *Xanthomonas campestris* pv. *campestris*: a cluster of 15 genes is involved in the biosynthesis of the LPS O-antigen and the LPS core. Mol Genet Genomics **266,** 79–95.

WATERS, V.L., and CROSA, J.H. (1991). Colicin V virulence plasmids. Microbiol Rev **55,** 437–450.

WELLS, J.M., RAJU, B.C., HUNG, H.Y., et al. (1987). *Xylella fastidiosa* gen. nov., sp. nov.: gram-negative, xylem-limited, fastidious plant bacteria related to *Xanthomonas* spp. Int J Syst Bacteriol **37,** 136–143.

WELLS, J.M., RAJU, B.C., THOMPSON, J.M., et al. (1981). Etiology of phony peach and plum leaf scald diseases. Phytopathology **71,** 1156–1161.

WIGGERICH, H.G., KLAUKE, B., KOPLIN, R., et al. (1997). Unusual structure of the tonB-exb DNA region of *Xanthomonas campestris* pv. *campestris*: tonB, exbB, and exbD1 are essential for ferric iron uptake, but exbD2 is not. J Bacteriol **179,** 7103–7110.

WIGGERICH, H.G., and PUHLER, A. (2000). The exbD2 gene as well as the iron-uptake genes tonB, exbB and exbD1 of *Xanthomonas campestris* pv. *campestris* are essential for the induction of a hypersensitive response on pepper (*Capsicum annuum*). Microbiology **146,** 1053–1060.

WILLIAMS, K.P. (2002). Integration sites for genetic elements in prokaryotic tRNA and tmRNA genes: sublocation preference of integrase subfamilies. Nucleic Acids Res **30,** 866–875.

WOOLFIT, M., and BROMHAM, L. (2003). Increased rates of sequence evolution in endosymbiotic bacteria and fungi with small effective population sizes. Mol Biol Evol **20,** 1545–1555.

Address reprint requests to:
Dr. João C. Setubal
Virginia Bioinformatics Institute and Department of Computer Science
Virginia Polytechnic Institute and State University
Bioinformatics I
Blacksburg, VA 24060–0477

E-mail: setubal@vbi.vt.edu.

MOREIRA

AU1
For Schaad & Alvarez ref, please provide name of publisher.

AU2
Are all column headings as you meant in Table 3?