OMICS A Journal of Integrative Biology
Volume 9, Number 2, 2005

Xylella and *Xanthomonas* Mobil'omics

CLAUDIA B. MONTEIRO-VITORELLO,[1] MARIANA C. DE OLIVEIRA,[2] MARCELO M. ZERILLO,[3] ALESSANDRO M. VARANI,[1] EDWIN CIVEROLO,[4] and MARIE-ANNE VAN SLUYS[2]

ABSTRACT

The gamma-proteobacterium Xanthomonadales groups two closely related genera of plant pathogens, *Xanthomonas* and *Xylella*. Whole genome sequencing and comparative analyses disclosed a high degree of identity and co-linearity of the chromosome backbone between species and strains. Differences observed are usually clustered into genomic islands, most of which are delimited by genetic mobile elements. Focus is given in this paper to describe which groups of mobile elements are found and what is the relative contribution of these elements to *Xanthomonas* and *Xylella* genomes. Insertion sequence (IS) elements have invaded the *Xanthomonas* genome several times, whereas *Xylella* is rich in phage-related regions. Also, different plasmids are found inhabiting the bacterial cells studied here. Altogether, these results suggest that the integrative elements such as phages and transposable elements as well as the episomal plasmids are important drivers of the genome evolution of this important group of plant pathogens.

INTRODUCTION

BACTERIAL GENOMES are compact and carry the essential genes for reproduction and survival. Reproduction is guaranteed by a large set of genes involved in basic functions associated with the generation of metabolic energy and the flow of genetic information. These functions depend on the interaction of groups of proteins, and correspond to what is considered to define the essential components of the minimal genome or the central core genome. These functions characterize the "stable" genome, through which it is possible to trace the evolutionary history of an organism (Liu et al., 1999).

Survival can be regarded in a broader context as the fitness of an organism to quickly adapt to particular environmental changes and, to ensure adaptation, bacteria count on invasive DNA brought in by horizontal gene transfer (HGT) mechanisms to generate the genetic diversity needed. This occurs by means of integration of accessory genetic elements, such as transposons, integrons, plasmids, and prophages of temperate viruses. Such elements can carry genes that code for a specific ecological adaptation or for new features associated with pathogenicity or bacterium–host interaction (Sullivan and Ronson, 1998; Wong and Golding, 2003). Whole genome sequence analysis is uncovering multiple regions of DNA that contain genes

[1]Labinfo-LNCC, Petrópolis, RJ, Brazil.
[2]Instituto de Biociências, Universidade de São Paulo, São Paulo, SP, Brazil.
[3]Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo, Piracicaba, SP, Brazil.
[4]USDA, ARS, Parlier, California.

with distinct GC content and codon usage. Most of these are associated with mobile genetic elements and are most probably the consequence of HGT.

Phages, plasmids, and transposable elements are important sources of genetic variability for bacteria genome evolution (Boucher et al., 2003; Canchaya et al., 2004; Mira et al., 2002). These reports reveal the impact of these mobile elements on the differentiation of closely related species and strains, raising the question of the concept of bacterial species. How similar and conserved should a bacterial chromosome be in order to define that two strains are from the same species? What is the minimal percentage of invasive DNA that would differentiate two strains? These questions can be raised at both species and strain level, and human pathogenic bacteria have been the main focus of these questions (Canchaya et al., 2004; Casjens, 2003; Nakagawa et al., 2003; Perna et al., 2001). Genome sequences of plant pathogens available support the observation that invasive genetic elements are tools for genome diversification (da Silva et al., 2002; Van Sluys et al., 2003).

This paper aims to review the genetic mobile elements associated with the differentiation of the closely related plant pathogens *Xylella fastidiosa* and two *Xanthomonas* species. To meet this objective, we have scanned the complete genome sequence of two *Xylella fastidiosa* strains (*Xylella fastidiosa* 9a5c and *Xylella fastidiosa* Temecula) and two *Xanthomonas* species (*Xanthomonas axonopodis* pv. *citri* and *Xanthomonas campestris* pv. *campestris*), which have been sequenced by the ONSA/FAPESP team (Simpson et al., 2000; da Silva et al., 2002; Van Sluys et al., 2003). We also have closed and annotated two other *Xylella* plasmids, pXF30 and pXF31, isolated from two different *Xylella* strains, Ann1 and Dixon, respectively. Draft sequences of these two *Xylella* strains were previously released (Bhattacharyya et al., 2002). Each genome has its own content of mobile genetic elements, circumscribing what will be referred here as "mobil'omics," and is composed in this particular study by plasmids, prophages, insertion sequence (IS) elements, and transposons.

MATERIALS AND METHODS

Comparative studies were mostly performed manually at a website (⟨www.lbi.ic.unicamp.br⟩) and in a relational database as described in Van Sluys et al. (2002). Similarity searches were performed using BLAST, either at NCBI (⟨www.ncbi.nlm.nih.gov/BLAST⟩) or at LBI (⟨www.lbi.ic.unicamp.br⟩). The sequences of IS*Xac*1–4 reported here have been deposited in ISfinder (⟨www-is.biotoul.fr⟩).

Assembly of plasmids (pXF30 and pXF31) was achieved using chromatogram files isolated from the main Xf genome database. These chromatograms were assembled using the Phred+Phrap+Consed (Gordon et al., 1998; Ewing and Grenn, 1998) package. All consensus bases have a Phred quality of at least 20, and the overall error estimate is less than one in every 10,000 bases. Circularization was achieved using PCR-based strategies, and the plasmids were named pXF30 from Xf-OLS (Ann1 strain) and pXF31 from Xf-ALS (Dixon strain). Annotation was carried using SABIA (⟨www.labinfo.br⟩).

RESULTS AND DISCUSSION

Mobil'omics in Xylella *and* Xanthomonas

Xylella and *Xanthomonas* belong to the Xanthomonadales, and their core chromosome backbone share a common origin. However, these species evolved and diverged by insertion and potential deletions (INDEL). If best reciprocal matches using BLASTP program is utilized for comparative analyses, it is observed that *Xylella* species share around 74% of its genes with *Xanthomonas* species, whereas *Xanthomonas*, having a larger genome, share almost 40% (Moreira et al., 2004). An interesting feature is that most of the regions larger than 10 kbp that correspond to INDEL regions present at one of their border IS elements in the case of *Xanthomonas* and phage-related integrases in the case of *Xylella*. Among other genes, these two genera differ in their content of mobile genetic elements, as depicted in Table 1. *Xylella* strains have several phage-related regions in their chromosomes and also harbor plasmids. *Xanthomonas* carries more than 40 IS element insertions and only *Xac* has two plasmids (Fig. 1).

TABLE 1. *XANTHOMONAS* AND *XYLELLA* MOBIL'OMICS

	Xanthomonas, *species (genome size, bp)*		Xylella fastidiosa, *strains (genome size, bp)*			
Mobil'omics	*Xac*[a] *(5,175,554)*	*Xcc*[b] *(5,076,187)*	*9a5c*[c] *(2,679,305)*	*Temecula*[d] *(2,519,802)*	*Dixon*[e] *(ND)*	*Ann-1*[f] *(ND)*
Phage-related regions	2	1	5	8	ND	ND
Plasmids	2	0	2	1	1	1
IS elements[g]	44	73	1	0	0	1
Transposons	4	1	ND	ND	ND	ND

[a]*Xanthomonas axonopodis* pv. *citri*; strain 306.
[b]*Xanthomonas campestris* pv *campestris*; strain ATCC33913.
[c]Citrus variegated chlorosis.
[d]Pierce's disease.
[e]Almond leaf scorch disease.
[f]Oleander leaf scorch disease.
[g]Only full-length unit.
ND, not determined; IS, insertion sequence.

Plasmids

The plasmids present in the Xanthomonadales strains analyzed in this paper are listed in Table 2. In *Xac* and *Xylella* strains, the type IV secretion system (T4SS) is most prominent. Bacteria use T4SS to export or transfer DNA or proteins to different target-cells (Ding et al., 2003). This secretion system is ancestrally related to the conjugal transfer proteins responsible for plasmid mobilization among bacteria. Following Ding et al. (2003), T4SS can be grouped as (i) conjugation systems, (ii) effector translocator, and (iii) DNA uptake or release systems from/to the extracellular milieu. Conjugation systems deliver DNA or proteins by establishing direct contact with target cells; examples are the VirB operon from *Agrobacterium tumefaciens* and the Tra operon from F plasmid (Dessaux et al., 1998; Zupan and Zambryski, 1997; Kado, 2000). VirB proteins together with VirD4 are involved in the transfer of the T-strand from the bacterial cell to the cytoplasm of the plant host cell. These protein-encoding genes are organized as a cluster of 11 genes in the large pTi plasmid.

There are two regions in the *Xac* genome homologous to the VirB operon. One of these regions is located in the main chromosome (3,090,847–3,070,546) and the other in pXAC64. The two regions have no significant similarity at the nucleotide level, and also differ in gene order and composition if *vir*B5 and *vir*D4 are considered (Fig. 2). The T4SS operon of the chromosome is flanked upstream by a copy of the

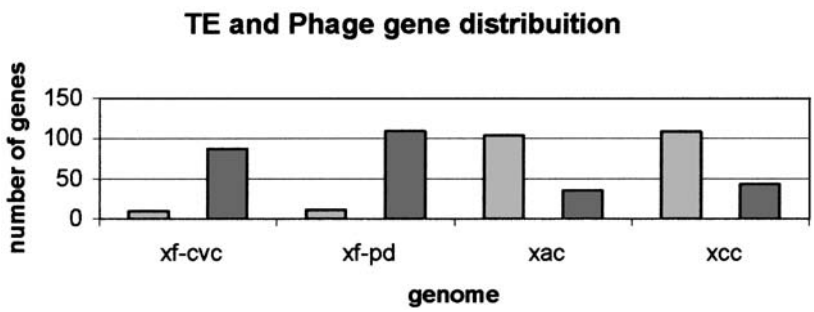

FIG. 1. Distribution of phage (dark gray) and transposable element (TE) (light gray) related genes in *Xanthomonas* and *Xylella* genomes.

TABLE 2. CHARACTERISTICS OF PLASMIDS PRESENT IN *XYLELLA FASTIDIOSA* STRAINS AND *XANTHOMONAS AXONOPODIS* PV. *CITRI*

Strain	*Plasmid*	*Size (bp)*	*GC (%)*[a]	*Number of predicted genes*	*Reference*
Xylella fastidiosa					
PD	pXFPD1.3	1,345	53.0	2	Van Sluys et al. (2003)
CVC	pXF1.3	1,285	55.6	2	Simpson et al. (2001)
	pXF51	51,158	49.6	65 (vapD)[b]	Simpson et al. (2001)
Ann1 (OLS)	pXF30	30,270	49.0	46 (vapD)	Bhattacharyya et al. (2002)
Dixon (ALS)	pXF31	31,572	49.0	36	This paper
Xanthomonas axonopodis pv. *citri*					
360	pXAC33	33,699	61.9	42 (pthA)	Silva et al. (2002)
	pXAC64	64,920	61.4	73 (pthA and avrPphE)	Salva et al. (2002)

[a]GC, content.
[b]Virulence-associated genes are in parentheses.

insertion sequence ISXac4 and downstream by a tRNA-VAL. The genomic version has no *vir*B5, which is postulated to be associated with T-pilus (Cascales and Christie, 2003) and present two copies of *vir*B6, one of which is partial.

Most of the predicted genes in the *Xanthomonas* system are more similar to those of *Bordetella* and *Brucella* T4SS. The major differences between the systems are the absence of *vir*B7, encoding a small protein (~55 aa) involved in T-pilus formation along with *vir*B2 and *vir*B5, and the general gene cluster organization (Fig. 2). Comparison with *Xylella fastidiosa* genes homologous to the VirB operon showed similar organization as the plasmid version of *Xanthomonas* system, with the exception that both *Xac* T4SS do not have the topoisomerase I between *vir*B1 and *vir*B2.

The function of this system in *Xanthomonas* remains to be established. However, most of the genes are similar to the system used in protein export, as in *Bordetella* and *Brucella*. None of the *Xac* T4SS regions, genomic or plasmidial, is complete. If these are functional, there may be some complementation between the operons to ensure system assembly. Two-hybrid studies may be a strategy to answer this question as it has been extensively used in other similar analyses (Cascales and Christie, 2003).

All *Xylella* strains sequenced to date harbor a plasmid. A small plasmid (pXF1.3) occurs in Xf-CVC

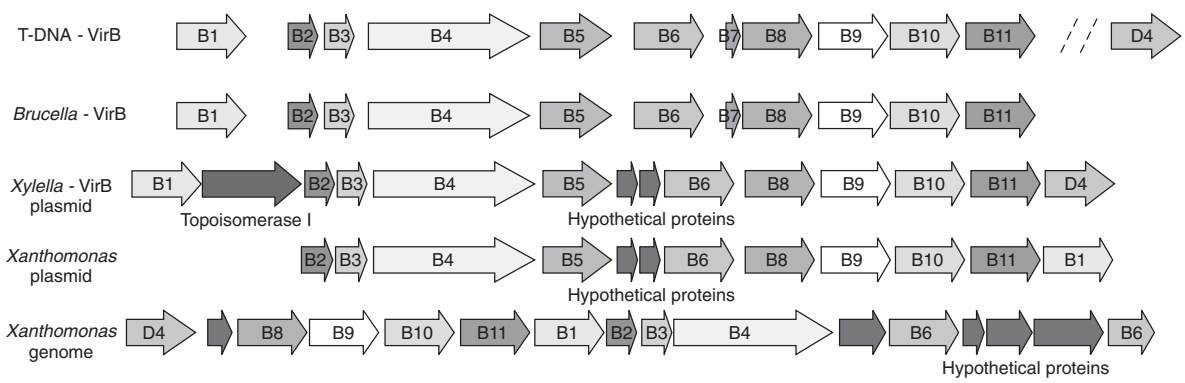

FIG. 2. Comparative organization of virB operon of *Agrobacterium, Brucella, Xylella fastidiosa,* and the two regions found in *Xanthomonas axonopodis* pv. *citri*.

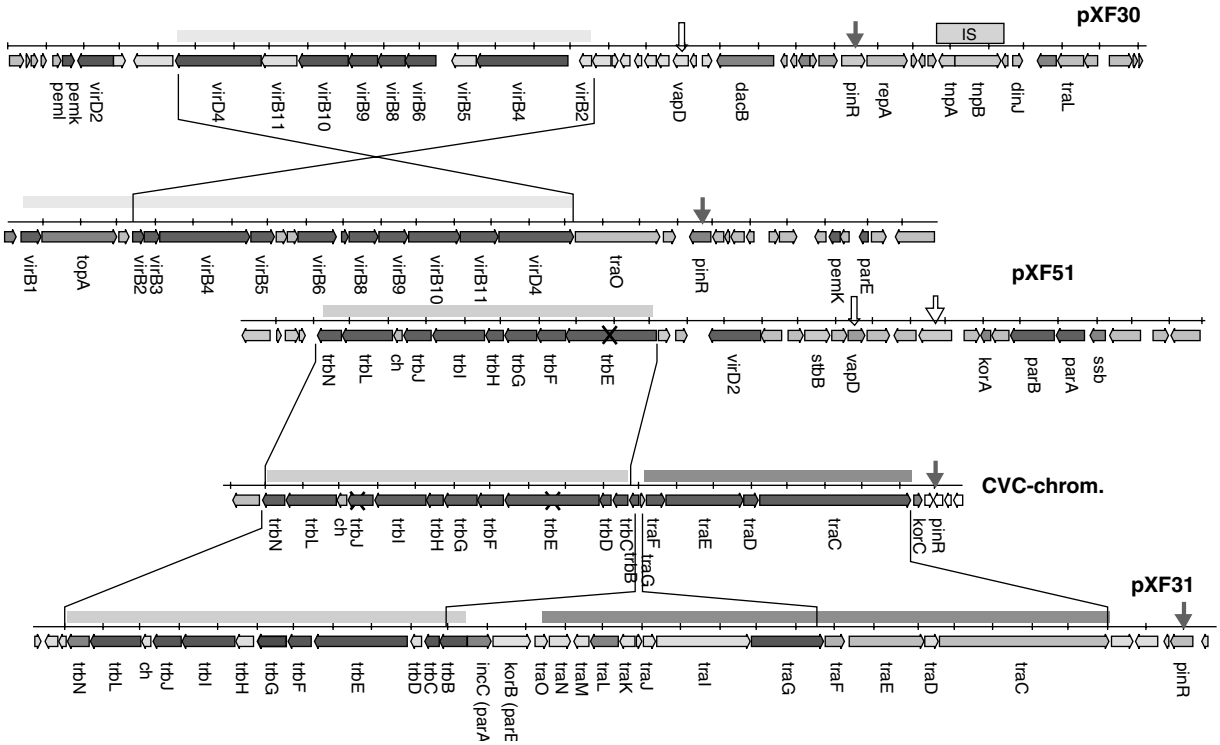

FIG. 3. Comparative organization of the Trb and VirB operon of the plasmids and chromosome of different *Xylella fastidiosa* strains.

(strain 9a5c), Xf-PD (strain Temecula) which share 87% nucleotide identity, and in several other strains of *X. fastidiosa* isolated from different plant hosts (Pooler et al., 1997; Hendson et al., 2001). This plasmid carries two predicted genes, one conserved hypothetical and the other one a replication protein.

In addition to the small plasmid, Xf-CVC (strain 9a5c) also carries a larger plasmid pXF51 where two distinct T4SS are found. The draft genome sequence of Xf-OLS (Ann1 strain), causing oleander leaf scorch (Purcell et al., 1999), describes a chromosome independent replicon (Bhattacharyya et al., 2002). We have circularized this sequence and named it pXF30. Another *Xylella* strain, Xf-ALS (Dixon strain), causing almond leaf scorch (Mircetich et al., 1976) carries pXF31. In order to evaluate sequence similarity between these replicons, MUMMER (Kurtz et al., 2004) based analyses were undertaken and clearly established that pXF30 has an independent origin from the other two (pXF51 and pXF31) as no sequence similarity was detected. Contrary to that, nucleotide similarity within the T4SS is detected in *Xf*-CVC (strain 9a5c) and *Xf*-ALS (Dixon strain) plasmids (Fig. 3).

Plasmids pXF51 and pXF1.3 of *Xf*-CVC (strain 9a5c) have been described elsewhere (Marques et al., 2001; Simpson et al., 2000; Van Sluys et al., 2002). Plasmid pXF51 harbors two T4SS, one that is similar to VirB operon found in *A. tumefaciens* pTi and another similar to Trb region from *E. coli* pRP4 (IncP). As in *Xac*, *Xf*-CVC (strain 9a5c) contains an extra T4SS integrated in the chromosome that shares 98% nucleotide identity with the Trb plasmidial region. The chromosomal location of this integrated plasmid corresponds to an INDEL associated with genome reorganization between *Xf*-CVC from *Xf*-PD.

pXF31

The newly described pXF31 found in *X. fastidiosa* (Dixon strain) contains 36 ORFs, of which 23 are related to T4SS, one is involved in plasmid stability and/or replication, one is a transcriptional regulator, four are conserved hypothetical proteins, six are hypothetical proteins and one is a phage-related DNA inver-

Table 3. Rearrangements Associated with Putative Bacteriophage Integrases

Genome position				
Start	*End*	*Length (kb)*	*Phage-integrase*	*tRNA*
Xf-CVC (9a5c)				
340,058	347.748	7.6	—	—
648,188	693.774	45.5	XF0678 in XfP1	tRNA-V
924,033	929.788	5.7	XF0968	tRNA-V
1,512,321	1.541.891	29.5	XF1555	—
2,008,487	2.024.917	16.4	XF2132 in XfP5	—
2,359,908	2.407.335	47.4	XF2478 and XF2530 in XfP2	—
Xf-PD (Temecula)				
338,499	341.667	3.1	—	—
944,301	951.351	7.0	PD0764	—
1,375,418	1.376.971	1.5	PD1196 in xpd4	—
1,736,666	1.738.827	2.1	PD1495	tRNA-V
2,002,896	2.021.110	18.2	PD1732 in xpd7	tRNA-V
2,246,927	2.251.145	4.2	—	—

tase/recombinase (*pin*R). The pXF31 contains Tra and Trb operons. The Trb operon includes *trb*B-J, a conserved hypothetical, *trb*L and *trb*N. The same arrangement of genes is found in Xf-CVC (strain 9a5c) both in the chromosome and in pXF51 (Fig. 3) with the exception of *trb*B, which is missing from the operon in the *Xf*-CVC (strain 9a5c) chromosome and *trb*B-D, which is missing from pXF51. These regions share high nucleotide sequence identity [*Xf*-CVC (strain 9a5c) − chromosome × pXF31 = 96.7%; pXF31 × pXF51 = 96%]. Besides the Trb operon, both plasmids and the *Xf*-CVC (strain 9a5c) chromosome also have a gene coding for *pin*R, but with a relatively low (44%) amino acid identity.

The Tra operon includes *tra*C-G and *tra*I-O. A similar region is found in the *Xf*-CVC (strain 9a5c) chromosome, containing *tra*C-F. The average amino acid identity between the *tra*C-F of the Xf-CVC (strain 9a5c) chromosome and pXF31 is 87.8% (±17.2%). Other ORFs present are *kor*B (*par*B family) and *inc*C (*par*A family), which are thought to be involved in plasmid maintenance and replication. Similarly pXF51 has *par*A and *par*B. The overall structure and gene content of pXF31 is most similar to the plasmid region integrated in the *Xf*-CVC (strain 9a5c) chromosome (Fig. 3), which could be indicative that pXF31 is well disseminated in *Xylella* populations and has integrated in *Xf*-CVC (strain 9a5c) genome. The dynamics of the mobilization of these plasmids among *Xylella* strains, as well as other bacterial species, still is to be demonstrated but the fact that some of the T4SS can share more than 95% nucleotide identity and carry different sets of genes (or rapidly evolving genes) is suggestive of their contribution to genome divergence.

pXF30

A putative episome of 30,270 bp in the *Xf*-OLS (Ann-1) genome was described in Bhattacharyya et al. (2002). This plasmid has a Vir operon and harbors an IS element between positions 24,500 and 26,700. Plasmid pXF30 contains 46 ORFs, of which 11 ORFs are related to T4SS, four are involved in plasmid stability or replication, seven are conserved hypothetical proteins, 18 are hypothetical proteins, two are similar to transposases and four have other functions. An addictive plasmid maintenance operon, *pem*I-*pem*K, is found where PemK is proposed to inhibit cell division of plasmid-free segregants, and PemI suppresses its activity. Although pXF51 has *pem*K and a putative *pem*I (Marques et al., 2001), there seems to be no similarity to the *Xf*-CVC (strain 9a5c) proteins/genes. Bhattacharyya et al. (2002) described a region containing the T4SS. The Vir operon in pXF30 includes the *vir*B2, *vir*B4-6, *vir*B8-11, and *vir*D4. The ORF for *vir*D2 is found a couple of genes upstream. These genes are arranged similarly to what is found in other plasmids, including pXF51 (Fig. 3) (Marques et al., 2001), but with low amino acid similarity (38% and lower) between these operons. The gene *vir*B3, which is part of this operon in other plasmids, is not pres-

ent in pXF30. Following the Vir operon, there is a region containing hypothetical ORFs and one conserved hypothetical and the gene coding for *vap*D (virulence-associated protein D), which is a homolog (98% amino acid and 93% nucleotide identity) of the XFa0052 from pXF51. Although this protein is related to virulence in various animal pathogens, there is no proposed mechanism for its action (Catani et al., 2004). Both ORFs from pXF30 and pXF51 are similar (64% and 51% amino acid identity) to *vap*D from the human pathogenic bacteria *Actinobacillus actinomycetemcomitans* and *Neisseria meningitidis* MC58. This gene was also found in other *X. fastidiosa* strains isolated from citrus and coffee, but seems to be absent from plum, mulberry, grape, elm, almond, and ragweed strains of this bacterium (Nunes et al., 2003). Although pXF30 carries a Vir operon similar to pXF51, no significant sequence (amino acid or nucleotide) similarity is found between the two plasmids, the exception being *vap*D, with a relatively high nucleotide similarity between them implying that these plasmids could exchange gene cassettes.

It is interesting to consider that these T4SS homologous regions present in both *Xanthomonas* and *Xylella* could be involved in the transport of specific macromolecules associated with virulence/pathogenesis as is the case for both *Bordetella pertussis* and *Brucella suis* and more distantly related T4SS described in *Helicobacter pylori* and responsible for IL-8 secretion (Selbach et al., 2002). Even though, the role of T4SS homologues in these Xanthomonadales-associated plasmids remains unclear.

Phages

The genomes of *Xanthomonas* and *Xylella* have at least one recognizable prophage region. Bacteriophages with dsDNA are found infecting a large number of unrelated bacterial species. With the sequencing of entire bacterial genomes, homologous genes of phages and prophages are described sometimes where no phage particles have been isolated. Each of these genomes harbors at least one copy of a prophage, or remnants of a prophage, related to P2 coliphage or ϕ CTX of the *Pseudomonas aeruginosa* genomes. The arrangements of these regions are similar in all cases.

In *Xac*, there are 35 predicted genes related to bacteriophage sequences, 15 of them are grouped and together resemble a defective prophage, most likely a P2-type phage. This region is 27 kbp long (3,093,912–3,121,023) inserted close to a tRNA-ASN, and contains putative genes similar to and with the same organization as the proteins for head assembly and part of tail formation of the ϕCTX phage particle of *Pseudomonas aeruginosa* (Nakayama et al., 1999) interrupted by an insertion of a copy of IS*xac3*. In *Xcc*, a P2-like 36 kb long prophage region is inserted into a tRNA-LYS (3,493,680–3,529,955). The *Xcc* genome also has an integrated ϕ-Lf DNA, a phage that specifically infects cell of this species. The 12-kbp region containing putative genes gI-gIX is located at position 2,435,600–2,447,919 of *Xcc* genome.

The genomes of the *Xylella* strains have the highest number of phage-related sequences dispersed in the chromosome, constituting 7% of the *Xf*-CVC (strain 9a5c) genome and 9.02% of the *Xf*-PD (Temecula strain) genome. The *Xf*-CVC (strain 9a5c) chromosome is composed of 2,679,305 bp (52.7% GC content). The *Xf*-PD (Temecula strain) genome is composed of 2,519,802 bp with a GC content of 51.7%. Different from *Xanthomonas*, *Xylella* genomes exhibit large-scale genomic rearrangements probably due to the presence of prophages or prophage-like elements (Van Sluys et al., 2002). Comparative analysis of the *Xf*-PD (Temecula strain) and *Xf*-CVC (strain 9a5c) genomes revealed a region of 68 kbp present in *Xf*-CVC (strain 9a5c) and absent from *Xf*-PD (Temecula strain), inserted in a tRNA-GLY with a 13-bp duplication adjacent to a phage integrase. This region is putatively associated with environmental adaptation (Van Sluys et al., 2002).

The sequences of the *Xf*-PD (Temecula strain) and *Xf*-CVC (strain 9a5c) genomes are to a large extent collinear. However, a major reorganization has taken place between the two sequences. In all cases, rearrangements are associated with putative bacteriophage integrases and other repeated sequences (Table 3). The *Xf*-CVC (strain 9a5c) genome includes five prophage regions (Simpson et al., 2000; Canchaya et al., 2004) with a different GC content (57%) and several other phage-related genes dispersed throughout the sequence, which result in a high percentage of repeated fragments. These regions and genes are organized differently in the *Xf*-PD (Temecula strain) genome and underlie its reorganization compared to the *Xf*-CVC (strain 9a5c). A total of eight clusters of phage-related genes were identified in *Xf*-PD (Temecula strain) none of which are present in the *Xf*-CVC (strain 9a5c) genome. In contrast, all other shared regions be-

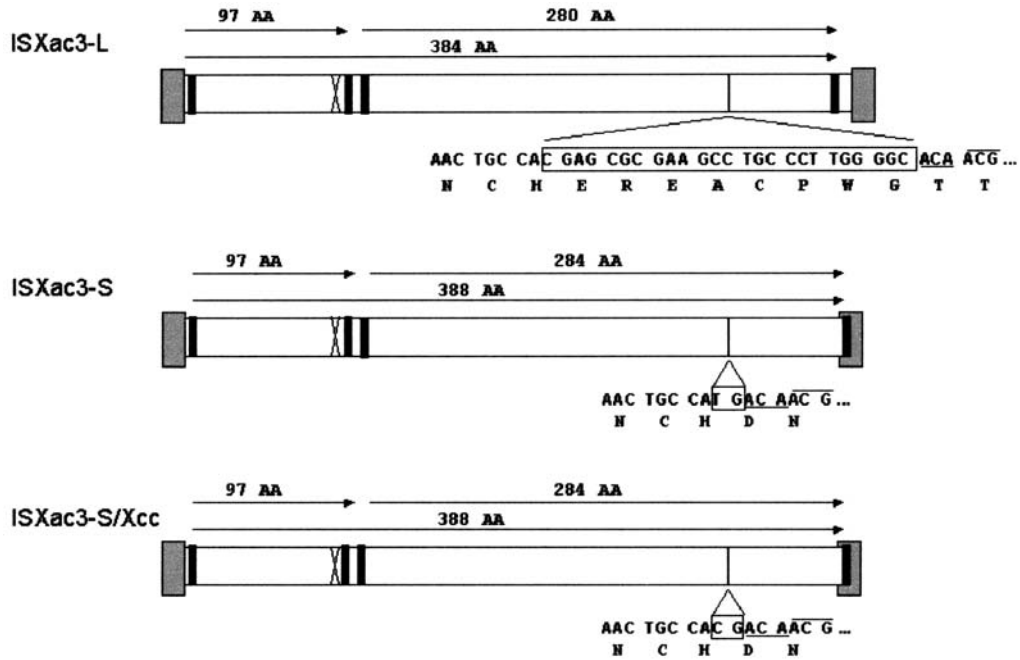

FIG. 4. Variants of ISXac3 in *Xac* and *Xcc*. White rectangles are the two putative coding regions of each insertion sequence and the respective size in amino acids. Gray boxes represent the 35-bp terminal inverted repeats; black lines represent the start and stop codon of each ORF; X's represent a potential frameshift window (nucleotides A6[G] 303–309), and the nucleotides shown indicate the differences found in each of the insertion sequence (IS) elements.

tween the *Xf*-CVC (strain 9a5c) and *Xf*-PD (Temecula strain) genomes are highly similar in gene order and nucleotide sequence. We speculate that these rearrangements are due to the presence of large phage or prophage-like regions that act as the substrate for recombination while in the case of the *Xanthomonas* repeated sequences these mostly consist of discrete IS elements.

Transposon and insertion sequence elements

Transposable elements (TE) are ubiquitous among all living organisms where they have been searched. Their activity was first detected due to unstable mutations, but now their presence is being associated as main components of the genome so that restructuring under stressed conditions may be guaranteed. It is interesting to note that there is a wide variation in the copy number of these elements between different organisms for instance in maize they can represent 50% of the genome as for yeast they represent on 3.1%. In the *Xac* genome, 1.36% of the genome is associated with a transposase. That is a total of 71,988 nucleotides coding for a transposase or transposon related functions. In this analysis, the inverted repeats and non-coding regions of the elements were not considered.

Comparative genomic analyses show that the overall organization and gene order within the two *Xanthomonas* species are similar. A total of 800 ORFs (18.5% of the genome) are present in citri and absent in campestris and 646 ORFs (15.4% of the genome) are present in *Xcc* and absent in *Xac*. These ORFs are distributed unevenly throughout the conserved backbone. Although some of the specific ORFs are isolated, most of them are clustered into islands of up to 100 kbp in size. Some of these regions present an altered GC content that can be as low as 46%, in contrast to the genome average of 64%. These clusters include the type IV secretion system not described before for this genus, genes encoding hemolysin, haemagglutinin, hemin storage proteins, syringomycin, avrPphE and cvgSY homologous proteins (da Silva et al., 2003). *Xac* and *Xcc* genomes were invaded by several insertion sequence (IS) elements and each genome has an element that is most represented (Table 4). There are more than 100 transposase

Table 4. Transposons and IS Elements Found in *Xac* and *Xcc* Genomes

Transposon and IS element (size)	*IS family*[a]	*Accession number*[b]	*Number of copies*[c]	
			Xac	Xcc
IS*Xac1* (1,352 bp)	IS4	Z73593	8	—
IS*Xac2* (1,195 bp)	IS1	AF263433	8	—
IS*Xac3* (1,255 bp)	IS3	AF327445	21	7
IS*Xac4*	IS1	AF034211	2	—
IS*1389*	IS3	U77781	1	—
IS*1404*	IS3	U45994	—	6
IS*1477*	IS3	M28557	—	7
IS*1478*	IS5	U59549	1	16
IS*1479*	IS5	AF077016	1	10
IS*1480*	IS5	U61260	—	7
IS*1481*	IS4	AF090837	—	11
IS*1595*	ISNCY	AF225215	—	1
IS*Xcd1*	IS1	AF263433	2	2
IX*Xcc1*	IS3	AF047478	—	5
ISD1	IS3	AF047478	—	1
Tn5041	Tn3	Z73593	—	1
Tn5044	Tn3	AF174129	2	—
Tn5045 (6,938 bp)	Tn3	Z73593	2	—

[a]As defined by Mahillon and Chandler (1998).

[b]Accession number of the most similar element (amino acid level) identified at <www-is.biotoul.fr>.

[c]Copy number represents full-length copies of the insertion sequence (IS) elements present in the chromosome and in the plasmids.

genes in each genome, and 15 distinct IS families were fully identified. All the genomic islands corresponding to an INDEL between the two genomes harbor IS*xac3* insertions in *Xac*. Other regions present remnants of plasmids and prophages.

Prokaryotic transposable elements can be either IS, ranging from 1,000 to 2,000 bp, or Transposons (Tn), with more than 2,000 bp. Transposons can be composite elements with IS sequences at the ends. Both have structural features that define them as TE; these are the transposase and the inverted repeats (TIR) that delimit these elements. The number of transposable elements, either transposons or IS elements, varies between *Xac* and *Xcc*. As for *Xylella*, only Xf-CVC (strain 9a5c) (Ferreira et al., 2002) and Xf-OLS (strain ANN-1) have complete IS element.

The IS element found in Xf-OLS is a member of the IS605 family and is located in the plasmid pXF31. This type of element encodes two predicted transposases translated in a divergent manner (Mahillon and Chandler, 1998). The element named IS*Xfo*1 is 2,039 bp long and is delimited by an imperfect TIR. The sequence *CAT*G (611–614) harbors the start codon of the two transposases translated in opposite orientation. The first predicted gene codes for an IS200-like transposase of 145 amino acids most similar to the putative transposase TnpA of the IS609 of *Escherichia coli* O157:H7. The second gene codes for a transposase of 404 amino acids, which is similar to the TnpB of same element in *E.coli*.

Four full-length units were identified in *Xac* genome (Table 5). Two of them were already described in X*anthomonas* and are Tn5045 (a composite element, originally named ISXC5) (Tu et al., 1989) and ISXCD1 (Lee and Lin, 2000).

IS*xac1*, IS*xac2*, and IS*xac3* are the most disseminated IS in *Xac* genome. Altogether they are present in 36 copies representing ~1% of the genome while the others are present in one or two copies. IS*Xac2* and IS*Xac3* are present in *Xac* in close association with genomic islands. In *Xcc,* IS*Xac1* and IS*Xac2* are very poorly represented being identified only as highly degenerated copies or truncated elements.

Table 5. General Features of Insertion Sequence Elements Present in the *Xac* Genome

Element	*Family*[a]	Xac	*p*Xac*33*	*p*Xac*64*	*Length (bp)*	*Transposase*	*DR*	*TIR*	
IS*Xac*1	IS4	8	1	—	1,351	404 AA	9 bp	5′ end	CTGAGGTTTCCCCACATTT
							aacCtgacc	3′ end	AAACGTGGGGATACCTAAG
IS*Xac*2	IS1	5	1	—	1,195	88 + 275	4 bp[b]	5′ end	GGTAATCCCCCCGCCCATTAGCAGACGCCAGAAGTGGA
						(362)		3′ end	TCCACTTCTGGCGACCGCTAGAAGTGGGGGGATTACC
IS*Xac*3-L	IS3	13	—	—	1,257	97 + 280	3 bp[b]	5′ end	TGATACGCCCAGGATTTCCTAGACATCTCAAGGCC
						(384)		3′ end	GGTCTTGAGTGTCTACGGAACCCTGGGCGTATCA
IS*Xac*3-S	IS3	6 + **2**	—	**1**	1,234	97 + 284	3 bp[c]	5′ end	TGATACGCCCAGGATTTCCTAGACATCTCAAGGCC
						(388)		3′ end	GGTCTTGAGTGTCTACGGAACCCTGGGCGTATCA
IS*Xac*4	IS1	1 + **2**	—	—	1,198	88 + 273	—	5′ end	CGTAAGCTCCCCCGATCCGCATACACCGCATAAGTAGATCTTTC
						(370)		3′ end	GAAAGCTCTACTTTTGAAGTGCCTGCTTGACGGGGGAGCTTACG

[a]As defined by Mahillon and Chandler (1998).
[b]Present only in a few copies.
[c]Present only in one copy.
Highlighted, not complemented.
Bold-faced, truncated versions.
IS*Xac*1 ~ IS1481 (81% nucleotide identity).
IS*Xac*2 ~ ISXcc1 (less than 44% nucleotide identity).
IS*Xac*3-S ~ IS *Xac*3-S-Xcc (91% nucleotide identity).
IS, insertion sequence; DR, duplicated region; TIR, terminal inverted repeat.

IS*Xac3* (Fig. 4) is closely related to an *Erwinia amylovora* transposon borne in the ubiquitous plasmid pEA29 and to an IS-like element present in the *Agrobacterium tumefaciens* pTI plasmid from the AB2/73 limited host range strain. A member of the IS3 family (Mahillon and Chandler, 1998), this element is found in two versions in *Xac* genome, and only one of them is found in *Xcc* genome. The two variants are 98% identical and one of them present a deletion of 23 bp. ISXac3-L (long) represented in 13 copies is 1,257 bp long and ISXac3-S (small) represented in six copies is 1,234 bp. We have also detected three other degenerated copies of these elements, two in the genome and another one in the pXac64. ISXac3-S is present in 8 full copies and one degenerated version in *Xcc*. The *Xcc* element shares 91% nucleotide identity with the IS*Xac*3-S. All variants have two consecutive open reading frames, which can be translated as one because of a potential frameshift along the sequence A6G (303-309) (Fig. 4). Most of the elements related to this family are translated by programmed frameshifting as previously described (Mahillon and Chandler, 1998). The 23-bp deletion is located at the C-terminal of the second putative translated protein generates, as a result, two distinct protein versions. The second ORF of the ISXac3-S is longer, coding for a protein of 284 AA with a stop codon within the 3′ end TIR. The stop codon of the second ORF in ISXac3-L is located upstream of the 3′ end TIR, resulting in a shorter version of the protein (280 aa). As a result ISXac3-L transposase is 4 amino acids shorter than the other variants in *Xac* and *Xcc* genomes. Also, the last 54 AA of the predicted protein is different losing its similarity with Tra5 domain (gnl|CDD|12158, COG2801) and rve domain (gnl|CDD|25582, pfam00665) (⟨www.ncbi.nlm.nih.gov/Structure/cdd⟩).

Xac also harbors IS*Xac*4, an element belonging to the IS1 family. The transposase of this element is most similar to the one in the ISXCD1 of *Xanthomonas axonopodis* pv. *dieffenbachiae* and to the OrfB of *Desulfovibrio vulgaris* subsp. *vulgaris*. As described for IS*Xac*3, the two consecutive ORFs are probably translated by programmed frameshifting, being the sequence A6C found at nucleotides 328–334 close to the end of the first ORF. This element is present in only one intact copy (3070606–3071803) by the side of the VirB operon. This region encompasses the largest number of *Xac* specific genes (123 genes) comparing *Xac* and *Xcc* whole genomes. There are two partial matches to IS*Xac*4 corresponding to the 5′ region (2765261–2764616) and 3′ region (2250174–2249941) separated by 500,000 bp. This region encompasses two genomic islands in which we identified *syr*E related genes, which in *Pseudomonas syringae* are responsible for the synthesis of the syringomycin phytotoxin (Menestrina and Semjén, 1999). Also in these regions, *Xac* has an insertion of 133,025 bp that harbors many non-gamma bacterial genes and 18 genes also found in *Xylella fastidiosa* CVC strain (da Silva et al., 2002; Lima et al., 2005) including a putative RTX like–toxin and hemagglutinins.

In the *Xcc* genome, IS*1478* is the most abundant (16 copies) element. It was previously described by Chen et al. (1999) in seven different *Xcc* strains when isolated after transposition within a target plasmid. IS*1478* belongs to IS5 family of transposases and, the 16 copies share 99–100% nucleotide identity to the published sequence (gi 1927196) indicating that these copies may be active. IS elements have been used to distinguish *Xanthomonas* strains and, only one of the six ISs present in *X. oryzae* pv *oryzae* has been detected in *Xac* and none in *Xcc*. Although no direct evidence being associated with the direct transmission of selective markers, it is interesting to note that most of the IS elements are either located close to or within the genomic islands described (da Silva et al., 2002; Lima et al., 2005).

Tn5045 is a composite transposon already described in *Xcc* as ISXC5 (Tu et al., 1989). This transposon was originally described to be born in pXW45J plasmid from *X. campestris* pv *cit*ri strain XW45. In *Xac,* a complete copy of the element is present in the pXAC33, while the IS unit (IS*Xac1*) is missing in pXAC64. In both plasmids, the element is in close association with the *pth*A gene. The transposon presents, between positions 5223 and 5235, a sequence (gccatcgccagca) that is also found in *pth*A repeats. This region could affect the variability in the numbers of repeats. We also find the same sequence as part of genes of unknown function in *Xylella* (71% AA identity) and *Xanthomonas campestris* (68% AA identity). In *Xylella*, this coding region is present in the pXF51 plasmid and in *Xanthomonas* is associated to Tn5053 family of transposons.

The contribution of these elements on these genomes can be inferred by the fact that only four IS elements are shared by *Xac* and *Xcc* (IS*1478*, IS*1479*, IS*Xcd1*, and IS*Xac3*) all the other 13 elements are par-

ticular elements to either of the two genomes. It is tempting to speculate that these IS elements may be drivers of the differentiation of the *Xanthomonas* genomes probably as a result of phage transduction or plasmid conjugation events.

CONCLUSION

Molecular phylogeny analysis (Van Sluys et al., 2002) showed the close relationship of *Xylella* and *Xanthomonas.* Whole genome analyses support this close relationship considering that *Xylella* species share around 74% of its genes with *Xanthomonas* species (Moreira et al., 2004). However, these genomes evolved and diverged by potential insertion and deletions of different DNA fragments. Many of those large (<10 kbp) INDEL regions are bordered by transposable elements in the case of *Xanthomonas* and by phage-related in the case of *Xylella* both acquired through lateral transfer. Whole genomic analyses clearly suggested a fundamental role of these mobile genetic elements in the differentiation of these organisms. Considering the above, T4SS and the plasmids that harbor these genes may play an important role in the DNA import and export. The clusters for T4SS were found in all the plasmids of *Xf* and *Xac*, with the exception of the small pXF1.3. The association of each class of genetic mobile elements to a specific group of genes in each genome varies and their contribution to organism diversification and adaptation is now being disclosed.

ACKNOWLEDGMENTS

We would like to thank all the technicians, students, and researchers of the ONSA/FAPESP consortium that made the Xylella and Xanthomonas project a reality (⟨http://watson.fapesp.br/onsa/Genoma3.htm⟩). S.R. Blancoand and E.Q. Pereira provided technical support for sequencing. L. Gonzaga, R. Paixao, R. de Souza, under A.T. Vasconcelos' supervision provided technical support for the bioinformatic web interface at Labinfo-LNCC. Genome sequencing projects received financial support from FAPESP and USDA-ARS.

REFERENCES

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

BOUCHER, Y., DOUADY, C.J., PAPKE, R.T., et al. (2003). Lateral gene transfer and the origins of prokaryotic groups. Annu Rev Genet **37,** 283–328.

BREEDVELD, M.W., and MILLER, K.J. (1994). Cyclic beta-glucans of members of the family Rhizobiaceae. Microbiol Rev **58,** 145–161.

BRIONES, G., IANNINO, N.I., ROSET, M., et al. (2001). *Brucella abortus* cyclic β-1,2-glucan mutants have reduced virulence in mice and are defective in intracellular replication in HeLa cells. Infect Immun **69,** 4528–4535.

CANCHAYA, C., FOURNOUS, G., and BRUSSOW, H. (2004). The impact of prophages on bacterial chromosomes. Mol Microbiol **53,** 9–18.

CASCALES, E., and CHRISTIE, P.J. (2003). The versatile bacterial type IV secretion systems. Nat Rev Microbiol **1,** 137–149.

CASJENS, S. (2003). Prophages and bacterial genomics: what have we learned so far? Mol Microbiol **49,** 277–300.

CATANI, C.F., AZZONI, A.R., PAULA, D.P., et al. (2004). Cloning, expression, and purification of the virulence-associated protein D from *Xylella fastidiosa*. Protein Expr Purif **37,** 320–326.

CHEN, J.H., HSIEH, Y.Y., HSIAU, S.L., et al. (1999). Characterization of insertions of IS476 and two newly identified insertion sequences, IS1478 and IS1479, in *Xanthomonas campestris* pv. *campestris*. J Bacteriol **181,** 1220–1228.

DA SILVA, A.C.R., 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.

DESSAUX, Y., PETIT, A., ELLIS, J.G., et al. (1989). Ti plasmid-controlled chromosome transfer in *Agrobacterium tumefaciens*. J Bacteriol **171,** 6363–6366.

DING, Z., ATMAKURI, K., and CHRISTIE, P.J. (2003). The outs and ins of bacterial type IV secretion substrates. Trends Microbiol **11,** 527–535.

EWING, B., and GREEN, P. (1998). Base-calling of automated sequencer traces using phred. II. Error probabilities. Genome Res **8,** 186–194.

FERREIRA, L.P., LEMOS, E.G., and LEMOS, M.V. (2002). Transposon Tn1721 distribution among strains of *Xylella fastidiosa*. FEMS Microbiol Lett **5,** 163–168.

GORDON, D., ABAJIAN, C., and GREEN, P. (1998). Consed: A graphical tool for sequence finishing. Genome Res **8,** 195–202.

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.

KADO, C.I. (2000). The role of the T-pilus in horizontal gene transfer and tumorigenesis. Curr Opin Microbiol **3,** 643–648.

KURTZ, S., PHILLIPPY, A., DELCHER, A.L., et al. (2004). Versatile and open software for comparing large genomes. Genome Biol **5,** R12.

LIMA, W.C., VAN SLUYS, M.-A., and MENCK, C.F.M. (2005). Non-gamma proteobacteria gene islands contribute to *Xanthomonas genome*. OMICS **9,** 160–172.

LIU, S.L., SCHRYVERS, A.B., SANDERSON, K.E., et al. (1999). Bacterial phylogenetic clusters revealed by genome structure. J Bacteriol **181,** 6747–6755.

MAHILLON, J., and CHANDLER, M. (1998). Insertion sequences. Microbiol Mol Biol Rev **62,** 725–774.

MARQUES, M.V., SILVA, A.M., and GOMES, S.L. (2001). Genetic organization of plasmid pXF51 from the plant pathogen *Xylella fastidiosa*. Plasmid **45,** 184–199.

MENESTRINA, G., and SEMJÉN, B.V. (1999). Biophysical methods and model membranes for the study of bacterial pore-forming toxins. In *The Comprehensive Sourcebook of Bacterial Protein Toxins*. J.E. Alouf and J.H. Freer, eds. (Academic Press, San Diego), pp. 287–309.

MIRA, A., KLASSON, L., and ANDERSSON, S.G. (2002). Microbial genome evolution: sources of variability. Curr Opin Microbiol **5,** 506–512.

MIRCETICH, S.M., LOWE, S.K., MOLLER, J.W., et al. (1976). Etiology of almond leaf scorch disease and transmission of the causal agent. Phytopathology **66,** 17–24.

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.

NAKAGAWA, I., KUROKAWA, K., YAMASHITA, A., et al. (2003). Genome sequence of an M3 strain of *Streptococcus pyogenes* reveals a large-scale genomic rearrangement in invasive strains and new insights into phage evolution. Genome Res **13,** 1042–1055.

NAKAYAMA, K., KANAYA, S., OHNISHI, M., et al. (1999). The complete nucleotide sequence of phi CTX, a cytotoxin-converting phage of *Pseudomonas aeruginosa*: implications for phage evolution and horizontal gene transfer via bacteriophages. Mol Microbiol **31,** 399–419.

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.

PERNA, N.T., PLUNKETT 3rd, G., BURLAND, V., et al. (2001). Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7. Nature **409,** 529–533.

POOLER, M.R., HARTUNG, J.S., and FENTON, R.G. (1997). Sequence analysis of a 1296-nucleotide plasmid from *Xylella fastidiosa*. FEMS Microbiol Lett **155,** 217–222.

PURCELL, A.H., SAUNDERS, S.R., HENDSON, M., et al. (1999). Causal role of *Xylella fastidiosa* in oleander leaf scorch disease. Phytopathology **89,** 53–58.

SELBACH, M., MOESE, S., MEYER, T.F., et al. (2002). Functional analysis of the *Helicobacter pylori* cag pathogenicity island reveals both VirD4-CagA-dependent and VirD4-CagA-independent mechanisms. Infect Immun **70,** 665–671.

SIMPSON, A.J.G., 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.

SULLIVAN, J.T., and ROBSON, C.W. (1998). Evolution of rhizobia by acquisition of a 500-kb symbiosis island that integrates into a phe-tRNA gene. Proc Natl Acad Sci USA **95,** 5145–5149.

TU, J., WANG, H.R., CHANG, S.F., et al. (1989). Transposable elements of *Xanthomonas campestris* pv. *citri* originating from indigenous plasmids. Mol Gen Genet **217,** 505–510.

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

VAN SLUYS, M.A., 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.

WONG, K., and GOLDING, G.B. (2003). A phylogenetic analysis of the pSymB replicon from the *Sinorhizobium meliloti* genome reveals a complex evolutionary history. Can J Microbiol **49,** 269–280.

ZUPAN, J.R., and ZAMBRYSKI, P. (1995). Transfer of T-DNA from Agrobacterium to the plant cell. Plant Physiol **107,** 1041–1047.

Address reprint requests to:
Marie-Anne Van Sluys, Ph.D.
Departamento de Botânica–IBUSP
Rua do Matão, 277
05508-900, São Paulo, SP, Brasil

E-mail: mavsluys@usp.br