Comparison of the genomes of two *Xanthomonas* pathogens with differing host specificities

A. C. R. da Silva[1], J. A. Ferro[2], F. C. Reinach[1,20], C. S. Farah[1], L. R. Furlan[3], R. B. Quaggio[1], C. B. Monteiro-Vitorello[4], M. A. Van Sluys[5], N. F. Almeida[6], L. M. C. Alves[2], A. M. do Amaral[7], M. C. Bertolini[8], L. E. A. Camargo[4], G. Camarotte[4], F. Cannavan[9], J. Cardozo[10], F. Chambergo[1], L. P. Ciapina[2], R. M. B. Cicarelli[10], L. L. Coutinho[4], J. R. Cursino-Santos[11], H. El-Dorry[1], J. B. Faria[12], A. J. S. Ferreira[1], R. C. C. Ferreira[12], M. I. T. Ferro[2], E. F. Formighieri[9], M. C. Franco[9], C. C. Greggio[2], A. Gruber[13], A. M. Katsuyama[13], L. T. Kishi[2], R. P. Leite[14], E. G. M. Lemos[2], M. V. F. Lemos[15], E. C. Locali[7], M. A. Machado[7], A. M. B. N. Madeira[13], N. M. Martinez-Rossi[11], E. C. Martins[10], J. Meidanis[16], C. F. M. Menck[12], C. Y. Miyaki[17], D. H. Moon[9], L. M. Moreira[1], M. T. M. Novo[8], V. K. Okura[16], M. C. Oliveira[5], V. R. Oliveira[12], H. A. Pereira[2], A. Rossi[18], J. A. D. Sena[15], C. Silva[13], R. F. de Souza[1], L. A. F. Spinola[13], M. A. Takita[7], R. E. Tamura[13], E. C. Teixeira[8], R. I. D. Tezza[2], M. Trindade dos Santos[1], D. Truffi[4], S. M. Tsai[9], F. F. White[2,20], J. C. Setubal[16] & J. P. Kitajima[19,20]

[1]*Departamento de Bioquímica, Instituto de Química, Av. Prof. Lineu Prestes 748;* [5]*Departamento de Botânica, Instituto de Biociências, Rua do Matão 277;* [12]*Departamento Microbiologia, Instituto de Ciências Biomédicas, Av. Prof. Lineu Prestes 1374;* [13]*Departamento de Patologia, Faculdade de Medicina Veterinária e Zootecnia, Av. Prof. Dr. Orlando Marques de Paiva 87; and* [17]*Departamento de Biologia, Instituto de Biociências, Rua do Matão 277, Universidade de São Paulo, 05508-900, São Paulo, SP, Brazil*

[2]*Departamento de Tecnologia, and* [15]*Departamento de Biologia Aplicada à Agropecuária, Faculdade de Ciências Agrárias e Veterinárias, Universidade Estadual Paulista, Via de Acesso Prof. Paulo Donato Castellane s/n, 14884-900, Jaboticabal, SP, Brazil*

[3]*Departamento de Melhoramento e Nutrição Animal, Faculdade de Medicina Veterinária e Zootecnia, Universidade Estadual Paulista, Fazenda Lageado, Caixa Postal 560, 18600-000, Botucatu, SP, Brazil*

[4]*Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo, Av. Pádua Dias 11, 13418-900, Piracicaba, SP, Brazil*

[6]*Departamento de Computação e Estatística, Universidade Federal de Mato Grosso do Sul, Cx Postal 549, 79070-900, Campo Grande, MS, Brazil*

[7]*Centro de Citricultura Sylvio Moreira—IAC, Rod. Anhanguera, km 158, Cx. Postal 04, 13.490-970, Cordeirópolis, SP, Brazil*

[8]*Departamento de Bioquímica e Tecnologia Química, Instituto de Química, Universidade Estadual Paulista, Cx. Postal 355, 14801-970, Araraquara, SP, Brazil*

[9]*Centro de Energia Nuclear na Agricultura, Av. Centenário 303, 13400-970, Piracicaba, SP, Brazil*

[10]*Departamento de Ciências Biológicas, Faculdade Ciências Farmacêuticas, Universidade Estadual Paulista, Rod. Araraquara-Jaú km 01, 14.801-902, Araraquara, SP, Brazil*

[11]*Departamento de Genética, and* [18]*Departamento de Bioquímica e Imunologia, Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, Av. Bandeirantes 3900, 14049-900, Ribeirão Preto, SP, Brazil*

[14]*Instituto de Tecnologia do Paraná, IAPAR, Cx. Postal 481, 86001-970, Londrina, PR, Brazil*

[16]*Instituto de Computação, and* [19]*Centro de Biologia Molecular e Engenharia Genética/Instituto de Computação, Universidade Estadual de Campinas, Cx Postal 6176, 13084-971, Campinas, SP, Brazil*

[20]*Present addresses: Alellyx Applied Genomics, Rua James Clark Maxwell 320, Techno Park, 13067-850, Campinas, SP, Brazil (F.C.R. and J.P.K.); Department of Plant Pathology, Kansas State University, 4024 Throckmorton Hall, Manhattan, Kansas 66506, USA (F.F.W.)*

The genus *Xanthomonas* is a diverse and economically important group of bacterial phytopathogens, belonging to the γ-subdivision of the Proteobacteria. *Xanthomonas axonopodis* pv. *citri* (*Xac*) causes citrus canker, which affects most commercial citrus cultivars, resulting in significant losses worldwide. Symptoms include canker lesions, leading to abscission of fruit and leaves and general tree decline[1]. *Xanthomonas campestris* pv. *campestris* (*Xcc*) causes black rot, which affects crucifers such as *Brassica* and *Arabidopsis*. Symptoms include marginal leaf chlorosis and darkening of vascular tissue, accompanied by extensive wilting and necrosis[2]. *Xanthomonas campestris* pv. *campestris* is grown commercially to produce the exopolysaccharide xanthan gum, which is used as a viscosifying and stabilizing agent in many industries[3]. Here we report and compare the complete genome sequences of *Xac* and *Xcc*. Their distinct disease phenotypes and host ranges belie a high degree of similarity at the genomic level. More than 80% of genes are shared, and gene order is conserved along most of their respective chromosomes. We identified several groups of strain-specific genes, and on the basis of these groups we propose mechanisms that may explain the differing host specificities and pathogenic processes.

Xanthomonas axonopodis pv. *citri* (strain 306) has one circular chromosome comprising 5,175,554 base pairs (bp), and two plasmids: pXAC33 (33,699 bp) and pXAC64 (64,920 bp). *Xanthomonas campestris* pv. *campestris* (strain ATCC33913) has a single circular chromosome comprising 5,076,187 bp, and no plasmids (Table 1). The chromosomes of the two organisms have a high degree of collinearity (Fig. 1). The alignment shown in Fig. 1 suggests that just three major rearrangement events have occurred. One is an inversion around the putative terminus of replication (Fig. 1; A); the others (Fig. 1; B1 and B2) are two inversions with translocations symmetrically located with respect to the putative origin of replication. A whole-chromosome alignment at the protein level, excluding transposable elements, paired 2,929 orthologous genes, corresponding to about 70% of the total of each (Fig. 2). There are 512 non-collinear (or translocated) pairs of orthologous genes scattered throughout the two chromosomes. *Xanthomonas campestris* pv. *campestris* contains 646 genes (15.4%) not found in *Xac*, of which 291 have functional assignment. *Xanthomonas axonopodis* pv. *citri* contains 800 genes (18.5%) not found in *Xcc*, of which 316 have functional assignment.

Table 1 General features of *X. campestris* pv. *campestris* and *X. axonopodis* pv. *citri* genomes

(**a**) General features of the chromosomes

	Xcc	*Xac*
Length (bp)	5,076,187	5,175,554
G+C content (%)	65.0	64.7
Protein-coding region (% chromosome size)	84.34	85.59
Protein-coding genes		
With assigned function	2,708	2,710
Conserved hypothetical	1,276	1,272
Hypothetical	198	331
Total	4,182	4,313
Transfer RNA	53	54
Ribosomal RNA operons	2	2
Plasmids	0	2
Insertion sequence elements (IS)	109	87

(**b**) Sequence similarity between protein-coding genes

≥80% amino acid identity (%)	82.1
≥90% amino acid identity (%)	48.9
≥80% nucleotide identity (%)	78.3
≥90% nucleotide identity (%)	18.3

(**c**) General features of *X. axonopodis* pv. *citri* plasmids

	pXAC33	pXAC64
Length (bp)	33,699	64,920
G+C content (%)	61.9	61.4
Coding region (% plasmid size)	81.59	89.26
Protein-coding genes		
With assigned function	21	39
Conserved hypothetical	7	7
Hypothetical	14	27
Total	42	73
Insertion sequence elements (IS)	11	10

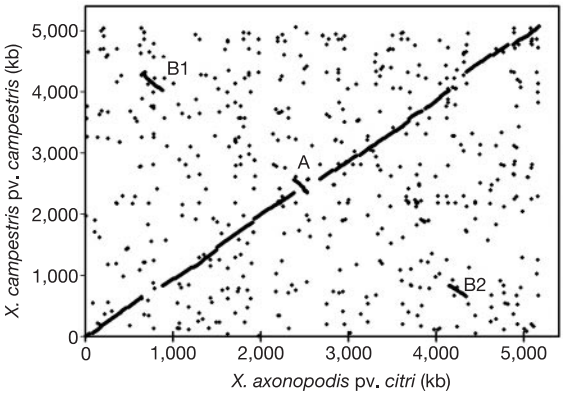

Figure 1 Nucleotide alignment between *Xac* and *Xcc*. Each point is a maximal unique match, that is, a maximal sequence of at least 25 bp that occurs only once in *Xac* and only once in *Xcc*. Each match is exact. Labels A, B1 and B2 are rearrangement events between the genomes.

In both genomes, the respective regions around the putative termini of replication (labelled A in Figs 1 and 2) contain a large number of strain-specific genes, suggesting that these regions are susceptible to gene acquisition and/or gene loss. Notably, some of these genes have functional assignments suggestive of involvement in host colonization and pathogenesis. Genes specific to *Xac* include: (1) *syrE*-related genes, similar to those responsible for the synthesis of the fungicide syringomycin in *Pseudomonas syringae*[4]; (2) a 133,025-bp insertion that harbours many genes not commonly found in γ-proteobacteria and 19 genes also found in the citrus pathogen *Xylella fastidiosa*, including an RTX-like toxin gene with its secretion system (*hlyBD*)[5]; and (3) genes related to the *nodV/nodW* two-component system, which in *Bradyrhizobium japonicum* respond to plant-derived isoflavonone signals and participate in an alternative regulatory pathway for nodulation[6]. Genes specific to *Xcc* include: (1) two copies of the ΦLf filamentous phage[7]; (2) a gene for cyclic β-1,2-glucan synthetase, associated with virulence/colonization in members of the *Rhizobiaceae* family (*Agrobacterium*, *Rhizobium* and *Brucella*)[8]; (3) the avirulence gene *avrBs1* (ref. 9); (4) a tannase, which allows microorganisms to overcome growth inhibition induced by plant tannins[10]; (5) four genes similar to those involved in antibiotic synthesis in *Streptomyces* (macrolides and streptomycin)[11]; and (6) a cluster of genes related to nitrate assimilation[12].

The two genomes are rich in transposable elements. We identified 109 genes in *Xcc* and 108 genes in *Xac* that are grouped in 27 different types of transposable element. Only four group types are present in both genomes. In *Xcc*, the IS*5* family is highly represented, with 16 copies of IS*1478* (ref. 13), whereas in *Xac* the IS*3* family is more abundant, with 21 copies of a member not previously described in *Xanthomonas* (ISXac*3*). Many of these insertion sequences (IS) are located near strain-specific genes, in regions with altered codon usage and G+C content, which suggests that these genes may have been acquired through lateral transfer.

Phylogenetic analyses place *Xac* and *Xcc* at the base of the γ-subdivision of proteobacteria, in the same clade as *X. fastidiosa*. We estimate the divergence time between *Xac* and *Xcc* to be 6.5 to 8.9 million years. Comparison of the predicted *Xac* proteome against other completed genomes (see Supplementary Information) indicates that approximately 40% of all genes found in *Xac* are most similar to genes from non-γ-proteobacteria. Many of these genes have orthologues in the α-proteobacteria *Mesorhizobium loti*, *Sinorhizobium meliloti* and *Agrobacterium tumefaciens*, which have plants as hosts, suggesting that they might be involved in plant–bacterium interactions.

The two *Xanthomonas* pathogens have a large and diversified set of pathways for intermediary, small molecule, and DNA metabolism. The DNA restriction/modification system of *Xac* has one gene cluster for a type I system and another for a type II system, whereas *Xcc* has two clusters for a type I system, but no type II system. This difference may be associated to the noted differences in mobile genetic element content in each genome. *Xanthomonas campestris* pv. *campestris* has a cluster of seven strain-specific genes (*cysG*, *nasTACDEF*) that are required to assimilate and convert nitrate and nitrite into ammonium in *Klebsiella pneumoniae*[12]. *Xanthomonas axonopodis* pv. *citri* has strain-specific genes for an ABC-type oligopeptide transport system (*oppA*, *oppB* and *oppC*), which may facilitate the entry of small oligopeptide products for use as a nitrogen source. *Xanthomonas axonopodis* pv. *citri* also contains

Table 2 Proposed *hrpX* regulons

Xanthomonas campestris pv. *campestris*				*Xanthomonas axonopodis* pv. *citri*			
PIP position	Distance (bp)*	Gene ID	Gene product	PIP position	Distance (bp)*	Gene ID	Gene product
hrp gene cluster				*hrp* gene cluster			
1,434,713	250	XCC1227	HrcQ	475,032	249	XAC0403	HrcQ
1,438,281	70	XCC1230	HrcU	478,627	70	XAC0406	HrcU
1,438,344	80	XCC1231	HrpB1	478,642	81	XAC0407	HrpB1
1,446,618	139	XCC1240	Hpa1	487,556	135	XAC0416	Hpa1
1,446,733	280	XCC1241	Hpa2	488,189	271	XAC0417	Hpa2
Extended Hrp conserved regulon				Extended Hrp conserved regulon			
323,682	65	XCC0258	*Xanthomonas* conserved hypothetical	332,388	64	XAC0277	*Xanthomonas* conserved hypothetical
1,962,252	103	XCC1686	Conserved hypothetical	1,966,043	149	XAC1706	Conserved hypothetical
2,172,034	123	XCC1864	β-K-adipate enol-lactone hydrolase	2,191,464	122	XAC1886	β-K-adipate enol-lactone hydrolase
2,666,140	241	XCC2266	Polygalacturonase	2,767,045	179	XAC2374	Polygalacturonase
2,848,531	184	XCC2399	Conserved hypothetical	2,982,678	153	XAC2534	Conserved hypothetical
3,737,359	64	XCC3157	Aminopeptidase	3,886,531	63	XAC3309	Aminopeptidase
4,118,186	118	XCC3459	Endopolygalacturonase	783,305	121	XAC0661	Endopolygalacturonase
Additional Hrp regulon candidate ORFs				Additional Hrp regulon candidate ORFs			
1,010,154	108	XCC0851	Extracellular protease	40,366	78	XAC0337	2-K-3-DdG permease
1,238,085	109	XCC1072	*Xanthomonas* conserved hypothetical	501,099	107	XAC0424	Conserved hypothetical with GGDEF domain
1,569,371	89	XCC1348	*Xanthomonas* conserved hypothetical	1,657,406	271	XAC1435	Iron receptor
3,190,239	217	XCC2693	Cysteine protease	2,763,222	227	XAC2370	Conserved hypothetical
				3,805,982	66	XAC3230	Hypothetical protein
3,871,764	65	XCC3258	Conserved hypothetical	3,428,498	165	XAC2922	HrpW
				4,769,953	147	XAC4074	Ribonucleotide-diphosphate reductase
				4,796,299	133	XAC4090	3-oxoacyl-[ACP] reductase

ORF open reading frame.
*Distance (bp) upstream of predicted start codon.

additional strain-specific genes related to proteases, including three genes similar to secreted xanthomonapepsins/pseudomonapepsins. These observations suggest that the two organisms have different nitrate assimilation and oligopeptide absorption capabilities, and this in turn may be linked to their different optimal environments for growth.

Both *Xac* and *Xcc* have an extensive repertoire of genes for cell-wall degradation. Both genomes code for enzymes with cellulolytic, pectinolytic and hemicellulolytic activities (Supplementary Information). *Xanthomonas campestris* pv. *campestris* has more genes involved in pectin and cellulose degradation than *Xac*. In addition, *Xcc* has two 1,4-β-cellobiosidases and two pectin esterases, none of which are found in *Xac*. The lack of these enzymes does not necessarily preclude the degradation of cellulose and pectin by *Xac*. In fact, in citrus canker there is tissue maceration, albeit localized. Black rot, on the other hand, presents massive degeneration of plant tissue. These differences in symptoms may be correlated to the noted differences in genes for cell-wall degradation.

The synthesis of extracellular degrading enzymes and exopolysaccharides is transcriptionally regulated by products of the *rpf* gene genes. This cluster in *Xcc* has an organization (*rpfABFCHGDIE*) similar to that previously described[14,15]. *Xanthomonas axonopodis* pv. *citri* lacks *rpfH* and *rpfI*, and *X. fastidiosa* also lacks the corresponding genes[16]. The expression levels of proteases and endoglucanases are reduced when the *rpfI* gene is inactivated in *Xcc*[15], suggesting that *rpfI* may have a function in the massive tissue degeneration observed in *Xcc* infection. The locus corresponding to *rpfI* in *Xac* is occupied by a truncated copy of an insertion sequence.

Motility in a number of plant pathogenic species is important for infection[17]. The genomic sequences of *Xac* and *Xcc* provide a detailed view of the flagellar and chemotactic biosynthetic pathways of *Xanthomonas* species. Chemotaxis receptors and flagella biogenesis genes are organized in four clusters spread over 150 kb in both *Xanthomonas* genomes. One of these clusters has several nearly identical tandemly repeated copies (ten in *Xac* and eight in *Xcc*) of a methyl-accepting chemotaxis protein gene (*mcp*). Such large numbers of tandemly repeated copies are unusual in bacteria, suggesting a prominent role for chemotaxis in the two *Xanthomonas* pathogens. In fact, *Xac* has another ten copies of the *mcp* gene scattered throughout its genome, whereas *Xcc* has eleven additional copies. Both genomes have genes for type IV fimbriae (pili) and glycine-rich outer membrane proteins, which are principal structures for host colonization and adhesion in many pathogenic bacteria[18]. However, the fimbrillin genes, whose products may come in direct contact with host cells or other bacteria, are different in *Xac* and *Xcc*, probably owing to adaptation to specific interactions. Both genomes have copies (two in *Xac*, one in *Xcc*) of a gene that codes for XadA, which is an outer membrane protein found in *Xanthomonas oryzae* pv. *oryzae* (where it is implicated in virulence) and in *Xanthomonas campestris* pv. *vesicatoria* (where it is repressed by *HrpG*[19]). The presence of these proteins in two other *Xanthomonas* species may indicate a general importance of these proteins to *Xanthomonas* pathogenicity.

Distinct gene clusters for O-antigen synthesis are present in the genomes of both *Xac* and *Xcc*. They are located in different regions of the respective chromosomes and lack significant sequence similarity. This is consistent with the observation that lipopolysaccharide (LPS) O antigen is pathovar/strain-specific and may be involved in host-range selection and pathogenicity by acting as a barrier against plant toxins[20]. In *Xcc*, these genes are organized in a single cluster as previously reported[21], whereas in *Xac*, the genes are divided into two regions. The first region contains genes that code for transferases, epimerases, translocases, and derived sugar transport proteins, and the second contains the *xanAB* and *rmlDABC* genes for nucleotide-sugar and dTDP-L rhamose biosynthesis[21].

In each genome there are two type II secretion systems that are involved in the secretion of degradative enzymes and toxins. One is the previously described *xps* cluster[22], whereas the genes in the second cluster are most similar to genes found in *Caulobacter crescentus*. These clusters are associated with the genomic rearrangement events B1 and B2 (Figs 1 and 2). The type III secretion system (*hrp* pathway) is critical for pathogenicity and initiation of disease. A cluster composed of 26 genes extending from *hpa2* to *hrpF* is found in both genomes. Although these clusters are located in different positions in each genome (Fig. 2), gene order is similar and consistent with the previously characterized *hrp* cluster[23]. One exception is *hrpW*, which is located between *hrpF* and *hrpE* in *Xcc*, but outside the *hrp* cluster in *Xac*. The presence of *hrpW* in *Erwinia amylovora*, *Pseudomonas syringae* and now in *Xac* and *Xcc* indicates that *hrpW* is a general core component of plant-pathogen type III systems. The amino acid identity between the orthologous type III genes in *Xac* and *Xcc* is high (greater than 80% on average). The least similar (less than 61% identity) orthologous genes are *HpaP*, *Hpa1*, *HpaA*, *HrpE*, *HrpF* and *HrpW*. It may be significant that some of the proteins coded for by these latter genes are predicted to be exposed components or secreted by the type III secretion system, and differences in these components may reflect different selective pressures due to strain-specific interactions with host plants. Type

Table 3 Putative effector/avirulence genes of *Xcc* and *Xac*

Gene ID	Name	Family	*Xcc*/*Xac**	PIP box†	Location‡	AME§	Motifs‖
XCC0052/XAC0076	*avrBs2*	*avrBs2*	Y/Y	Y/Y	C	N	n
XCC1629/XAC0286	*avrXccE1/avrXacE1*	*avrPphE*	Y/Y	Y/Y	C	Y/N	m
XAC3224	*avrXacE2*	*avrPphE*	N/Y	N	C	Y	m,n
XACb0011	*avrXacE3*	*avrPphE*	N/Y	Y	P	Y	m,n
XACa0022	*pthA1*	*avrBs3*	N/Y	N	P	Y	n
XACa0039	*pthA2*	*avrBs3*	N/Y	N	P	Y	n
XACb0015	*pthA3*	*avrBs3*	N/Y	N	P	Y	n
XACb0065	*pthA4*	*avrBs3*	N/Y	N	P	Y	n
XCC2100	*avrBs1*	*avrBs1*	Y/N	N	C	Y	n
XCC2099	*avrBs1.1*	*avrBs1*	Y/N	N	C	Y	
XCC2109	*avrXccC*	*avrC*	Y/N	Y	C	Y	
XCC3731	*avrXccB*	*yopJ*	Y/N	Y	C	Y	
XCC4229	*avrXccA1*	*avrXca*	Y/N	N	C	N	sp
XCC2396	*avrXccA2*	*avrXca*	Y/N	N	C	N	n
XCC4186/XAC3090	Leucine-rich protein	*popC*	Y/Y	Y/N	C	N/N	lrr
XAC0393	*hpaF*	*popC*	Y/N	Y	C	N	lrr
XCC2565	Leucine-rich protein	*popC*	N/Y	Y	C	N	lrr

*Presence in *Xcc* or *Xac* genome: Y, yes; N, No.
†Presence of PIP motif near promoter: Y, yes; N, No.
‡Location in genome: C, chromosome; P, plasmid.
§AME (associated mobile genetic elements): Y, yes; N, no.
‖Amino acid sequence motifs as determined by PSORT or BLAST. m, myristolyation; n, nuclear localization; lrr, leucine-rich receptor; sp, bacterial signal peptide cleavage site.

IV secretion systems have been well characterized in organisms such as *Agrobacterium tumefaciens*, where the *virB* operon is required for the transfer of the T-DNA to the plant cell[24]. Other bacteria use this system for conjugal transfer and secretion of toxins or other proteins[24]. *Xanthomonas axonopodis* pv. *citri* has two type IV systems, one in the chromosome and one in the pXAC64 plasmid. Gene content in these clusters is different from that of the *Agrobacterium virB* operon: *virB5* and *virB7* are missing from both clusters, and *virD4* is missing in the plasmid. *Xanthomonas campestris* pv. *campestris* has only one *virB* cluster, which is almost identical to that of the *Xac* chromosome, except that *virB6* is outside the cluster. This is the first report to our knowledge of a type IV system in any *Xanthomonas* species, and its precise function remains to be elucidated.

In the genus *Xanthomonas*, expression of structural genes of the type III secretion system and some effector genes are mediated, in part, by the *hrpX* and *hrpG* gene products[25], both of which are present in *Xac* and *Xcc*. Several genes that are regulated in a HrpX-dependent manner possess the consensus nucleotide sequence TTCGC...N_{15}...TTCGC, which has been termed the plant-inducible-promoter box, or PIP box[26]. Detection of a PIP box or a sequence similar to it provides a tool for identification of candidate genes in the *hrpX* regulon as well as effector protein genes of the type III pathway. We found 17 (*Xcc*) and 20 (*Xac*) copies of the sequence TTCG...N_{16}...TTCG (a putative PIP box; separated by 16 nucleotide sequences) in promoter regions (Table 2). Among these, 12 are associated with the same genes in both genomes. Five of these genes are in the *hrp* gene cluster and seven are scattered on the chromosomes, including three with amino-terminal type II signal peptide sequences and similarity to cell-wall-degrading enzymes. Of the five other putative PIP boxes in *Xcc*, two are associated with protease genes. Of the eight other putative PIP boxes in *Xac*, one is associated with an iron receptor gene.

Both genomes have a variety of putative avirulence/effector protein-coding genes, which are fundamental to the development or restriction of the disease (Table 3). *Xanthomonas campestris* pv. *campestris* contains a more diverse array of known types of effectors. However, it lacks *avrBs3/pth*, which is found in *Xac* as well as in many *Xanthomonas* species, and in *Ralstonia solanacearum*[27]. *Xanthomonas axonopodis* pv. *citri* contains four copies of *avrBs3/pth*, located in the plasmids and without recognizable PIP boxes. These four copies are distinct from each other and from the previously described *pthA*[28]. *Xanthomonas axonopodis* pv. *citri* also contains three genes with similarity to *avrPphE*, originally identified in *Pseudomonas syringae* pv. *phaseolicola*—*Xcc* has only one *avrPphE* gene. Notably, *Xac* lacks a recognizable YopJ homologue, a protease-coding gene that has been found in many other pathogenic bacteria[27] (including animal pathogens). It would be interesting to verify whether the differences noted above contribute significantly to the different infection modes of the two *Xanthomonas* pathogens. Both *Xac* and *Xcc* contain genes coding for leucine-rich-repeat (LRR) proteins. LRR motifs are commonly involved in protein–protein interactions and are found in the three major classes of plant-resistance genes[29] and in the PopC protein of *Ralstonia solanacearum*[30]. The presence of distinct genes for LRR proteins in these genomes, under control of putative PIP boxes, suggests that they may have an important function at some stage of plant colonization.

Comparative analysis between the two *Xanthomonas* strains allowed us to identify a set of strain-specific genes, some of which are probably responsible for the distinct pathogenicity and host specificity profiles of these organisms. *Xanthomonas campestris* pv. *campestris* is uniquely suited to invade and colonize host tissue, a fact that may partially explain the systemic nature of its infection. In contrast, *Xac* induces a strong local response with cell proliferation and necrosis, but shows little spontaneous dissemination, probably due to a smaller number of genes capable of causing a massive degeneration of host tissue. Besides the already known *pthA* genes, other candidates involved in eliciting this strong host response should emerge experimentally by testing the roles of a selected list of *Xac*-specific genes currently without functional assignment. Host specificity may result from a combination of differential subsets of genes found in each genome, such as *avr*, type III secretion system, *rpf*, type IV fimbriae, and LPS O-antigen operons. Future experimental tests of the several hypotheses put forward here will be facilitated now that the genomes of a plant pathogen (*Xcc*) and one of its hosts (*Arabidopsis thaliana*) are publicly available. □

Figure 2 Comparative genome map. The organizations of *Xcc* (upper) and the *Xac* (lower) genomes are represented side by side. Predicted proteins are indicated by vertical bars coloured according to orthologue pairing and collinearity. The genomes were aligned by determining a longest common subsequence (LCS) of orthologues, with manual adjustments. Light blue, collinear orthologous proteins; dark blue, translocated (that is, not belonging to the LCS) orthologous proteins; yellow, *Xcc*-specific proteins; pink, *Xac*-specific proteins. Green flags indicate transposases. Rearrangement events are labelled A, B1 and B2, and are indicated by red lines. Numbers in black represent genome coordinates in base pairs. Features in each genome are labelled as follows: O antigen, O-antigen biosynthesis cluster; Xcs and Xps, secretory pathway II; Hrp, secretory pathway III; SyrE, syringomicin-related proteins; GS, cyclic β-1,2-glucan synthetase; Tan, tannase protein; RTX, haemolysin calcium-binding protein; RM.I or .II, restriction and modification system type I or II; XccP1 or XacP1, ΦCTX-type prophage; Vir, VirB proteins; Lf, ΦLf prophage; RpfI, RpfI protein; RpfH, RpfH protein; Nas, Nitrate assimilation proteins; Opp, oligopeptide transport system; M/S, macrolide/streptomycin-related proteins; VW, NodV and NodW proteins.

Methods

The sequencing and analysis in this work were carried out by the *Xanthomonas* consortium of the Organization for Nucleotide Sequencing and Analysis (ONSA), a group of 13 biology laboratories and one bioinformatics centre, located in the state of São Paulo, Brazil.

Sequencing and assembly

We sequenced both genomes using shotgun methodologies. Several pUC18 libraries were generated with inserts ranging from 1 to 4 kb in size, and sequenced from both ends. From these libraries 205,408 (*Xac*) and 174,051 (*Xcc*) reads were sequenced. Several cosmid libraries (using lawrist vector) were generated, with inserts ranging from 30 to 45 kb in size. Both ends of these inserts were sequenced, resulting in 7,380 (*Xac*) and 5,476 (*Xcc*) ends. Assembly was carried out using the phred/phrap/consed package and a scaffolding program developed by our bioinformatics team. Assembly was confirmed by comparing *Ceu*I, *Swa*I and *Pme*I restriction maps to computational predictions. Genome assembly yielded 103 virtual gaps in *Xac* and 135 in *Xcc*. Many of these gaps were a result of the repeat screening process required for scaffolding. Gaps were closed by whole-plasmid and cosmid sequencing, plus refill of screened sequences. The total number of whole plasmids sequenced was 20 (*Xac*) and 67 (*Xcc*). The total number of whole cosmids sequenced was 49 (*Xac*) and 13 (*Xcc*). The final *Xac* sequence has an estimated overall error rate of less than 1 in 10,000 bp, and no base has phred error quality below 20. The *Xcc* sequence has 15 uncertain bases due to resolution errors within regions of secondary structure.

Putative origins of replication

On the basis of similarity with *X. fastidiosa*, we hypothesize that the origins of replication for both chromosomes are located in a region between the 50S ribosomal protein L34 gene and the *gyrB* gene. This is supported by predictions based on GC-skew inversion. The origins of the *Xac* plasmids were chosen according to GC-skew inversion and within a region that did not contain predicted genes.

Genome annotation

Annotation was carried out using a system developed by our bioinformatics team. Putative protein-coding genes were identified using GeneMark and Glimmer. Curators assigned functions to these genes based on comparisons to sequences available in public databases. RNA species were identified using BLASTN and tRNAscan-SE. Metabolic pathways were analysed using the KEGG website (http://www.jaist.genome.ad.jp/kegg). Transporter proteins were annotated based on BLAST comparisons of predicted proteins in each proteome against a custom database of transporters, which was created based on data available from http://www-biology.ucsd.edu/ ~ msaier/transport/toc.html#1A.

Phylogeny

Phylogenetic analyses were performed based on 16S ribosomal RNA sequences using the program DNADIST from PHYLIP, with a distance matrix based on the two-parameter Kimura model. Divergence time was estimated based on the 16S rRNA sequences. A constant rate of 1–2% per 50 million years was assumed.


300,000
400,000
300,000
400,000
B1
700,000
700,000
800,000
Xcs
B2
1,000,000
1,101,000
1,100,000
1,202,000
Hrp
1,500,000
1,600,000
1,503,000
1,600,000
1,700,000
2,001,000
2,102,000
2,001,000
2,106,000
A
Nas
M/S
GS
Lf
Tan
2,400,000
2,501,000
SyrE
2,502,000
VW
A
2,701,000
2,801,000
2,900,000
2,800,000
RM.II
2,901,000
3,001,000
3,100,000
3,300,000
3,402,000
3,500,000
RM.I
3,700,000
3,800,000
3,900,000
Xcs
B2
4,100,000
4,200,000
B1
O antigen.II
4,300,000
B1
4,500,000
4,600,000
4,500,000
4,600,000
4,900,000
5,000,000
5,075,000
Xcc
Xac
5,001,000
5,100,000
5,175,000

Whole-genome comparisons between *Xac* and *Xcc*

We compared the genomes at the nucleotide level using the program MUMmer with default values. At the amino acid level the genomes were compared using programs developed by our bioinformatics team. Genes *g* and *h* were considered orthologues if *h* is the best BLASTP hit for *g* and vice versa, with *e*-values less than or equal to 10^{-20}. A gene was considered strain-specific if it had no hits with an *e*-value 10^{-5} or less.

Received 11 December 2001; accepted 26 March 2002.

1. Gottwald, T. R. & Graham, J. H. in *Compendium of Citrus Diseases* (eds Timmer, L. W., Garnsey, S. M. & Graham, J. H.) 5–7 (Am. Phytopathol. Soc., St. Paul, 2000).
2. Hayward, A. C. in *Xanthomonas* (eds Swings, J. G. & Civerolo, E. L.) 51–54 (Chapman & Hall, London, 1993).
3. Becker, A., Katzen, F., Puhler, A. & Ielpi, L. Xanthan gum biosynthesis and application: a biochemical/genetic perspective. *Appl. Microbiol. Biotechnol.* **50,** 145–152 (1998).
4. Menestrina, G. & Semjen, B. V. in *The Comprehensive Sourcebook of Bacterial Proteins Toxins* (ed. Alouf, J. E.) 287–309 (Academic, London, 1999).
5. Simpson, A. J. *et al.* The genome sequence of the plant pathogen *Xylella fastidiosa*. *Nature* **406,** 151–157 (2000).
6. Göttfert, M., Grob, P. & Hennecke, H. Proposed regulatory pathway encoded by the *nodV* and *nodW* genes, determinants of host specificity in *Bradyrhizobium japonicum*. *Proc. Natl Acad. Sci. USA* **87,** 2680–2684 (1990).
7. Chang, K. H. *et al.* Sequence analysis and expression of the filamentous phage phi Lf gene I encoding a 48-kDa protein associated with host cell membrane. *Biochem. Biophys. Res. Commun.* **245,** 313–318 (1998).
8. Breedveld, M. W., Hadley, J. A. & Miller, K. J. A novel cyclic β-1,2-glucan mutant of *Rhizobium meliloti*. *J. Bacteriol.* **177,** 6346–6351 (1995).
9. Ronald, P. C. & Staskawicz, B. J. The avirulence gene *avrBs1* from *Xanthomonas campestris* pv. *vesicatoria* encodes a 50-kD protein. *Mol. Plant-Microbe Interact.* **1,** 191–198 (1988).
10. Bhat, T. K., Singh, B. & Sharma, O. P. Microbial degradation of tannins—a current perspective. *Biodegradation* **9,** 343–357 (1998).
11. Beyer, S., Distler, J. & Piepersberg, W. The *str* gene cluster for the biosynthesis of 5′-hydroxy-streptomycin in *Streptomyces glaucescens* GLA.0 (ETH 22794): new operons and evidence for pathway-specific regulation by *StrR*. *Mol. Gen. Genet.* **250,** 775–784 (1996).
12. Lin, J. T., Goldman, B. S. & Stewart, V. The *nasFEDCBA* operon for nitrate and nitrite assimilation in *Klebsiella pneumoniae* M5al. *J. Bacteriol.* **176,** 2551–2559 (1994).
13. Chen, J. H., Hsieh, Y. Y., Hsiau, S. L., Lo, T. C. & Shau, C. C. Characterization of insertions of IS476 and two newly identified insertion sequences, IS1478 and IS1479, in *Xanthomonas campestris* pv. *campestris*. *J. Bacteriol.* **181,** 1220–1228 (1999).
14. Tang, J. L. *et al.* 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 (1991).
15. Dow, J. M., Feng, J. X., Barber, C. E., Tang, J. L. & Daniels, M. J. Novel genes involved in the regulation of pathogenicity factor production within the rpf gene cluster of *Xanthomonas campestris*. *Microbiology* **146,** 885–891 (2000).
16. Dow, J. M. & Daniels, M. J. *Xylella* genomics and bacterial pathogenicity to plants. *Yeast* **17,** 263–271 (2000).
17. Rudolf, K. in *Xanthomonas* (eds Swings, J. G. & Civerolo, E. L.) 193–264 (Chapman & Hall, London, 1993).
18. Cao, H., Baldini, R. L. & Rahme, L. G. Common mechanisms for pathogens of plants and animals. *Annu. Rev. Phytopathol.* **39,** 259–284 (2001).
19. Noël, L., Thieme, F., Nennstiel, D. & Bonas, U. cDNA-AFLP analysis unravels a genome-wide *hrpG*-regulon in the plant pathogen *Xanthomonas campestris* pv. *vesicatoria*. *Mol. Microbiol.* **41,** 1271–1281 (2001).
20. Dow, J. M., Osbourn, A. E., Wilson, T. J. & Daniels, M. J. A locus determining pathogenicity of *Xanthomonas campestris* is involved in lipopolysaccharide biosynthesis. *Mol. Plant-Microbe Interact.* **8,** 768–777 (1995).
21. Vorhölter, F. J., Niehaus, K. & Puhler, A. 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 (2001).
22. Dums, F., Dow, J. M. & Daniels, M. J. Structural characterization of protein secretion genes of the bacterial phytopathogen *Xanthomonas campestris* pathovar *campestris*: relatedness to secretion systems of other gram-negative bacteria. *Mol. Gen. Genet.* **229,** 357–364 (1991).
23. Bonas, U. *et al.* Isolation of a gene cluster from *Xanthomonas campestris* pv *vesicatoria* that determines pathogenicity and the hypersensitive response on pepper and tomato. *Mol. Plant-Microbe Interact.* **4,** 81–88 (1991).
24. Burns, D. L. Biochemistry of type IV secretion. *Curr. Opin. Microbiol.* **2,** 25–29 (1999).
25. Wengelnik, K. & Bonas, U. HrpXv, an AraC-type regulator, activates expression of five of the six loci in the hrp cluster of *Xanthomonas campestris* pv. *vesicatoria*. *J. Bacteriol.* **178,** 3462–3469 (1996).
26. Fenselau, S. & Bonas, U. Sequence and expression analysis of the hrpB pathogenicity operon of *Xanthomonas campestris* pv. *vesicatoria* which encodes eight proteins with similarity to components of the Hrp, Ysc, Spa, and Fli secretion systems. *Mol. Plant-Microbe Interact.* **8,** 845–854 (1995).
27. Lahaye, T. & Bonas, U. Molecular secrets of bacterial type III effector proteins. *Trends Plant Sci.* **6,** 479–485 (2001).
28. Yang, Y. & Gabriel, D. W. Intragenic recombination of a single plant pathogen gene provides a mechanism for the evolution of new host specificities. *J. Bacteriol.* **177,** 4963–4968 (1995).
29. Young, N. D. The genetic architecture of resistance. *Curr. Opin. Plant Biol.* **3,** 285–290 (2000).
30. Gueneron, M., Timmers, A. C., Boucher, C. & Arlat, M. Two novel proteins, PopB, which has functional nuclear localization signals, and PopC, which has a large leucine-rich repeat domain, are secreted through the hrp-secretion apparatus of *Ralstonia solanacearum*. *Mol. Microbiol.* **36,** 261–277 (2000).

Supplementary Information accompanies the paper on *Nature*'s website (http://www.nature.com).

Acknowledgements

We thank A. M. da Silva, S. Verjovski-Almeida and D. W. Wood for reading of the manuscript; our steering committee (S. Oliver, A. Goffeau, J. Sgouros, A. C. M. Paiva and J. L. Azevedo) for their critical accompaniment; and all the technicians in the sequencing laboratories of ONSA involved in this project. Project funding was from FAPESP, Fundecitrus, Fundect-MS and CNPq.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to A.C.R.d.S. (e-mail: acrasera@iq.usp.br). The sequences have been deposited in GenBank under accession numbers AE008922 (*Xcc* chromosome), AE008923 (*Xac* chromosome), AE008924 (pXAC33) and AE008925 (pXAC64).

The catalytic pathway of horseradish peroxidase at high resolution

Gunnar I. Berglund*†, Gunilla H. Carlsson*†, Andrew T. Smith‡, Hanna Szöke*§, Anette Henriksen§‖ & Janos Hajdu*

* *Department of Biochemistry, Uppsala University, Biomedical Center, Box 576, S-751 23 Uppsala, Sweden*
‡ *School of Biological Sciences, University of Sussex, Brighton BN1 9QG, UK*
‖ *Protein Structure Group, Department of Chemistry, University of Copenhagen, Universitetsparken 5, DK-2100 København, Denmark*
† *These authors contributed equally to this work*

A molecular description of oxygen and peroxide activation in biological systems is difficult, because electrons liberated during X-ray data collection reduce the active centres of redox enzymes catalysing these reactions[1–5]. Here we describe an effective strategy to obtain crystal structures for high-valency redox intermediates and present a three-dimensional movie of the X-ray-driven catalytic reduction of a bound dioxygen species in horseradish peroxidase (HRP). We also describe separate experiments in which high-resolution structures could be obtained for all five oxidation states of HRP, showing such structures with preserved redox states for the first time.

In 1810, Planche[6] reported that a tincture of guaiacum developed a stronger colour when a piece of fresh horseradish root was soaked in it. During the next two centuries, horseradish peroxidase (HRP), the haem-containing enzyme responsible for the colour change in Planche's experiment, was used extensively to develop ideas on redox catalysis. Haem-containing redox enzymes participate in a strikingly diverse range of chemistry, yet all biological oxidation reactions catalysed by these enzymes involve very similar high-oxidation-state intermediates (Fig. 1) whose reactivity is modulated by the protein environment[4,5,7,8]. An understanding of these enzymes in structural terms requires detailed structures for the redox intermediates. This is not a trivial task because electrons liberated in the sample by X-rays during crystallographic data collection alter the redox state of the active site[1–5]. Redox enzymes have evolved to channel electrons efficiently into an oxidized active site. Attempts to minimize radiation-induced structural changes in crystals[1,2,4,9–11] include soaking them in excess electron scavengers[12] and using short-wavelength X-rays[4]. Neither of these measures proved to be sufficient to preserve the redox state of oxidized intermediates in HRP during conventional X-ray data collection. Figure 2a and b shows that the ferric and the compound III forms of HRP became reduced by X-rays at doses smaller than those required

§ Present addresses: Lawrence Livermore National Laboratory, PO Box 808, L-41, Livermore, California 94551, USA (H.S.); Carlsberg Laboratory, Gamle Carlsberg Vej 10, DK-2100 Valby, Denmark (A.H.).