

The genome sequence of the plant pathogen *Xylella fastidiosa*

The *Xylella fastidiosa* Consortium of the Organization for Nucleotide Sequencing and Analysis*, São Paulo, Brazil

* A full list of authors appears at the end of this paper

Xylella fastidiosa is a fastidious, xylem-limited bacterium that causes a range of economically important plant diseases. Here we report the complete genome sequence of *X. fastidiosa* clone 9a5c, which causes citrus variegated chlorosis—a serious disease of orange trees. The genome comprises a 52.7% GC-rich 2,679,305-base-pair (bp) circular chromosome and two plasmids of 51,158 bp and 1,285 bp. We can assign putative functions to 47% of the 2,904 predicted coding regions. Efficient metabolic functions are predicted, with sugars as the principal energy and carbon source, supporting existence in the nutrient-poor xylem sap. The mechanisms associated with pathogenicity and virulence involve toxins, antibiotics and ion sequestration systems, as well as bacterium–bacterium and bacterium–host interactions mediated by a range of proteins. Orthologues of some of these proteins have only been identified in animal and human pathogens; their presence in *X. fastidiosa* indicates that the molecular basis for bacterial pathogenicity is both conserved and independent of host. At least 83 genes are bacteriophage-derived and include virulence-associated genes from other bacteria, providing direct evidence of phage-mediated horizontal gene transfer.

Citrus variegated chlorosis (CVC), which was first recorded in Brazil in 1987, affects all commercial sweet orange varieties¹. Symptoms include conspicuous variegations on older leaves, with chlorotic areas on the upper side and corresponding light brown lesions, with gum-like material on the lower side. Affected fruits are small, hardened and of no commercial value. A strain of *Xylella fastidiosa* was first identified as the causal bacterium in 1993 (ref. 2) and found to be transmitted by sharpshooter leafhoppers in 1996 (ref. 3). CVC control is at present limited to removing infected shoots by pruning, the application of insecticides and the use of healthy plants for new orchards. In addition to CVC, other strains of *X. fastidiosa* cause a range of economically important plant diseases including Pierce's disease of grapevine, alfalfa dwarf, phony peach disease, periwinkle wilt and leaf scorch of plum, and are also associated with diseases in mulberry, pear, almond, elm, sycamore, oak, maple, pecan and coffee⁴. The triply cloned *X. fastidiosa* 9a5c, sequenced here, was derived from the pathogenic culture 8.1b obtained in 1992 in Bordeaux (France) from CVC-affected Valencia sweet orange twigs collected in Macaúbal (São Paulo, Brazil) on May 21, 1992 (ref. 2). Strain 9a5c produces typical CVC symptoms on inoculation into experimental citrus plants⁵, and into *Nicotiana tabacum* (S. A. Lopes, personal communication) and *Catharantus roseus* (P. Brant-Monteiro, personal communication)—two novel experimental hosts.

General features of the genome

The basic features of the genome are listed in Table 1 and a detailed map is shown in Fig. 1. The conserved origin of replication of the large chromosome has been identified in a region between the putative 50S ribosomal protein L34 and *gyrB* genes containing *dhnaA*, *dhnaN* and *recP*. The *Escherichia coli* DnaA box consensus sequence TTATCCACA is found on both DNA strands close to *dhnaA*. In addition, there are typical 13-nucleotide (ACCACACAC-CACCA) and 9-nucleotide (two TTTCATTGG and two TTTTATATT) sequences in other intergenic sequences of this region. This region is coincident with the calculated GC-skew signal inversion⁷. We have designated base 1 of the *X. fastidiosa* genome as the first T of the only TTTTAT sequence found between the ribosomal protein L34 gene and *dhnaA*.

The overall percentage of open reading frames (ORFs) for which a putative biological function could be assigned (47%) was slightly below that for other sequenced genomes such as *Thermotoga maritima*⁸ (54%), *Deinococcus radiodurans*⁹ (52.5%) and *Neisseria*

*meningitidis*¹⁰ (53.7%). This may reflect the lack of previous complete genome sequences from phytopathogenic bacteria. Plasmid pXF1.3 contains only two ORFs, one of which encodes a replication-associated protein. Plasmid pXF51 contains 64 ORFs, of which 5 encode proteins involved in replication or plasmid stability and 20 encode proteins potentially involved in conjugative transfer. One ORF encodes a protein similar to the virulence-associated protein D (VapD), found in many other bacterial pathogens¹¹. Four regions of pXF51 present significant DNA similarity to parts of transposons found in plasmids from other bacteria, suggesting interspecific horizontal exchange of genetic material.

The principal paralogous families are summarized in Table 2. The complete list of ORFs with assigned function is shown in Table 3. Seventy-five proteins present in the 21 completely sequenced genomes in the COG database¹² (as of 15th March 2000) were also found in *X. fastidiosa*. Each of these sequences was used to

Table 1 General features of the *Xylella fastidiosa* 9a5c genome

Main chromosome	2,679,305
Length (bp)	52.7%
G+C ratio	2.782
Open reading frames (ORFs)	88.0%
Coding region (% of chromosome size)	799
Average ORF length (bp)	1,283
ORFs with functional assignment	310
ORFs with matches to conserved hypothetical proteins	1,083
ORFs without significant data base match	2
Ribosomal RNA operons	2
rRNAs	(1SS/rRNA-Ala-TGC-rRNA-lle-GAT-rRNA-23S/rRNA-5S/rRNA)
tRNAs	49 (46 different sequences corresponding to all 20 amino acids)
tRNA ^{Met}	1
Plasmid pXF51	51,158
Length (bp)	49.6%
G+C ratio	64
Open reading frames (ORFs)	86.9%
Protein coding region (% of plasmid size)	30
ORFs with functional assignment	8
ORFs with matches to conserved hypothetical proteins	24
ORFs without significant data base match	
Plasmid pXF1.3	1,285
Length (bp)	55.6%
G+C ratio	2
Open reading frames (ORFs)	
ORFs with functional assignment	1



Table 2 Largest families of paralogous genes

Family (total number of families = 312)	Number of genes (total number of genes = 853)
ATP-binding subunits of ABC transporters	23
Reductases/dehydrogenases	12
Two-component system, regulatory proteins	12
Hypothetical proteins	10
Transcriptional regulators	9
Fimbrial proteins	9
Two-component system, sensor proteins	9

generate a phylogenetic tree of the 22 organisms. In 69% of such trees, *X. fastidiosus* was grouped with *Haemophilus influenzae* and *E. coli*, consistent with a phylogenetic analysis undertaken with the 16S rRNA gene¹³.

One ORF, a cytosine methyltransferase (XFI1774), is interrupted by a Group II intron. The intron was identified on the basis of the presence of a reverse transcriptase-like gene (as in other Group II introns), conserved splice sites, conserved sequence in structure V and conserved elements of secondary structure¹⁴. Group II introns are rare in prokaryotes, but have been found in different evolutionary lineages including *E. coli*, cyanobacteria and proteobacteria¹⁵.

Transcription, translation and repair

The basic transcriptional and translational machinery of *X. fastidiosus* is similar to that of *E. coli*¹⁶. Recombinational repair, nucleotide and base-excision repair, and transcription-coupled repair are present with some noteworthy features. For example, no photolyase was found, indicating exclusively dark repair. Although the main genes of the SOS pathway, *recA* and *lexA*, are present, ORFs corresponding to the three DNA polymerases induced by SOS in *E. coli* (DNA polymerases II, IV and V)¹⁷ are missing, indicating that the mutational pathway itself may be distinct.

Energy metabolism

Even though *X. fastidiosus* is, as its name suggests, a fastidious organism, energy production is apparently efficient. In addition to all the genes for the glycolytic pathway, all genes for the tricarboxylic acid cycle and oxidative and electron transport chains are present. ATP synthesis is driven by the resulting chemiosmotic proton gradient and occurs by an F-type ATP synthase. Fructose, mannose and glycerol can be utilized in addition to glucose in the glycolytic pathway. There is a complete pathway for hydrolysis of cellulose to glucose, consisting of 1,4- β -cellobiosidase, endo-1,4- β -glucanase and β -glucosidase, suggesting that cellulose breakdown may supplement the often low concentrations of monosaccharides in the xylem¹⁸. Two lipases are encoded in the genome, but there is no β -oxidation pathway for the hydrolysis of fatty acids, presumably precluding their utilization as an alternative carbon and energy source. Likewise, although enzymes required for the breakdown of threonine, serine, glycine, alanine, aspartate and glutamate are present, pathways for the catabolism of the other naturally occurring amino acids are incomplete or absent.

The gluconeogenesis pathway appears to be incomplete. Phosphoenolpyruvate carboxykinase and the gluconeogenic enzyme fructose-1,6-bisphosphatase, which are required to bypass the irreversible step in glycolysis, are not present. The absence of the first is compensated by the presence of phosphoenolpyruvate synthase and malate oxidoreductase, which together can generate phosphoenolpyruvate from malate. There appears, however, to be no known compensating pathway for the absence of fructose-1,6-bisphosphatase. It is possible that among the large number of unidentified *X. fastidiosus* genes there are non-homologous genes that compensate for steps in such critical pathways. Barring this possibility, however, the absence of a functional gluconeogenesis pathway implies a strict dependence on carbohydrates both as a

source of energy and anabolic precursors. The glyoxylate cycle is absent and the pentose phosphate pathway is incomplete. In the latter pathway, genes for neither 6-phosphogluconic dehydrogenase nor transaldolase were identified.

Small molecule metabolism

X. fastidiosus exhibits extensive biosynthetic capabilities, presumably an absolute requirement for a xylem-dwelling bacterium. Most of the genes found in *E. coli* necessary for the synthesis of all amino acids from chorismate, pyruvate, 3-phosphoglycerate, glutamate and oxaloacetic acid¹⁶ were identified. However, some genes in *X. fastidiosus* are bi-functional, such as phosphoribosyl-AMP cyclohydrolase/phosphoribosyl-ATP pyrophosphatase (XF2213), aspartokinase/homoserine dehydrogenase I (XF2225), imidazole-glycerolphosphate dehydratase/histidinol-phosphate phosphatase (XF2217) and a new diaminoipimate decarboxylase/aspartate kinase (XF1116) that would catalyse the first and the last steps of lysine biosynthesis. In addition, the gene for acetylglutamate kinase (XF1001) has an acetyltransferase domain at its carboxy-terminal end that would compensate for the missing acetyltransferase in the arginine biosynthesis pathway. Other missing genes include phosphoserine phosphatase, cystathionine β -lyase, homoserine O-succinyltransferase and 2,4,5-methyltetrahydrofolate-homocysteine methyltransferase. The first two enzymes are also absent in the *Bacillus subtilis* genome, the third is absent in *Haemophilus influenzae* and the fourth is missing in both genomes¹². We thus presume that alternative, unidentified enzymes complete the biosynthetic pathways in these organisms and in *X. fastidiosus*.

The pathways for the synthesis of purines, pyrimidines and nucleotides are all complete. *X. fastidiosus* is also apparently capable of both synthesizing and elongating fatty acids from acetate. Again, however, some *E. coli* enzymes were not found, such as holo acyl-carrier-protein synthase (also absent in *Synechocystis* sp., *H. influenzae* and *Mycoplasma genitalium*) and enoyl-ACP reductase (NADPH) (*FabI*) (also absent from *M. genitalium*, *Borrelia burgdorferi* and *Treponema pallidum*)¹².

X. fastidiosus appears to be capable of synthesizing an extensive variety of enzyme cofactors and prosthetic groups, including biotin, folic acid, pantothenate and coenzyme A, ubiquinone, glutathione, thioredoxin, glutaredoxin, riboflavin, FMN, FAD, pyrimidine nucleotides, porphyrin, thiamin, pyridoxal 5'-phosphate and lipolate. In a number of the synthetic pathways, one or more of the enzymes present in *E. coli* are absent, but this is also true for at least one other sequenced Gram-negative bacterial genome in each case¹². We therefore again infer that the missing enzymes are either not essential or replaced by unknown proteins with novel structures.

Transport-related proteins

A total of 140 genes encoding transport-related proteins were identified, representing 4.8% of all ORFs. For comparison, *E. coli*, *B. subtilis* and *M. genitalium* have around 10% of genes encoding transport proteins, whereas *Helicobacter pylori*, *Synechocystis* sp. and *Methanococcus jannaschii* have 3.5–5.4% (ref. 19). Transport systems are central components of the host-pathogen relationship (Fig. 2). There are a number of ion transporters and transporters for the uptake of carbohydrates, amino acids, peptides, nitrate/nitrite, sulphate, phosphate and vitamin B12. Many different transport

Figure 1 Linear representation of the main chromosome and plasmids pXF51 and pXF1.3 of the *Xyella fastidiosus* genome. Genes are coloured according to their biological role. Arrows indicate the direction of transcription. Genes with frameshift and point mutations are indicated with an X. Ribosomal RNA genes, the trnRNA, the principal repeats, prophages and the group II intron are indicated by coloured lines. Transfer RNAs are identified by a single letter identifying the amino acid. Pie chart represents the distribution of the number of genes according to biological role. The numbers below protein-producing genes correspond to gene IDs.

families are represented and include both small and large mechanosensitive conductance ion channels, a monovalent cation:proton antiporter (CAP-2) and a glycerol facilitator belonging to the major intrinsic protein (MIP) family. In addition, 23 ABC transport systems comprising 41 genes can be identified. *X. fastidiosa* appears to possess a phosphotransferase system (PTS) that typically mediates small carbohydrate uptake. There are both the enzyme I and HPr components of this system, as well as a gene supposedly involved in its regulation (*psfK* or *hprK*); however, there is no PTS permease—an essential component of the phosphotransferase complex. The functionality of the system therefore remains in question.

There are five outer membrane receptors, including siderophores, ferrichrome-iron and haemin receptors, which are all associated with iron transport. The energizing complexes, TomB–ExbB–ExbD and the paralogous TolA–TolR–TolQ, essential for the functioning of the outer membrane receptors, are also present. In all, 67 genes encode proteins involved in iron metabolism. We propose that in *X. fastidiosa* the uptake of iron and possibly of other transition metal ions such as manganese causes a reduction in essential micronutrients in the plant xylem, contributing to the typical symptoms of leaf variegation.

The *X. fastidiosa* genome encodes a battery of proteins that mediate drug inactivation and detoxification, alteration of potential drug targets, prevention of drug entry and active extrusion of drugs and toxins. These include ABC transporters and transport processes driven by a proton gradient. Of the latter, eight belong to the hydrophobe/amphiphile efflux-1 (HAEl) family, which act as multidrug resistance factors.

Adhesion

X. fastidiosa is characteristically observed embedded in an extracellular translucent matrix *in planta*²⁰. Clumps of bacteria form within the xylem vessels leading to their blockage and symptoms of the disease such as water-stress leaf curling. We deduce, from our analysis of the complete genome sequence, that the matrix is composed of extracellular polysaccharides (EPSs) synthesized by enzymes closely related to those of *Xanthomonas campestris* pv *campestris* (Xcc) that produce what is commercially known as xanthan gum. In comparison with Xcc, however, we did not find *gumI* (encoding glycosyltransferase V, which incorporates the terminal mannose), *gumL* (encoding ketalase which adds pyruvate to the polymer) or *gumG* (encoding acetyltransferase which adds acetate), suggesting that *Xylella* gum may be less viscous than its *Xanthomonas* counterpart.

Positive regulation of the synthesis of extracellular enzymes and EPS in *Xanthomonas* is effected by proteins coded by the *rpf* (regulation of pathogenicity factors) gene cluster²¹. Mutations in any of these genes in *Xanthomonas* results in failure to synthesize the EPS. In consequence, the strain becomes non-pathogenic²¹. *X. fastidiosa* contains genes that encode RpfA, RpfB, RpfC and RpfF, suggesting that both bacteria may regulate the synthesis of pathogenic EPS factors through similar mechanisms.

Fimbria-like structures are readily apparent upon electron microscopical observation of *X. fastidiosa* within both its plant and insect hosts²². Because of the high velocity of xylem sap passing through narrow portions of the insect foregut, fimbria-mediated attachment may be essential for insect colonization. Indeed, in the insect mouthparts the bacteria are attached in ordered arrays, indicating specific and polarized adhesion²³. In addition, fimbriae are thought to be involved in both plant–bacterium and bacterium–bacterium interactions during colonization of the xylem itself. We identified 26 genes encoding proteins responsible for the biogenesis and function of Type 4 fimbriae filaments. This type of fimbria is found at the poles of a wide range of bacterial pathogens where they act to mediate adhesion and translocation along epithelial surfaces²⁴. The genes include *pilS* and *pilR* homologues, which encode a two-component

system controlling transcription of fimbrial subunits, presumably in response to host cues, and *pilG*, *H*, *I*, *J* and *clpA*, which encode a chemotactic system transducing environmental signals to the pilus machinery.

In addition to the EPS and fimbriae, which are likely to have central roles in the clumping of bacteria and in adhesion to the xylem walls, we also identified outer membrane protein homologues for afimbrial adhesins. Although fimbrial adhesins are well characterized as crucial virulence factors in both plant and human pathogens²⁵, afimbrial adhesins, which are directly associated with the bacterial cell surface, have been hitherto associated only with human and animal pathogens, where they promote adherence to epithelial tissue. Of the three putative adhesins of this kind identified in *X. fastidiosa*, two exhibit significant similarity to each other (XF1981, XF1529) and to the *hsf* and *hna* gene products of *H. influenzae*²⁶. The third (XF1516) is similar to the *uspA1* gene product of *Moraxella catarrhalis*²⁷. All these proteins share the common C-terminal domain of the autotransporter family²⁸. Direct experimentation will be required to establish whether these adhesins promote binding to plant cell structures or components of the insect vector foregut, or both. Nevertheless, their presence in the *X. fastidiosa* genome adds to the increasing evidence for the generality of mechanisms of bacterial pathogenicity, irrespective of the host organism²⁹.

We also identified three different haemagglutinin-like genes. Again, similar genes have not previously been identified in plant pathogens. These genes (XF2775, XF2196, XF0889) are the largest in the genome and exhibit highest similarity to a *Neisseria meningitidis* putative secreted protein¹⁰.

Intervessel migration

Movement between individual xylem vessels is crucial for effective colonization by *X. fastidiosa*. For this to occur, degradation of the pit membrane of the xylem vessel is required. Of the known pectolytic enzymes capable of this function, a polygalacturonase precursor and a cellulase were identified, although the former contains an authentic frameshift. These genes exhibited highest similarity to orthologues in *Ralstonia solanacearum*—which causes wilt disease in tomatoes—where the polygalacturonase genes are required for wild-type virulence.

Toxicity

We identified five haemolysin-like genes: haemolysin III (XF0175), which belongs to an uncharacterized protein family, and four others (XF0668, XF1011, XF2407, XF2759) which belong to the RTX toxin family that contains tandemly repeated glycine-rich nonapeptide motifs at the C-terminal domain. One of these ORFs is closely related to bacteriocin, an RTX toxin also found in the plant bacterium *Rhizobium leguminosarum*³⁰. RTX or RTX-like proteins are important virulence factors widely distributed among Gram-negative pathogenic bacteria³¹.

There are two Colicin-V-like precursor proteins. Colicin V is an antibacterial polypeptide toxin produced by *E. coli*, which acts against closely related sensitive bacteria³². The precursors consist of 102-amino-acid peptides (XF0262, XF0263) that have the typical conserved leader 15-amino-acid motif, and have some similarity with Colicin V from *E. coli* at the remaining C-terminal portion. The necessary apparatus for Colicin biosynthesis and secretion is also present. Interestingly, in *E. coli* most of the genes necessary for biogenesis and export of Colicin V are in a gene cluster present in a plasmid, whereas in *X. fastidiosa* these genes are dispersed in the chromosome.

We found four genes that may function in polyketide biogenesis: polyketide synthase (PKS), peridine-dependent deoxygenase, daunorubicin C-13 ketoreductase and a NonF-related protein. These genes belong to the synthesis pathways of frenolicin, rapamycin, daunorubicin and nonactin, respectively. These pathways



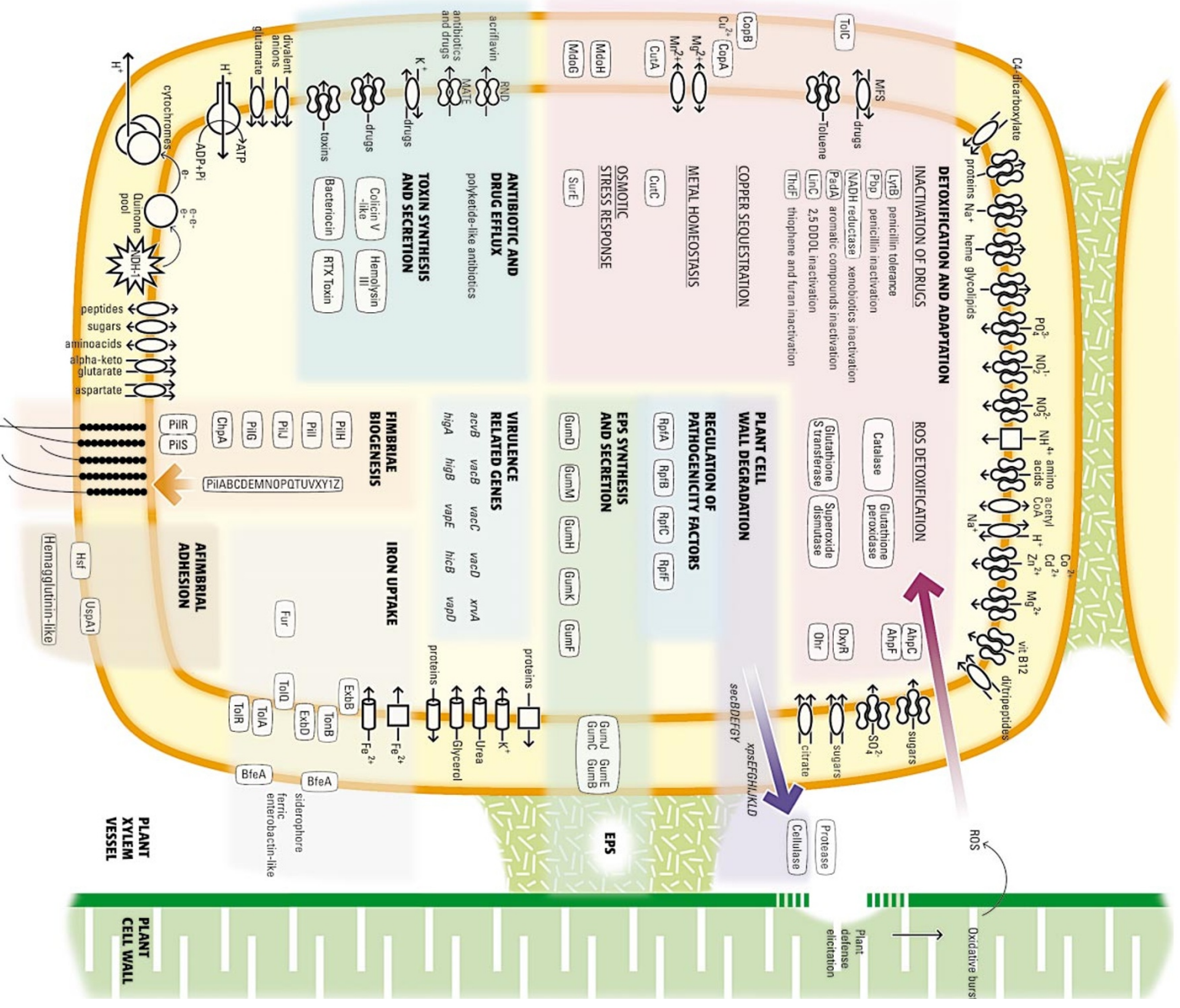


Figure 2 A comprehensive view of the biochemical processes involved in *Xylella fastidiosa* pathogenicity and survival in the host xylem. The principal functional categories are shown in bold, and the bacterial genes and gene products related to that function are arranged within the coloured section containing the bold heading. Transporters are indicated as follows: cylinders, channels; ovals, secondary carriers, including the MFS family; paired dumbbells, secondary carriers for drug extrusion; triple dumbbells, ABC transporters; bulb-like icon, F-type ATP synthase; squares, other transporters. Icons with two arrows

represent symporters and antiporters (H^+ or Na^+ porters, unless noted otherwise). 2,5-DDDL, 2,5-dichloro-2,5-cyclohexadiene-1,4-diol; EPS, exopolysaccharides; MATE, multi-antimicrobial extrusion family of transporters multidrug efflux gene (KF2686); MFS, major facilitator superfamily of transporters; Pcbp, β -lactamase-like penicillin-binding protein (XF1621); RND, resistance-nodulation-cell division superfamily of transporters; ROS, reactive oxygen species.

include many more enzymes, which we did not find; however, some of the genes listed lie close to ORFs without significant database matches, suggesting that at least one (as yet undiscovered) polypeptide pathway may be functional.

Prophages

Bacteriophages can mediate the evolution and transfer of virulence factors and occasional acquisition of new traits by the bacterial host. Because as much as 7% of the *X. fastidiosa* genome sequenced corresponds to double-stranded (ds) DNA phage sequences, mostly from the Lambda group, we suspect that this route may have been of particular importance for this bacterium. It is noteworthy that a very high percentage of phage-related sequences has also been detected in a second vascular-restricted plant pathogen, *Spiroplasma citri*³³. We identified four regions, with a high density of ORFs homologous to phage sequences, that we considered to be prophages, in addition to isolated phage sequences dispersed throughout the genome. Two of these prophages (each ~42 kbp, designated XFP1 and XFP2) are similar to each other, lie in opposite orientations in distinct regions and appear to belong to the dsDNA, tailed-phage group. Both appear to contain most of the genes responsible for particle assembly, although we know of no reports of phage particle release from *X. fastidiosa* cultures. In prophage XFP1, we found two ORFs from the virulence tail genes V and W that are similar to ORF118 and *wspA* from the virulence-associated region of the animal pathogen *Dichelobacter nodosus*, which by homology encode a killer and a suppressor protein³⁴. Interestingly, in prophage XFP2, we found two other ORFs also between tail genes V and W that are similar to hypothetical ORFs of *Kalstonia eutropha* transposon Tn4371 (ref. 35). The other two identified prophages, XFP3 and XFP4, are also similar in sequence to each other and to the *H. influenzae* cryptic prophage phiU (ref. 36). They both contain a 14,317-bp exact repeat. Few particle-assembly genes were found in these regions, suggesting that these prophages are defective. An ORF similar to *hiccB* from *H. influenzae*, a component of the major pilus gene cluster in some isolates, was found in XFP4 (ref. 37).

The presence of virulence-associated genes from other organisms within the prophage sequences is strong evidence for a direct role for bacteriophage-mediated horizontal gene transfer in the definition of the bacterial phenotype.

Absence of avirulence genes

Phytopathogenic bacteria generally have a limited host range, often confined to members of a single species or genus. This specificity is defined by the products of the so-called avirulence (*avr*) genes present in the pathogen, which are injected directly into host cells, on infection, through a type III secretory system^{38–40}. BLAST⁴¹ searches with all known *avr* and type III secretory system sequences failed to identify genes encoding proteins with significant similarities in the genome of *X. fastidiosa*. Although the variability of *avr* genes amongst bacteria might account for this apparent lack, the high level of similarity of some components of the type III secretory system argues against this. We suspect that these genes are, in fact, not required because of the insect-mediated transmission and vascular restriction of the bacterium that obviates the necessity of host cell infection. Furthermore, if the differing host ranges of *X. fastidiosa* are molecularly defined, this may be by a quite different mechanism not involving *avr* proteins.

Conclusions

Before the elucidation of its complete genome sequence, very little was known of the molecular mechanisms of *X. fastidiosa* pathogenicity. Indeed, this bacterium was probably the least characterized of all organisms that have been fully sequenced. Our complete genetic analysis has determined not only the basic metabolic and replicative characteristics of the bacterium, but also a number of potential

pathogenicity mechanisms. Some of these have not previously been postulated to occur in phytopathogens, providing new insights into the generality of these processes. Indeed, the availability of this first complete plant pathogen genome sequence will now allow the initiation of the detailed comparison of animal and plant pathogens at the whole-genome level. In addition, the information contained in the sequence should provide the basis for an accelerated and rational experimental dissection of the interactions between *X. fastidiosa* and its hosts that might lead to fresh insights into potential approaches to the control of CYC. □

Methods

The sequencing and analysis in this project were carried out by a network of 34 biology laboratories and one bioinformatics centre. The network is called the Organization for Nucleotide Sequencing and Analysis (ONSA)⁴², and is entirely located in the state of São Paulo, Brazil.

Sequencing and assembly

The sequence was generated using a combination of ordered cosmid and shotgun strategies⁴³. A cosmid library was constructed, providing roughly 15-fold genome coverage, containing 1,056 clones with average insert size of 40 kilobases (kb). High-density colony filters of the library were made, and a physical map of the genome was constructed using a strategy of hybridization without replacement⁴⁴. A total of 113 cosmid clones was selected for sequencing on the basis of the hybridization map and end-sequence analysis. The cosmid sequences were assembled into 15 contigs covering 90% of the genome. Additionally, shotgun libraries with different insert sizes (0.8–2.0 kb and 2.0–4.5 kb) were constructed from sheared or restricted genomic DNA cloned into plasmids, and sequenced to achieve a 3.74-fold coverage of high-quality sequence (29,140 reads). Most of the sequencing was performed with BigDye terminators on ABI Prism 377 DNA sequencers.

Cosmid and shotgun sequences were assembled into six contigs. We identified sequence gaps by linking information from forward and reverse reads, and closed either by primer walking or insert subcloning. The remaining physical gaps were closed by combinatorial PCR and by lambda clones selected from a λ Dash library by end-sequencing. The collinearity between the genome and the obtained sequence was confirmed by digestion of genomic DNA with *Ascl*, *NciI*, *SfiI*, *SmaI* and *SrfI*, followed by comparison of the digestion pattern with the electronic digestion of the generated sequence. In addition, sequences from both ends of most cosmid clones and 236 λ clones were used to confirm the orientation and integrity of the contigs. The sequence was assembled using phred+phrap+consed⁴⁵. All consensus bases have quality with Phred value of at least 20. There are no unexplained high quality discrepancies, each consensus base is confirmed by at least one read from each strand, and the overall error estimate is less than 1 in every 10,000 bases.

ORF prediction and annotation

ORFs were determined using glimmer 2.0 (ref. 46) and the glimmer post-processor RBSfinder (S. L. Salzberg, personal communication). A few ORFs were found by hand guided by BLAST⁴¹ results. Annotation was carried out in a cooperative way, mostly by comparison with sequences in public databases, using BLAST⁴¹ and tRNAscan-SE (ref. 47) and was based on the functional categories for *E. coli*⁴⁸. Only one tmRNA was located (K. Williams, personal communication). To help annotate transport proteins, we built a custom BLAST⁴¹ database using sequences from <http://www-biology.ucsd.edu/~msaiter/transport/toc.html> and compared our ORFs with these sequences. Phylogenetic trees for conserved COG⁴² were built using ClustalX⁴⁹ for multiple alignment and Phylip⁵⁰. Paralogous gene families (Table 2) were determined using BLASTX with the E-value cut-off equal to 4×5 and such that at least 60% of the query sequence and at least 30% of the subject sequence were aligned.

Received 24 March; accepted 24 May 2000.

- Rossari, V. *et al.* Présence de bactéries dans le xylème d'orangers atteints de chlorose variée, une nouvelle maladie des agrumes au Brésil. *C. R. Acad. Sci. Paris série III*, **310**, 345–349 (1990).
- Chang, C. J. *et al.* Culture and serological detection of the xylem-limited bacterium causing citrus variegated chlorosis and its identification as a strain of *Xylella fastidiosa*. *Curr. Microbiol.* **27**, 137–142 (1993).
- Roberto, S. R., Coutinho, A., De Lima, J. E. O., Miranda, V. S. & Carlos, E. F. Transmissão de *Xylella fastidiosa* pelas cigarrinhas *Diholopertus castilliani*, *Acrargonia terminalis* e *Oncotropia fasciis* em citros. *Fitopatol. Bras.* **21**, 517–518 (1996).
- Purcell, A. H. & Hopkins, D. L. Fastidious xylem-limited bacterial plant pathogens. *Annu. Rev. Phytopathol.* **34**, 131–151 (1996).
- Li, W. B. *et al.* A triply cloned strain of *Xylella fastidiosa* multiplies and induces symptoms of citrus variegated chlorosis in sweet orange. *Curr. Microbiol.* **39**, 106–108 (1999).
- Ye, F., Renaudin, J., Bové, J. M. & Laigret, F. Cloning and sequencing of the replication origin (oriC) of the *Spiroplasma citri* chromosome and construction of autonomously replicating artificial plasmids. *Curr. Microbiol.* **29**, 23–29 (1994).
- Francino, M. P. & Ochman, H. Strand asymmetries in DNA evolution. *Trends Genet.* **13**, 240–245 (1997).
- Nelson, K. E. *et al.* Evidence for lateral gene transfer between Archaea and bacteria from genome sequence of *Thermoplasma murinum*. *Nature* **399**, 323–329 (1999).



9. White, O. *et al.* Genome sequence of the radiotolerant bacterium *Deinococcus radiodurans* RI. *Science* **286**, 1571–1577 (1999).

10. Tetelin, H. *et al.* Complete genome sequence of *Nesteria meningitidis* serogroup B strain MC58. *Science* **287**, 1809–1815 (2000).

11. Katz, M. E., Strugnell, R. A. & Rood, J. I. Molecular characterization of a genomic region associated with virulence in *Dichelobacter nodosus*. *Infect. Immun.* **60**, 4586–4592 (1992).

12. Tatusov, R. L., Galperin, M. Y., Natale, D. A. & Koonin, E. V. The COG database: a tool for genome scale analysis of protein functions and evolution. *Nucleic Acids Res.* **28**, 33–36 (2000).

13. Preston, G. M., Hamblin, B. & Rainey, P. B. Bacterial genomics and adaptation to life on plants: implications for the evolution of pathogenicity and symbiosis. *Curr. Opin. Microbiol.* **1**, 589–597 (1998).

14. Knopp, V., Koskela, S. & Brennickle, A. On the identification of group II introns in nucleotide sequence data. *J. Mol. Biol.* **242**, 389–396 (1994).

15. Fean, J. L. & Michal, F. Group II self-splicing introns in bacteria. *Nature* **364**, 356–361 (1993).

16. Blatner, F. R. *et al.* The complete genome sequence of *Escherichia coli* K-12. *Science* **277**, 1453–1474 (1997).

17. Bridges, B. A. DNA repair: Polymerases for passing lesions. *Curr. Biol.* **9**, R475–R477 (1999).

18. Brodbeck, B. V., Anderson, P. C. & Mizell, R. F. Effects of total dietary nitrogen and nitrogen form on the development of xylophagous lepidoptera. *Arch. Insect Biochem. Physiol.* **42**, 37–50 (1999).

19. Paulsen, I. T., Sliwinski, M. K. & Sater, M. H. J. Microbial genome analyses: global comparisons of transport capabilities based on phylogenies, bioenergetics and substrate specificities. *J. Mol. Biol.* **277**, 573–592 (1996).

20. Chagas, C. M., Rossetti, V. & Beretta, M. J. G. Electron-microscopy studies of a xylem-limited bacterium in sweet orange affected with citrus variegated chlorosis disease in Brazil. *J. Phytopathol.* **134**, 306–312 (1992).

21. Tang, J. L. *et al.* Genetic and molecular analysis of a cluster of *ryf* genes involved in positive regulation of synthesis of extracellular enzymes and polysaccharide in *Xanthomonas campestris* pathovar *caespertis*. *Mol. Gen. Genet.* **226**, 409–417 (1991).

22. Rajin, C. B. & Wells, J. M. Diseases caused by fastidious xylem-limited bacteria. *Plant Disease* **70**, 182–186 (1986).

23. Branstetter, R. H., Timmer, I. W., French, W. J. & McCoy, R. E. Colonization of the sharpshooter vectors, *Oncometopia nigricans* and *Hemulidictya conglutinata*, by xylem-limited bacteria. *Phytopathology* **73**, 530–535 (1983).

24. Fernandez, L. A. & Bergenger, J. Secretion and assembly of regular surface structures in Gram-negative bacteria. *FEBS Microbiol. Rev.* **24**, 21–44 (2000).

25. Soto, G. E. & Hudegger, S. J. Bacterial adhesins: common themes and variations in architecture and assembly. *J. Bacteriol.* **181**, 1059–1071 (1999).

26. Gene, J. W. Molecular determinants of the interaction between *Haemophilus influenzae* and human cells. *Am. J. Respir. Crit. Care Med.* **154**, S192–S196 (1996).

27. Cope, L. D. *et al.* Characterization of the *Monaxella calanthalis uspiA1* and *uspiA2* genes and their encoded products. *J. Bacteriol.* **181**, 4026–4034 (1999).

28. Henderson, I. R., Navarro-Garcia, F. & Nataro, J. P. The great escape: structure and function of the autotransporter proteins. *Trends Microbiol.* **6**, 370–378 (1998).

29. Rahme, L. G. *et al.* Common virulence factors for bacterial pathogenicity in plants and animals. *Science* **268**, 1899–1902 (1995).

30. Oresank, I. J., Twelker, S. & Hyres, M. E. Cloning and characterization of a *Rhizobium leguminosarum* gene encoding a bacteriotoxin with similarities to RTX toxins. *Appl. Environ. Microbiol.* **65**, 2833–2840 (1999).

31. Lally, E. T., Hill, R. B., Kleba, I. R. & Korostoff, J. The interaction between RTX toxins and target cells. *Trends Microbiol.* **7**, 356–361 (1999).

32. Havarstein, L. S., Hollo, H. & Nes, I. E. The leader peptide of colicin V shares consensus sequences with leader peptides that are common among peptide bacteriocins produced by gram-positive bacteria. *Microbiology* **140**, 2383–2389 (1994).

33. Ye, F. *et al.* A physical and genetic map of the *Spiriplasma citri* genome. *Nucleic Acids Res.* **20**, 1559–1565 (1992).

34. Billington, S. J., Johnston, J. L. & Rood, J. I. Virulence regions and virulence factors of the ovine footrot

pathogen, *Dichelobacter nodosus*. *FEBS Microbiol. Lett.* **145**, 147–156 (1996).

35. Meriin, C., Spingard, D. & Toussaint, A. Trn371: A modular structure encoding a phage-like integrase, a pseudomonas-like catabolic pathway, and RPr/Trf-like transfer functions. *Plasmid* **41**, 40–54 (1999).

36. Hendrix, R. W., Smith, M. C., Burns, R. N., Ford, M. E. & Hatfull, G. E. Evolutionary relationships among diverse bacteriophages and prophages: All the world's a phage. *Proc. Natl Acad. Sci. USA* **96**, 2192–2197 (1999).

37. Mhangara-Mhangadura, T., Morlin, G., Smith, A. L., Eisenstark, A. & Golomb, M. Evolution of the major pilus gene cluster of *Haemophilus influenzae*. *J. Bacteriol.* **180**, 4693–4703 (1998).

38. Altano, J. R. & Collmer, A. The type III (Hrp) secretion pathway of plant pathogenic bacteria: trafficking harpins, Avr proteins, and death. *J. Bacteriol.* **179**, 5655–5662 (1997).

39. Galan, J. E. & Collmer, A. Type III secretion machines: bacterial devices for protein delivery into host cells. *Science* **284**, 1322–1328 (1999).

40. Young, G. M., Schmidt, D. H. & Miller, V. L. A new pathway for the secretion of virulence factors by bacteria: the flagellar export apparatus functions as a protein-secretion system. *Proc. Natl Acad. Sci. USA* **96**, 6456–6461 (1999).

41. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).

42. Simpson, A. J. & Perez, J. F. ONSA, the São Paulo Viral Genomics Institute. *Nature Biotechnol.* **16**, 795–796 (1998).

43. Fleischmann, R. D. *et al.* Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* **269**, 496–512 (1995).

44. Hohsied, J. D. *et al.* High resolution cosmid and P1 maps spanning the 14 Mb genome of the fission yeast *S. pombe*. *Cell* **73**, 109–120 (1993).

45. Gordon, D., Abajian, C. & Green, P. Consed: a graphical tool for sequence finishing. *Genome Res.* **8**, 195–202 (1998).

46. Dacler, A. L., Harmon, D., Kasif, S., White, O. & Szaberg, S. L. Improved microbial gene identification with GLIMMER. *Nucleic Acids Res.* **27**, 4636–4641 (1999).

47. Lowe, T. M. & Eddy, S. R. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequences. *Nucleic Acids Res.* **25**, 975–984 (1997).

48. Riley, M. Functions of the gene products of *Escherichia coli*. *Microbiol. Rev.* **57**, 862–952 (1993).

49. Thompson, J. D., Gibson, T. J., Plewniak, E., Jeanmougin, F. & Higgins, D. G. The CLUSTAL windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* **25**, 4876–4882 (1997).

50. Felsenstein, J. PHYLIP—Phylogeny Inference Package (Version 3.2). *Cladistics* **5**, 164–166 (1989).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>), or on the author's World-Wide Web site (<http://www.lbi.ic.unicamp.br/xfi>) or as paper copy from the London editorial office of Nature.

Acknowledgements

The consortium is indebted to J. F. Perez, Scientific Director of Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), for his strategic vision in creating and nurturing this project as well as C. A. de Pan and Jucara Parra for their administrative coordination. We thank our Steering Committee: S. Oliver, A. Coffeau, J. Sgourou, A. C. M. Paiva and J. L. Azevedo for their critical accompaniment of the work. We also thank R. Fulton and P. Minx for their timely contribution and advice. Project funding was from FAPESP, the RHAÉ programme of the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and Fundectrus. For the full list of individuals who contributed to the completion of this project see (<http://www.lbi.ic.unicamp.br/xfi>).

Correspondence and requests for materials should be addressed to J. C.S. (e-mail: sethbal@ic.unicamp.br). The sequence has been deposited in GenBank with accession numbers AF003849 (chromosome), AF003850 (pXF1_3) and AF003851 (pXF51).

Authors: A. J. G. Simpson¹, F. C. Reinach², P. Arruda³, F. A. Abreu⁴, M. Agencio⁵, R. Alvarenga², L. M. C. Alves⁶, J. E. Araya⁷, G. S. Bata², G. S. Baptista⁸, M. H. Barros⁸, E. D. Bonaccorsi², S. Bordin⁹, J. M. Bove¹⁰, M. R. S. Briones¹, M. R. P. Bueno¹¹, A. A. Camargo¹, L. E. A. Camargo¹², D. M. Carraro¹², H. Carrer¹², M. B. Colauto¹³, C. Colombo¹⁴, F. F. Costa³, M. C. R. Costa¹⁵, G. M. Costa-Neto¹⁶, L. L. Coutinho¹², M. Cristofani¹⁷, E. Dias-Neto¹, G. Docena², H. El-Dorri², A. P. Facincani¹, A. J. S. Ferreira², V. C. A. Ferreira¹⁸, J. A. Ferro⁶, J. S. Fraga⁴, S. C. Franca¹⁹, M. G. Franco²⁰, M. Frohme²¹, L. R. Furian²², M. Garnier¹⁰, G. H. Goldman²³, M. H. S. Goldman²⁴, S. L. Gomes², A. Gruber⁴, P. L. Ho²⁵, J. D. Høihsel²¹, M. L. Junqueira²⁶, E. L. Kemper³, J. P. Kitajima²⁷, J. E. Krieger²⁶, E. E. Kuramae²⁸, F. Laigret¹⁰, M. R. Lambais¹², L. C. G. Leite²⁵, E. G. M. Lenos⁶, M. V. F. Lemos⁸, S. A. Lopes¹⁹, C. R. Lopes¹³, J. A. Machado³⁰, M. A. M. B. N. M. Madeira⁴, H. M. F. Madeira¹²⁺, G. L. Marinho¹³, M. V. F. Marques⁸, E. A. L. T. O. Nascimento²⁵, L. E. S. Netto¹¹, A. M. Menck⁸, E. G. Miracca⁵, C. Y. Miyaki¹¹, G. B. Montelero-Vitorello¹², D. H. Moon²⁰, M. A. Nagai¹, A. L. T. O. Nascimento²⁵, L. E. S. Netto¹¹, A. Nhani Jr⁶, F. G. Nobrega⁸, L. R. Nunes³¹, M. A. Oliveira³², M. C. de Oliveira³³, R. G. de Oliveira³¹, D. A. Palmieri¹³, A. Paris¹³, B. R. Peixoto², G. A. G. Pereira³², H. A. Pereira Jr⁶, J. B. Pesquero¹⁶, R. B. Quaggio², P. G. Roberto¹⁹, V. Rodrigues³⁴, A. J. de M. Rosa³⁴, V. E. de Rosa Jr²⁸, R. G. de Sa³⁴, R. V. Santelli², H. E. Sawasaki¹⁴, A. C. R. da Silva², A. M. da Silva², F. R. da Silva^{3,27}, W. A. Silva Jr¹⁵, J. F. da Silveira⁷, M. L. Z. Silvestri², W. J. Siqueira¹⁴, A. A. de Souza¹⁷, A. P. de Souza³, M. F. Terenzi²³, D. Truffi¹², S. M. Tsai²⁰, M. H. Tshunako¹⁸, H. Vallada³⁵, M. A. Van Sluys³³, S. Verjovski-Almeida², A. L. Vettore³, M. A. Zago¹⁵, M. Zatz¹¹, J. Meldanis²⁷ & J. C. Setubal¹²⁷.

Addresses: 1. Instituto Ludwig de Pesquisa sobre o Câncer, Rua Prof. Antonio Prudente, 109 - 4º andar, 01509-010, São Paulo - SP, Brazil; 2. Departamento de Bioquímica, Instituto de Química, Universidade de São Paulo, Av. Prof. Lineu Prestes, 748, 05508-900, São Paulo - SP, Brazil; 3. Centro de Biologia Molecular e Engenharia Genética, Universidade Estadual de Campinas, Caixa Postal 6010, 13083-970, Campinas - SP, Brazil; 4. Laboratório de Biologia Molecular, Departamento de Patologia, Faculdade de Medicina Veterinária e Zootecnia, Universidade de São Paulo, Av. Prof. Dr. Orlando Marques de Paiva, 87, 05508-000, São Paulo - SP, Brazil; 5. Disciplina de Oncologia, Departamento de Radiologia, Faculdade de Medicina, Universidade de São Paulo, Av. Dr. Arnaldo, 455, 01296-903, São Paulo - SP, Brazil; 6. Departamento de Tecnologia, Faculdade de Ciências Agrárias e Veterinárias de Jaboticabal, Universidade Estadual Paulista, Via



de Acesso Prof. Paulo D. Castellane s/n, Km 5, 14884-900, Jaboticabal – SP, Brazil; 7, Departamento de Microbiologia, Imunologia e Parasitologia, Escola Paulista de Medicina, Universidade Federal de São Paulo, Rua Botucatu, 862, 04023-062, São Paulo – SP, Brazil; 8, Departamento de Microbiologia, Instituto de Ciências Biomédicas, Universidade de São Paulo, Av. Prof. Lineu Prestes, 1374, 05508-900, São Paulo – SP, Brazil; 9, Hemocentro, Faculdade de Ciências Médicas, Universidade Estadual de Campinas, 13083-97, Campinas – SP, Brazil; 10, Institut National de la Recherche Agronomique et Université Victor Ségalen Bordeaux 2, 71 Avenue Edouard Bourleaux, Laboratoire de Biologie Cellulaire et Moléculaire, 33883 Villenave d'Ornon Cedex, France; 11, Département de Biologie, Institut de Biochimie, Université de São Paulo, Rua do Matão, 277, 05508-900, São Paulo – SP, Brazil; 12, Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo, Av. Pádua Dias, 11, 13418-900, Piracicaba – SP, Brazil; 13, Departamento de Genética, Instituto de Biociências, Universidade Estadual Paulista, Distrito de Rubião Junior, 18618-000, Botucatu – SP, Brazil; 14, Centro de Genética, Biologia Molecular e Fitoquímica, Instituto Agronômico de Campinas, Av. Barão de Itapuruá, 1481, Caixa Postal 28, 13001-970, Campinas – SP, Brazil; 15, Departamento de Clínica Médica, Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, Av. Bandeirantes, 3900, 14049-900, Ribeirão Preto – SP, Brazil; 16, Departamento de Biofísica, Escola Paulista de Medicina, Universidade Federal de São Paulo, Rua Botucatu, 862, 04023-062, São Paulo – SP, Brazil; 17, Centro de Citricultura Sylvio Moreira, Instituto Agronômico de Campinas, Caixa Postal 04, 13490-970, Cordeirópolis – SP, Brazil; 18, Laboratório de Bioquímica Fitopatológica e Laboratório de Imunologia, Instituto Biológico, Av. Cons. Rodrigues Alves, 1252, 04014-002, São Paulo – SP, Brazil; 19, Departamento de Biotecnologia de Plantas Medicinais, Universidade de Ribeirão Preto, Av. Costáble Romano, 2201, 14096-380, Ribeirão Preto – SP, Brazil; 20, Centro de Energia Nuclear na Agricultura, Universidade de São Paulo, Av. Centenário, 303, Caixa Postal 96, 13400-970, Piracicaba – SP, Brazil; 21, Funktionelle Genomanalyse, Deutsches Krebsforschungszentrum, Im Neuenheimer Feld 506, D-69120, Heidelberg, Germany; 22, 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; 23, Departamento de Ciências Farmacéuticas, Faculdade de Ciências Farmacéuticas de Ribeirão Preto, Universidade de São Paulo, Av. do Café s/n, 14040-903, Ribeirão Preto – SP, Brazil; 24, Departamento de Biologia, Faculdade de Filosofia, Ciências e Letras de Biotecnologia, Instituto Butantan, Av. Vital Brasil, 1500, 05503-900, São Paulo – SP, Brazil; 25, Centro de Biologia Aplicada a Agropecuária, Faculdade de Ciências Agrárias e Veterinárias de Jaboticabal, 14884-900, Jaboticabal – SP, Brazil; 26, Laboratório de Genética e Cardiologia Molecular/LJM 13, Instituto do Coração (InCor), Faculdade de Medicina, Universidade de São Paulo, Av. Dr. Enes de Carvalho Aguiar, 44, 05403-000, São Paulo – SP, Brazil; 27, Instituto de Computação, Universidade Estadual de Campinas, Caixa Postal 6176, 13083-970, Campinas – SP, Brazil; 28, Departamento de Produção Vegetal, Faculdade de Ciências Agronômicas, Universidade Estadual Paulista, Fazenda Lageado, Caixa Postal 237, 18603-970, Botucatu – SP, Brazil; 29, Departamento de Biologia Aplicada a Agropecuária, Faculdade de Ciências Agrárias e Veterinárias de Jaboticabal, 14884-900, Jaboticabal – SP, Brazil; 30, Hospital do Câncer – A.C. Camargo, R. Antonio Prudente, 211, 01509-010, São Paulo – SP, Brazil; 31, Núcleo Integrado de Biotecnologia, Universidade de Mogi das Cruzes, Av. Dr. Cândido Xavier de Almeida Souza, 200, 08780-911, Mogi das Cruzes – SP, Brazil; 32, Departamento de Genética e Evolução, Instituto de Biologia, Universidade Estadual de Campinas, Caixa Postal 6010, 13083-970, Campinas – SP, Brazil; 33, Departamento de Botânica, Instituto de Biociências, Universidade de São Paulo, Rua do Matão, 277, 05508-900, São Paulo – SP, Brazil; 34, Departamento de Parasitologia, Microbiologia e Imunologia, Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, Av. Bandeirantes, 3900, 14049-900, Ribeirão Preto – SP, Brazil; 35, Departamento de Psiquiatria, Instituto de Psiquiatria, Faculdade de Medicina, Universidade de São Paulo, Rua Dr. Ovídeo Pires de Campos s/n, Sala 4051 (3º. andar), 05403-010, São Paulo – SP, Brazil.

† Present addresses: Norartis Seeds LTDA, Av. Prof. Vicente Rios, 90, 04706-900, São Paulo – SP, Brazil (J. A. Maclhado); Centro de Ciências Agrárias e Ambientais, Pontifícia Universidade Católica do Paraná, BR-376, Km 14, Caixa Postal 129, 83010-500, São José dos Pinhais – PR, Brazil (H. M. F. Maddeira); Instituto de Pesquisa e Desenvolvimento, Universidade do Vale do Paraíba, Av. Shishimi Hifumi, 2911, 12244-000, São José dos Campos – SP, Brazil (F. G. Nobrega).





Table 3 Putative identification of Xylem/Endothelial genes. Gene numbers correspond to those in Fig. 1. Percentages represent % identities relative to the most relevant database hit (conserved hypotheticals and % identities of genes with known/shaded codons in frame not shown).

Gene	Accession	Identity	Gene	Accession	Identity	Gene	Accession	Identity	Gene	Accession	Identity	Gene	Accession	Identity
10000	U00000	100%	10000	U00000	100%	10000	U00000	100%	10000	U00000	100%	10000	U00000	100%
10001	U00001	95%	10001	U00001	95%	10001	U00001	95%	10001	U00001	95%	10001	U00001	95%
10002	U00002	90%	10002	U00002	90%	10002	U00002	90%	10002	U00002	90%	10002	U00002	90%
10003	U00003	85%	10003	U00003	85%	10003	U00003	85%	10003	U00003	85%	10003	U00003	85%
10004	U00004	80%	10004	U00004	80%	10004	U00004	80%	10004	U00004	80%	10004	U00004	80%
10005	U00005	75%	10005	U00005	75%	10005	U00005	75%	10005	U00005	75%	10005	U00005	75%
10006	U00006	70%	10006	U00006	70%	10006	U00006	70%	10006	U00006	70%	10006	U00006	70%
10007	U00007	65%	10007	U00007	65%	10007	U00007	65%	10007	U00007	65%	10007	U00007	65%
10008	U00008	60%	10008	U00008	60%	10008	U00008	60%	10008	U00008	60%	10008	U00008	60%
10009	U00009	55%	10009	U00009	55%	10009	U00009	55%	10009	U00009	55%	10009	U00009	55%
10010	U00010	50%	10010	U00010	50%	10010	U00010	50%	10010	U00010	50%	10010	U00010	50%
10011	U00011	45%	10011	U00011	45%	10011	U00011	45%	10011	U00011	45%	10011	U00011	45%
10012	U00012	40%	10012	U00012	40%	10012	U00012	40%	10012	U00012	40%	10012	U00012	40%
10013	U00013	35%	10013	U00013	35%	10013	U00013	35%	10013	U00013	35%	10013	U00013	35%
10014	U00014	30%	10014	U00014	30%	10014	U00014	30%	10014	U00014	30%	10014	U00014	30%
10015	U00015	25%	10015	U00015	25%	10015	U00015	25%	10015	U00015	25%	10015	U00015	25%
10016	U00016	20%	10016	U00016	20%	10016	U00016	20%	10016	U00016	20%	10016	U00016	20%
10017	U00017	15%	10017	U00017	15%	10017	U00017	15%	10017	U00017	15%	10017	U00017	15%
10018	U00018	10%	10018	U00018	10%	10018	U00018	10%	10018	U00018	10%	10018	U00018	10%
10019	U00019	5%	10019	U00019	5%	10019	U00019	5%	10019	U00019	5%	10019	U00019	5%
10020	U00020	0%	10020	U00020	0%	10020	U00020	0%	10020	U00020	0%	10020	U00020	0%