Koide et al., 2004

Supplemental Material

Supplemental Methods

Pathogenicity tests. Six-month-old uniform sized Pera sweet orange trees (*Citrus sinensis*) were maintained without watering for 2 days before inoculation for easier uptake of the bacterial suspension. Plants were inoculated as follows: 50 µl drops of the bacterial culture were applied in 20 different places along the stem and the drops were introduced in the xylem by puncturing with a 22 gauge x 1" needle. For each strain, a total of 15 plants were inoculated. Four plants were treated with *X. fastidiosa*-free PW medium in the same way and used as negative control. In addition, four other plants were maintained without inoculation. Inoculated and control plants were kept in a greenhouse and watered every 2 days. Five plants (1 month old seedlings) of *Nicotiana tabaccum* (accession clevelandii) were inoculated with each culture as described above and kept in a greenhouse. Detection of *X. fastidiosa* in host plants followed the procedure described in (1).

DNA Microarray Construction. PCR primers were designed to amplify unique internal fragments of 200-1000 bp of each predicted CDS described in the annotated genome sequence of *X. fastidiosa* strain 9a5c (http://aeg.lbi.ic.unicamp.br/xf). Primers (18-23mers) with equivalent predicted melting temperature were designed with the use of a perl program that ran PRIMER3 (http://www-genome.wi.mit.edu/genome_software/other/primer3.html) for the complete CDS list, automatically testing many parameter settings and also guaranteeing that primers hybridized only to a single genome location. Oligonucleotides were synthesized by MWG and Operon Technologies. Genomic or cosmid DNA, obtained in the *X. fastidiosa* genome sequencing project (3), were used as template in the first round of PCR amplification, and 200-fold-diluted PCR products were used as templates for PCR reamplification to increase product concentration when necessary. The reactions were done in 96-well plates. The mixture in each well contained 100 ng of DNA, 0.5 U of Biolase Taq polymerase (Bioline), 0.2 mM of each dNTP (Invitrogen), 1.5 mM $MgCl_2$ and the primers at 0.5 µM, in a total volume of 100 µl. A 5min denaturing step at 95˚C was applied, followed by 40 cycles of 95˚C for 45s, 50˚C for 30s, 72˚C for 1min and a final step at 72˚C for 10min. 4 µl of each PCR reaction were checked for product size and concentration by

electrophoresis in 1.2% agarose gels. The amplicons were then purified with 96-well MultiScreen purification plates (Millipore) and an equal volume of dimethyl sulfoxide was added to the purified products (~100 ng/µl final concentration). Generation III DNA spotter (Amersham Biosciences) was used to array the samples onto coated type-7 glass slides (Amersham Biosciences). This spotter arrays two technical replicas of each sample, one in each longitudinal half of the slides. Thus, a 6152-element array was printed, representing 2692 CDS spotted at least in duplicate. After deposition, the spotted DNA samples were crosslinked to the coated slides by applying 50 mJ of UV light and the slides were stored desiccated at ~10% relative humidity at room temperature until use.

DNA and RNA Labeling . Whole genomic DNA from either 9a5c or J1a12 strain was fragmented by shearing and used in the synthesis of targets by direct incorporation of fluorescent nucleotide analogs (Cy3-dCTP or Cy5-dCTP, Amersham Biosciences) as described in http://cmgm.stanford.edu/pbrown/protocols/4_genomic.html and using random nonamers. Total RNA was extracted using hot phenol procedure adapted from http://caulobacter.stanford.edu/CellCycle/protocols/RNAIsolation.htm and contaminating DNA was removed with RQ1-RNAse free DNAse (Promega). 20 µg of total RNA was used for cDNA labeling using CyScribe Post-Labelling Kit (Amersham Biosciences) and random nonamers.

Hybridization conditions. Labeled DNA or cDNA fragments from both strains were combined in the hybridization mixture containing 50% formamide and hybridization buffer (Amersham Biosciences) in a final volume of 54µl. The mixture was heated to 92˚C for 2min, cooled on ice and applied to the microarray. A cover slip was used to spread the solution throughout the microarray and the slide was then placed in a 50mL Falcon tube that was sealed and horizontally positioned for hybridization in a 42˚C water bath for 16 hours. After hybridization, slides were washed at 55˚C for 10 min in 1x SSC buffer containing 0.2% SDS and twice for 10min in 0.1x SSC buffer containing 0.2% SDS, followed by 1min at room temperature in 0.1x SSC and a quick rinse in ddH_2O. After drying with N_2 in a clean room, the slides were ready for scanning.

Data Acquisition and Analysis. A Generation III DNA scanner (Amersham Biosciences) was used to acquire monochromatic images of 10μm/pixel from the microarray slides, corresponding to channels Cy3 (532 nm laser and PMT at 700 V) and Cy5 (633 nm and PMT at 750 V). For each slide, images of the left and right replicas were recovered. Images were analyzed with ArrayVision 6.0 software (Imaging Research Inc.). The artifact-removed mean density (ARMdens) of signal intensity was the measure chosen for signal quantification. This measure removes pixels with signals above 4 MADs (median of absolute deviation) of the mean signal intensity of all pixels within a spot, such as those representing dust particles. The median of the background intensity was calculated for a frame of 24000 μm^2 around each spot of 180 μm in diameter and the value was subtracted from the spot's ARMdens value. This represents the raw data used in the normalization procedure.

Normalization Procedure. Several imbalance errors affect the true ratio measure. We have assumed that all imbalance, due to enzyme efficiency, wavelength detection, dye brightness, etc, can be approximated by multiplicative factors that are contained into just one normalization constant that depends non-linearly on signal intensities. In order to normalize the Cy3 and Cy5 signal intensities (I_{Cy3} and I_{Cy5}) we have used the hypothesis that the great majority of genes are equally present in both strains (same number of copies) and therefore the predominant ratio must be one. This is a reasonable hypothesis given that our microarray contains fragments of all CDS of the 9a5c strain genome. We have performed the LOWESS fitting (included in the R package, available at http://www.r-project.org) on M vs A plot in order to obtain, locally and non-parametrically, the normalization constant and thus normalized ratios, following (5). As defined in that work, $A = 1/2 * \log_2(I_{Cy5} * I_{Cy3})$ and $M = \log2(I_{Cy5} / I_{Cy3})$. Lowess normalization eliminated the dye bias of the ratios, therefore these were all calculated as J1a12 / 9a5c (independent of dye) for a better visualization in M-A plots.

Intensity-dependent ratio cutoff level. We have performed homotypical hybridizations (9a5c strain DNA labeled with either Cy3-dCTP or Cy5-dCTP and hybridized simultaneously to the same microarray) in order to derive intrinsic experimental variability

of the 1:1 ratio (noise) and to set an upper and lower limit for this noise. With this approach, we have detected a clear dependence between ratio, estimated by M, and total foreground intensity of each CDS, estimated by A. CDS with intensities above but close to background strongly varied inside [0;∞], or in logarithm scale, inside [-∞;∞] when the expected result is ratio = 1 or $\log_2$(ratio) = 0. We used 3 independent hybridization experiments (3 slides, 6 images) as samplings from experimental error around 1:1 bi-dimensional probability density distribution. We conditioned this distribution in arbitrary A intervals to make it one-dimensional and estimated the density distribution using Kernel Density Estimators (2). Finally, we integrated this density around mode peak until 0.995 probability was reached, to determine an intensity-dependent ratio cutoff level. These cutoffs levels were subsequently used in the analysis of replicas of 9a5c *vs* J1a12 hybridization experiments; spots outside these credibility intervals present strong evidence against 1:1 ratio.

CDS classification process. Four categories were defined for the CDS in the J1a12 genome based on its orthologous 9a5c counterpart: (i) equally present in both strains, (ii) highly divergent or absent in J1a12 (iii) divergent and (iv) higher copy number in J1a12. Category (i) includes all the CDS for which ≥60% of the replicas were inside the credibility intervals. In order to separate categories (ii) and (iii), we performed 4 control hybridizations with DNA from the sequenced *X. fastidiosa* strain Temecula (Tc) (4). The 9a5c amplicons were BLASTed locally against the Tc genome (http://aeg.lbi.ic.unicamp.br/world/xfpd). Amplicons with ≤20% nucleotide identities were used to determine the "highly divergent or absent" M value cut-off threshold. The M value with the least false callings was –1.7. At this threshold, 19 false positives and 23 false negatives were observed (0.76% and 0.92%, respectively). Category (ii) includes CDS with *P*-value smaller than 0.05 in a *t*-test against the null hypothesis H_0: M ≥ –1.7. This threshold was further supported by sequencing some amplicons from J1a12 that were outside the credibility intervals and checking the divergence between 9a5c and J1a12 sequences. The remaining CDS were considered (iii) divergent if M-values were negative or (iv) present at higher copy number in J1a12 (positive M values). CDS with low replicas´ consistency (less than 60% of replicas in a single category) were excluded from the analysis.

Validation of microarray data. PCR and RT-PCR were performed using CDS specific primers were performed to confirm the status of CDS in the genome of J1a12 strain determined using the microarray data. The reactions were carried out with genomic DNA or cDNA from strain 9a5c or J1a12 using 35 cycles of amplification.A 4µl sample of each reaction was electrophoresed in 1.2% agarose gels and DNA was stained with ethidium bromide. The amplicons were then classified by visual inspection as absent, same copy number or more abundant in J1a12 strain in relation to 9a5c. In addition, DNA sequence determination was carried out for a few CDS. For that, the PCR products were cloned in pGEM-TEasy vector (Promega) and dideoxy sequencing reactions were performed using 100 ng of plasmid DNA in Big Dye Terminator sequencing reactions (Applied Biosystems) according to the manufacturer's instructions. Sequencing reactions were carried out using either CDS specific or T7 promoter primers.

References:

1. **Pooler, M. R., and J. S. Hartung.** 1995. Specific PCR detection and identification of *Xylella fastidiosa* strains causing citrus variegated chlorosis. Curr. Microbiol. **31:**377-381.
2. **Silverman, B. W.** 1986. Density estimation. Chapman and Hall, London.
3. **Simpson, A. J., F. C. Reinach, P. Arruda, F. A. Abreu, M. Acencio, R. Alvarenga, L. M. Alves, J. E. Araya, G. S. Baia, C. S. Baptista, M. H. Barros, E. D. Bonaccorsi, S. Bordin, J. M. Bove, M. R. Briones, M. R. Bueno, A. A. Camargo, L. E. Camargo, D. M. Carraro, H. Carrer, N. B. Colauto, C. Colombo, F. F. Costa, M. C. Costa, C. M. Costa-Neto, L. L. Coutinho, M. Cristofani, E. Dias-Neto, C. Docena, H. El-Dorry, A. P. Facincani, A. J. Ferreira, V. C. Ferreira, J. A. Ferro, J. S. Fraga, S. C. Franca, M. C. Franco, M. Frohme, L. R. Furlan, M. Garnier, G. H. Goldman, M. H. Goldman, S. L. Gomes, A. Gruber, P. L. Ho, J. D. Hoheisel, M. L. Junqueira, E. L. Kemper, J. P. Kitajima, J. E. Krieger, E. E. Kuramae, F. Laigret, M. R. Lambais, L. C. Leite, E. G. Lemos, M. V. Lemos, S. A. Lopes, C. R. Lopes, J. A. Machado, M. A. Machado, A. M. Madeira, H. M. Madeira, C. L. Marino, M. V. Marques, E. A. Martins, E. M. Martins, A. Y. Matsukuma, C. F. Menck, E. C. Miracca, C. Y. Miyaki, C. B. Monteriro-Vitorello, D. H. Moon, M. A. Nagai, A. L. Nascimento, L. E. Netto, A. Nhani, Jr., F. G. Nobrega, L. R. Nunes, M. A. Oliveira, M. C. de Oliveira, R. C. de Oliveira, D. A. Palmieri, A. Paris, B. R. Peixoto, G. A. Pereira, H. A. Pereira, Jr., J. B. Pesquero, R. B. Quaggio, P. G. Roberto, V. Rodrigues, M. R. A. J. de, V. E. de Rosa, Jr., R. G. de Sa, R. V. Santelli, H. E. Sawasaki, A. C. da Silva, A. M. da Silva, F. R. da Silva, W. A. da**

Silva, Jr., J. F. da Silveira, et al. 2000. The genome sequence of the plant pathogen *Xylella fastidiosa*. Nature **406:**151-157.

4. **Van Sluys, M. A., M. C. de Oliveira, C. B. Monteiro-Vitorello, C. Y. Miyaki, L. R. Furlan, L. E. Camargo, A. C. da Silva, D. H. Moon, M. A. Takita, E. G. Lemos, M. A. Machado, M. I. Ferro, F. R. da Silva, M. H. Goldman, G. H. Goldman, M. V. Lemos, H. El-Dorry, S. M. Tsai, H. Carrer, D. M. Carraro, R. C. de Oliveira, L. R. Nunes, W. J. Siqueira, L. L. Coutinho, E. T. Kimura, E. S. Ferro, R. Harakava, E. E. Kuramae, C. L. Marino, E. Giglioti, I. L. Abreu, L. M. Alves, A. M. do Amaral, G. S. Baia, S. R. Blanco, M. S. Brito, F. S. Cannavan, A. V. Celestino, A. F. da Cunha, R. C. Fenille, J. A. Ferro, E. F. Formighieri, L. T. Kishi, S. G. Leoni, A. R. Oliveira, V. E. Rosa, Jr., F. T. Sassaki, J. A. Sena, A. A. de Souza, D. Truffi, F. Tsukumo, G. M. Yanai, L. G. Zaros, E. L. Civerolo, A. J. Simpson, N. F. Almeida, Jr., J. C. Setubal, and J. P. Kitajima.** 2003. Comparative analyses of the complete genome sequences of Pierce's disease and citrus variegated chlorosis strains of *Xylella fastidiosa*. J. Bacteriol. **185:**1018-1026.

5. **Yang, Y. H., S. Dudoit, P. Luu, D. M. Lin, V. Peng, J. Ngai, and T. P. Speed.** 2002. Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. Nucleic Acids Res. **30:**e15.

Supplementary Tables

Table 1S: CDS divergent in strain J1a12

Gene ID[a]	Product	M[c]
XF0075	hypothetical protein	-0.56
XF0078	fimbrial adhesin precursor	-0.77
XF0157	hypothetical protein	-0.91
XF0262	colicin V precursor	-0.92
XF0263	colicin V precursor	-0.75
XF0500	phage-related repressor protein	-1.51
XF0501	conserved hypothetical protein	-1.30
XF0614	hypothetical protein	-0.35
XF0626	hypothetical protein	-1.00
XF0659	cell cycle protein	-0.41
XF0663	hypothetical protein	-1.53
XF0665	hypothetical protein	-1.19
XF0666	hypothetical protein	-1.19
XF0668	hemolysin-type calcium binding protein	-0.68
XF0684	phage-related protein	-0.70
XF0696	phage-related repressor protein[b]	-1.78
XF1057	hypothetical protein	-0.58
XF1306	hypothetical protein	-1.33
XF1588	hypothetical protein[b]	-1.71
XF1589	plasmid stabilization protein	-1.62
XF1590	plasmid stabilization protein	-0.91
XF1609	glucose/galactose transporter	-0.74
XF1664	hypothetical protein[b]	-1.71
XF1720	hypothetical protein	-0.40
XF1733	tryptophan repressor binding protein	-0.38
XF1746	alcohol dehydrogenase	-0.70
XF1756	hypothetical protein	-0.60
XF1758	hypothetical protein	-0.58
XF1786	phage-related protein	-0.39
XF1851	serine protease	-0.54

XF1859	hypothetical protein	-0.60
XF1862	conserved hypothetical protein	-1.19
XF1873	conserved hypothetical protein	-0.70
XF1877	hypothetical protein	-0.76
XF1883	hypothetical protein	-0.73
XF1968	methyltransferase	-1.44
XF2193	hypothetical protein	-1.47
XF2194	hypothetical protein	-1.10
XF2195	hypothetical protein[b]	-1.75
XF2217	imidazoleglycerolphosphate dehydratase/histidinol-phosphate phosphatase bifunctional enzyme	-0.54
XF2307	hypothetical protein[b]	-1.73
XF2406	hypothetical protein	-1.02
XF2407	bacteriocin	-0.93
XF2542	fimbrial protein	-1.07
XF2722	type I restriction-modification system specificity determinant	-1.25
XF2726	type I restriction-modification system specificity determinant[b]	-1.73
XF2744	hypothetical protein	-0.42
XF2768	hypothetical protein	-1.38
XF2770	hypothetical protein[b]	-1.78
XF2772	hypothetical protein[b]	-1.78

[a] the CDS shown here have a hybridization intensity ratio $-1.7 < M < -0.3$.

[b] these CDS have a *P*-value greater than 0.05 in the *t*-test and therefore were not classified in category (ii), i.e. absent/highly divergent in J1a12.

[c] DNA hybridization intensity ratios $M = \log_2(J1a12/9a5c)$ were calculated. For each gene, the values shown are the mean from at least 8 replicas

Table 2S: CDS possibly presenting higher number of copies in J1a12.

Gene ID[a]	Product	M[b]
XF0391	hypothetical protein	0.37
XF0487	fimbrillin	0.68
XF0488	hypothetical protein	0.77
XF0490	hypothetical protein	0.44
XF0493	hypothetical protein	0.67
XF0512	hypothetical protein	0.65
XF0513	phage-related endolysin	0.68
XF0514	hypothetical protein	0.76
XF0515	hypothetical protein	0.68
XF0516	hypothetical protein	0.82
XF0517	hypothetical protein	0.56
XF0518	hypothetical protein	0.69
XF0519	hypothetical protein	0.76
XF0521	conserved hypothetical protein	0.58
XF0523	hypothetical protein	0.56
XF0533	conserved hypothetical protein	0.63
XF0730	phage-related tail protein	0.32
XF1655	hypothetical protein	0.36
XF1932	hypothetical protein	1.37
XF1933	exodeoxyribonuclease III	1.23
XF1934	HetI protein	1.17
XF1935	glucose inhibited division protein	1.01
XF1936	transketolase 1	1.23
XF1937	proton glutamate symport protein	1.10
XF1938	hypothetical protein	1.07
XF2112	conserved hypothetical protein	0.52
XF2113	hypothetical protein	0.70
XF2114	hypothetical protein	0.61
XF2122	hypothetical protein	0.35
XF2190	hypothetical protein	0.60
XF2288	phage-related integrase	0.66
XF2289	hypothetical protein	0.66
XF2480	phage-related tail protein	0.36

XF2481	phage-related tail protein	0.38
XF2498	phage-related portal protein	0.34
XF2499	hypothetical protein	0.38
XF2501	phage-related protein	0.38
XF2512	hypothetical protein	0.85
XF2514	hypothetical protein	0.48
XFa0027	plasmid maintenance protein	1.03

[a] the CDS shown here have a hybridization intensity ratio M > 0.3.

[b] DNA hybridization intensity ratios $M = \log_2(J1a12/9a5c)$ were calculated. For each gene, the values shown are the mean from at least 8 replicas

Table 3S: CDS presenting higher RNA expression level (2-fold or more) in strain 9a5c.

Gene ID[a]	Product	M[b]
XF0172	hypothetical protein	-1.57
XF0278	hypothetical protein	-2.27
XF0302	hypothetical protein	-2.08
XF0395	bacterioferritin	-1.76
XF0562	sec-independent protein translocase	-1.22
XF0692	hypothetical protein	-1.27
XF0968	hypothetical protein	-1.73
XF1056	hypothetical protein	-2.35
XF1063	6-phosphogluconolactonase	-1.76
XF1107	carbamoyl-phosphate synthase large chain	-1.14
XF1111	peptide chain release factor 2	-2.09
XF1205	hypothetical protein	-2.30
XF1389	cytochrome O ubiquinol oxidase, subunit I	-1.55
XF1390	cytochrome O ubiquinol oxidase, subunit II	-1.83
XF1676	conserved hypothetical protein	-1.61
XF1789	hypothetical protein	-1.43
XF1797	porphyrin biosynthesis protein	-1.33
XF1798	hypothetical protein	-1.37
XF2377	hypothetical protein	-1.31
XF2475	tRNA/rRNA methyltransferase	-1.18
XF2506	hypothetical protein	-2.69
XF2516	hypothetical protein	-1.79
XF2526	phage-related protein	-1.70
XFa0015	conjugal transfer protein	-1.54
XFa0060	plasmid replication protein	-2.81

[a] the CDS shown here have a *P*-value smaller than 0.05 in a *t*-test for the null hypothesis H_0: $M \geq -1.0$.

[b] DNA hybridization intensity ratios $M = \log_2(J1a12/9a5c)$ were calculated. For each gene, the values shown are the mean from at least 4 replicas

Table 4S: CDS presenting higher RNA expression level (2-fold or more) in strain J1a12.

Gene ID[a]	Product	M[b]
XF0002	DNA polymerase III, beta chain	2.20
XF0123	recombination protein RecA	1.34
XF0155	hypothetical protein	1.82
XF0275	adenylate kinase	1.21
XF0280	leucine aminopeptidase	1.36
XF0301	hypothetical protein	1.75
XF0339	conserved hypothetical protein	1.58
XF0377	cytochrome P450-like enzyme	1.69
XF0449	conserved hypothetical protein	1.49
XF0615	60kDa chaperonin	2.25
XF0616	10kDa chaperonin	2.30
XF0872	outer membrane protein	1.96
XF0974	hypothetical protein	1.49
XF0975	polyphosphate-selective porin O	1.72
XF1007	hypothetical protein	1.71
XF1009	hypothetical protein	1.80
XF1024	outer membrane protein H.8 precursor	2.05
XF1164	50S ribosomal protein L5	1.29
XF1165	30S ribosomal protein S14	1.55
XF1168	50S ribosomal protein L18	1.34
XF1169	30S ribosomal protein S5	1.44
XF1175	30S ribosomal protein S4	1.16
XF1177	50S ribosomal protein L17	1.48
XF1210	glutathione S-transferase	1.50
XF1216	colicin V secretion protein	1.44
XF1248	hypothetical protein	3.40
XF1476	ABC transporter membrane protein	1.37
XF1500	ATP sulfurylase, small subunit	1.58
XF1693	hypothetical protein	2.45
XF1881	hypothetical protein	1.44
XF1891	di-tripeptide ABC transporter membrane protein	2.06
XF2151	hypothetical protein	1.21

XF2237	conserved hypothetical protein	1.49
XF2241	periplasmic protease	1.60
XF2336	two-component system, regulatory protein	1.28
XF2402	hypothetical protein	1.63
XF2561	30S ribosomal protein S6	1.51
XF2625	heat shock protein	1.13
XF2750	hypothetical protein	1.13
XF2755	hypothetical protein	1.55
XF2773	hypothetical protein	2.86
XF2779	hypothetical protein	2.10
XFa0031	hypothetical protein	1.54
XFa0054	hypothetical protein	1.65

[a] the CDS shown here have a *P*-value smaller than 0.05 in a *t*-test for the null hypothesis H_0: $M \leq 1.0$.

[b] DNA hybridization intensity ratios $M = \log_2(J1a12/9a5c)$ were calculated. For each gene, the values shown are the mean from at least 4 replicas