Comparative and functional genomic analyses of the pathogenicity of phytopathogen *Xanthomonas campestris* pv. *campestris*

Wei Qian,[1,9] Yantao Jia,[1,9] Shuang-Xi Ren,[2,3,9] Yong-Qiang He,[4,9] Jia-Xun Feng,[4,9] Ling-Feng Lu,[2,9] Qihong Sun,[1] Ge Ying,[1] Dong-Jie Tang,[4] Hua Tang,[5] Wei Wu,[1] Pei Hao,[6,7] Lifeng Wang,[1] Bo-Le Jiang,[4] Shenyan Zeng,[1] Wen-Yi Gu,[2] Gang Lu,[2] Li Rong,[8] Yingchuan Tian,[1] Zhijian Yao,[8] Gang Fu,[2] Baoshan Chen,[4] Rongxiang Fang,[1] Boqin Qiang,[8] Zhu Chen,[2] Guo-Ping Zhao,[2,3,7,10] Ji-Liang Tang,[4,10] and Chaozu He[1,10]

[1]National Key Laboratory of Plant Genomics, Institute of Microbiology, Chinese Academy of Sciences, Beijing 100080, P. R. China; [2]Laboratory of Health and Disease Genomics, Chinese National Human Genome Center at Shanghai, Shanghai 201203, P. R. China; [3]National Key Laboratory of Genetic Engineering/Department of Microbiology, School of Life Science, Fudan University, Shanghai 200433, P. R. China; [4]Guangxi Key Laboratory of Subtropical Bioresources Conservation and Utilization, Guangxi University, Nanning 530004, Guangxi, P. R. China; [5]Department of Ecology and Evolution, University of Chicago, Chicago, Illinois 60637, USA; [6]Shanghai Center for Bioinformation Technology, Shanghai 200235, P. R. China; [7]Bioinformation Center/Institute of Plant Physiology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200032, P. R. China; [8]Chinese National Human Genome Center at Beijing, Beijing 100176, P. R. China

Xanthomonas campestris pathovar *campestris* (*Xcc*) is the causative agent of crucifer black rot disease, which causes severe losses in agricultural yield world-wide. This bacterium is a model organism for studying plant–bacteria interactions. We sequenced the complete genome of *Xcc* 8004 (5,148,708 bp), which is highly conserved relative to that of *Xcc* ATCC 33913. Comparative genomics analysis indicated that, in addition to a significant genomic-scale rearrangement cross the replication axis between two IS*1478* elements, loss and acquisition of blocks of genes, rather than point mutations, constitute the main genetic variation between the two *Xcc* strains. Screening of a high-density transposon insertional mutant library (16,512 clones) of *Xcc* 8004 against a host plant (*Brassica oleraceae*) identified 75 nonredundant, single-copy insertions in protein-coding sequences (CDSs) and intergenic regions. In addition to known virulence factors, full virulence was found to require several additional metabolic pathways and regulatory systems, such as fatty acid degradation, type IV secretion system, cell signaling, and amino acids and nucleotide metabolism. Among the identified pathogenicity-related genes, three of unknown function were found in *Xcc* 8004-specific chromosomal segments, revealing a direct correlation between genomic dynamics and *Xcc* virulence. The present combination of comparative and functional genomic analyses provides valuable information about the genetic basis of *Xcc* pathogenicity, which may offer novel insight toward the development of efficient methods for prevention of this important plant disease.

[Supplemental material is available online at www.genome.org. The sequence data of *Xcc* 8004 from this study have been submitted to GenBank under accession no. CP000050.]

Xanthomonas campestris is a gram-negative, pathogenic bacterium that belongs to the γ-subdivision of Proteobacteria. It was genetically differentiated into over 141 pathovars (pv.), each with a specific host range (Dye et al. 1980; Swings and Civerolo 1993). The variant *X. campestris* pv. *campestris* (*Xcc*) generally invades and multiplies in cruciferous plant vascular tissues, resulting in the characteristic "black rot" symptoms of blackened veins and V-shaped necrotic lesions at the foliar margin (Alvarez 2000). *Xcc* invades into plant tissues through hydathodes, stomates, roots, or wounds, and infects a wide range of plants in the crucifer family (*Brassicaceae*), including broccoli, cabbage, cauliflower, radish, and the model plant *Arabidopsis thaliana*. Thus, this bacterium represents a problematic bacterial pathogen. Periodic epidemics of black rot disease have occurred world-wide, especially in the developing regions of Africa and Asia, where high temperatures and humidity can aggravate the damage (Swings and Civerolo 1993). Each epidemic can cause substantial yield loss in agricultural production. However, an efficient, low-pollutive control treatment has not, as of yet, been developed (Swings and Civerolo 1993).

[9]These authors contributed equally to this work.
[10]Corresponding authors.
E-mail hecz@im.ac.cn; fax 86-10-62548243.
E-mail gpzhao@sibs.ac.cn; fax 86-21-64837495.
E-mail jltang@gxu.edu.cn; fax 86-771-3237873.
Article and publication date are at http://www.genome.org/cgi/doi/10.1101/gr.3378705. Article published online before print in May 2005.

The recent rapid development of genomics has brought a paradigm shift to bacterial pathogenesis research (Wren 2000). Although the complete genomic sequence of *Xcc* ATCC 33913 (da Silva et al. 2002) provided a profile of genetic information to explore *Xcc*'s biological characteristics, the functions of about one-third of the protein-coding sequences (CDSs) are yet to be assigned, and a large repertoire of genes, including virulence determinants, have not been experimentally defined. Thorough annotations using comparative and functional genomic technologies are essential for gaining a systematic understanding of the biology of black rot disease and for facilitating the development of better preventive techniques.

We choose *Xcc* 8004 and its susceptible host plant cabbage (*Brassica oleraceae*) as the pathosystem for the present study. *Xcc* 8004 is a spontaneous rifampicin-resistant strain derived from *Xcc* NCPPB No. 1145 (isolated from infected cauliflower, *B. oleracea* var. *botrytis*, in Sussex, UK, 1958). It is a genetically amendable strain widely used for phytopathological studies, such as studies of secretion of extracellular enzymes and exopolysaccharides (Tang et al. 1991; Wilson et al. 1998), cell–cell signaling (Slater et al. 2000), and biofilm formation (Dow et al. 2003). *Xcc* ATCC 33913 was isolated from cabbage (*B. oleracea* var. *gemmifera*) and its genome served as a reference for comparison. Although *Xcc* ATCC 33913 and *Xcc* 8004 cause similar necrogenic symptoms, these two strains differ in their growth rate on minimal medium and in their host range in the five commercial cruciferous plants (Table 1). In particular, *Xcc* ATCC 33913 exhibited weak or little virulence against the tested cultivars (cv) of cabbage (*B. oleracea* cv. Jingfeng 1) and radish (*Raphanus sativus* cv. Huaye).

We sequenced and annotated the complete genome of *Xcc* 8004, and compared it with that of *Xcc* ATCC 33913 (da Silva et al. 2002). Subsequently, we screened an *Xcc* 8004 transposon insertional mutant library with 4× genomic coverage (Sun et al. 2003) against its host plant *B. oleraceae* for pathogenicity-deficient mutants, from which 75 genes with diverse functions were identified, including previously undefined genes and three strain-specific genes of unknown function. This screening is one of the largest in scale for either animal or plant bacterial pathogens (Kavermann et al. 2003; Kurz et al. 2003; Bae et al. 2004).

Results

General genomic features

The *Xcc* 8004 genome resides on a single circular chromosome (5,148,708 bp, Fig. 1.) with a G+C content of 64.94%, which is larger than that of the *Xcc* ATCC 33913 (5,076,187 bp, Table 2). Like *Xcc* ATCC 33913 (da Silva et al. 2002), *Xcc* 8004 was found not to have a plasmid. The replication origin of the *Xcc* 8004 genome was predicted to be at the intergenic region of *dnaA* and *dnaN* by GC-skew analysis and a similarity search. Approximately 63% of the CDSs (2671) were assigned to biological functions (Table 2). Thirty-two phage-related genes, including two copies of ϕ*Lf* filamentous phages that specifically infect *Xcc* (Lin et al. 1996), were identified in the genome. In addition, there were 115 insertion sequence elements, ranging from 0.8 to 1.6 kb and belonging to 15 groups, distributed throughout the genome. The 4273 predicted CDSs in the *Xcc* 8004 genome are depicted in Supplemental Figure 1.

Comparative genomics

A high degree of gene content conservation was observed between the genomes of *Xcc* 8004 and *Xcc* ATCC 33913. The majority of the CDSs (3467) found in *Xcc* 8004 are identical to those in *Xcc* ATCC 33913, not only at the amino-acid level, but also at the nucleic-acid sequence level. However, single-nucleotide polymorphisms (SNPs) were detected in 498 CDSs with identical gene length, which account for about 12% of the total CDSs. In order to assess the effect of nonsynonymous mutations upon the intraspecies differentiation of the two strains, we analyzed the ratios of total nonsynonymous SNPs vs. total synonymous SNPs (A/S) of these 498 genes in four functional categories regarding pathogenicity ([I] normal physiology, [II] unknown functions, [III] virulence factors, and [IV] pathogenicity-related, Table 3). The A/S ratio for pathogenicity-associated genes (category III and IV) did not differ significantly from that of the other groups (Table 3). Therefore, the overall intensity of the selection force for the genes among the different groups was very similar. In other words, pathogenicity-associated genes did not show any significant bias in response to selection compared with the other

Table 1. Pathogenicity properties of wild-type strains of *Xanthomonas campestris* pv. *campestris* and the *Xcc* 8004 strain-specific mutants tested on various host plants

	Average virulence level (standard deviation)[a]				
	Cabbage	Radish (Huaye)	Radish (Xiaojingzhong)	Chinese cabbage	Pakchoi cabbage
Wild-type strains					
Xcc 8004	4 (0.00)	4 (0.00)	4 (0.00)	4 (0.00)	4 (0.00)
Xcc ATCC 33913	0.67 (0.49)[b]	0 (0.00)	4 (0.00)	3.7 (0.49)[b]	4 (0.00)
Xcc 8004 strain-specific segment mutants					
XC2055	1.9 (0.20)[b]	4 (0.00)	4 (0.00)	1.9 (0.20)[b]	2.9 (0.20)[b]
XC2068	1.8 (0.39)[b]	2.3 (0.65)[b]	3.3 (0.87)[b]	3.8 (0.62)	3.8 (0.45)
XC2416	2.0 (0.14)[b]	0.3 (0.49)[b]	3.8 (0.39)	2.9 (0.52)[b]	4 (0.00)
Control	0	0	0	0	0

[a]Virulence level scale and plant inoculation approach are described in Methods (Dow et al. 1990). Each average virulence level was calculated from a data set consisting of ≥36 inoculation sites distributed over ≥3 leaves on ≥6 individual plants. Symptoms were scored on the 10th day post-inoculation. Standard deviations are shown in the parentheses.
[b]$P < 0.05$ vs. *Xcc* 8004, by *t*-test.

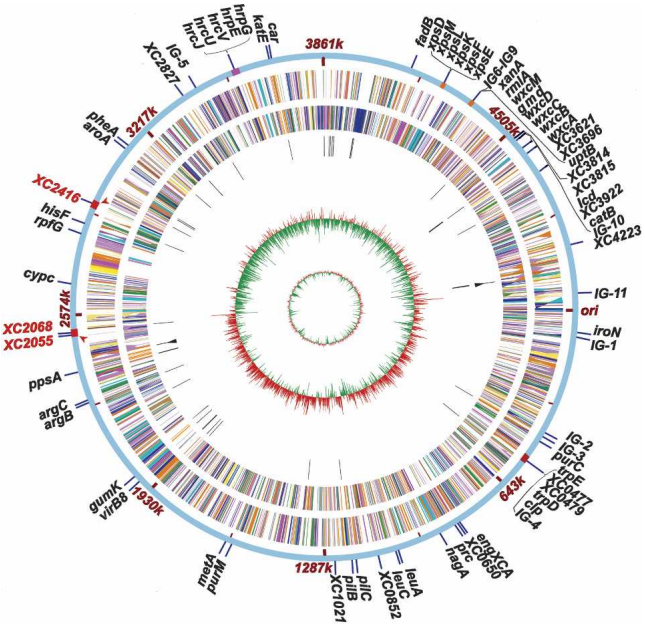

Figure 1. Circular representation of the *Xcc* 8004 genome and EZ::TN transposon insertion sites of virulence-reduced mutants. Circles range from 1 (*outer* circle) to 7 (*inner* circle). (Circle 1) EZ::TN insertional locations, gene names, or codes are according to Supplemental Table 3; (IG) intergenic region; *ori*, the origin of DNA replication; Red arrowheads indicate *Xcc* 8004 strain-specific chromosomal segments. (Circle 2) CDSs on forward strand. (Circle 3) CDSs on reverse strand. (Circle 4) tRNA genes. (Circle 5) rRNA genes. (Circle 6) GC skew [(G-C)/(G+C)], red indicates values >0 and green <0. (Circle 7) G+C contents. Colors in circle 2 and 3 represent functional categories; orange for intermediary metabolism; green for biosynthesis of small molecules; blue for macromolecule metabolism; magenta for cell structure; khaki for cellular processes; green for biosynthesis of small molecules; light-blue for mobile genetic elements; pink for pathogenicity, virulence, and adaptation; yellow for hypothetical and conserved hypothetical proteins; dark-green for ORFs with an undefined category.

functional groups. However, the exact contribution of point mutation in phenotypic difference between the two strains can only be determined by examining the functional role of each alteration.

In addition to the SNP mutations noted, insertion/deletion and related variations were observed in 200 pairs of genes (Table 2). We also identified 108 and 62 CDSs unique to *Xcc* 8004 and *Xcc* ATCC 33913, respectively (Supplemental Table 1). The majority of these genes were localized to strain-specific chromosomal segments (Fig. 2).

Segment M (60 kb) and N (58 kb) are unique to *Xcc* 8004. Segment M possesses 44 CDSs. Within segment M are genes relevant to plasmid mobilization, recombination (e.g., integrase and DNA helicase), and cellular regulation, such as those encoding hemolysin activation protein (XC2404), ankyrin repeat protein (XC2428), and a chaperone-like ATPase (XC2433) with AAA^+ domain, as well as 28 genes of unknown function. The G+C content value at the synonymously variable third position (GC3s) of segment M is 0.580 ± 0.091, which is lower than that of the whole *Xcc* 8004 genome (0.801 ± 0.081). In addition, the codon usage of genes within this segment is distinctly biased from the whole genome (Supplemental Table 2). Therefore, segment M was likely to be acquired via horizontal gene transfer. Although most of the 52 CDSs in segment N encode hypothetical proteins, 39 of them have orthologs (BLASTP search, $e < 10^{-5}$) in the genome of a closely related species, *X. axonopodis* pv. *citri* (*Xac*) strain 306 (da Silva et al. 2002). A deletion event may have occurred in the genome of *Xcc* ATCC 33913 during its evolution, which eventually made this segment unique in *Xcc* 8004.

Segments D (18 kb) and G (37 kb) are *Xcc* ATCC 33913 specific (Fig. 2). Segment D contains seven genes, with gene products that include an oxidoreductase, a DNA methylase, and two DNA helicases. Segment G possesses 51 CDSs, which includes genes encoding two polymerase V subunits and a site-specific DNA-methyltransferase. Because most of the genes in this segment are phage related, it likely originated from phage integration.

The majority of the genes encoded by the two genomes were identical, and the nonsynonymous variations are unlikely to be the main driving force of evolution (see above). Thus, the identified strain-specific genes might contribute to the biological differences between the two *Xcc* strains. It is worth emphasizing that among the *Xcc* 8004 strain-specific genes, there are two type IV secretary pathway related genes (XC2031 and XC2040), and both of them possess TraG/TraD and ATPase domains. Findings with the gastric pathogen *Helicobacter pylori* (Israel et al. 2001) and the plant crown gall disease pathogen *Agrobacterium tumefaciens* (Cascales and Christie 2003) suggested that proteins of this kind are putative NTPases functioning as coupling factors to facilitate communication between a transport pore and its substrate molecules. In addition, *Xcc* 8004 contains several strain-specific genes that encode proteins with putative transport functions, including a transport transmembrane protein (XC2405), an arsenite efflux pump (resistant to arsenic toxicity, XC2294), and a high-affinity Fe^{2+}/Pb^{2+} permease (pumping out heavy metals, XC2295).

Genomic colinearity

Comparison of *Xcc* (strains 8004 and ATCC 33913) and *Xac* (strain 306) genomes revealed a dramatic pattern of genomic rearrangement (Fig. 2). In contrast to the considerable colinearity between the genomes of *Xcc* ATCC 33913 and *Xac*, significant rearrangements, including translocation, inversion, and insertion/deletion, were clearly recognized between the genomes of *Xcc* 8004 and *Xcc* ATCC 33913. These findings are consistent with the view that *Xac* and *Xcc* ATCC 33913 may represent the prototypes of these xanthmonads, while *Xcc* 8004 may have originated via recent recombination events. By comparing the pattern of genomic organization, one may notice that except for the DNA segments flanking the replication origin (segments A, B, and L, Fig. 2), the segments of the two *Xcc* genomes are located symmetrically at mirror-image positions across the replication axis. This significant rearrangement may have resulted from a single round of homologous recombination. There are two identical IS*1478*-related genes (*Xcc*0535 and *Xcc*3627) present at each extreme end of the predicted recombination site (on the end of segments L and B) on the *Xcc* ATCC 33913 genome. They may be the preferred sites for homologous recombination during replication (Tillier and Collins 2000). Segments I and K are also translocated across the replication axis, but they are inverted in their orientations relative to the other segments, which indicates that more than one round of recombination events occurred on them. A mechanism of this putative genomic rearrangement process is thus deduced and shown in Supplemental Figure 2 as a detailed, animated depiction.

Screen for virulence-reduced mutants

Using a highly efficient transposon-based mutagenesis system, we previously constructed a *Xcc* 8004 random insertional mutant library that covers the genome approximately four times (Sun et al. 2003). A total of 16,512 transformants were screened individually on the susceptible host plant cabbage (*B. oleracae* cv. Jingfeng 1) and 172 pathogenicity-deficient mutants were obtained. The virulence capacities of these mutants were confirmed by at least six rounds of independent inoculation experiments, with no less than 36 inoculation sites per mutant (see Methods). These multiple replications and numerous sites enabled growth phase variation and stochastic factors to be excluded. Among the pathogenicity-related mutations, 75 nonredundant, single-copy transposon insertional disrupted CDSs or intergenic regions were identified by Southern blotting and flanking sequence analysis, and then mapped to the *Xcc* 8004 genome (Fig. 1, the outermost circle). Although transposon insertion may cause polarity on downstream genes within an operon and thereby complicate the subsequent genetic analysis, these mutants remain informative, as most of the operons comprise functionally related genes. These mutated genes were assigned to diverse functional categories (Fig. 3; details in Supplemental Table 3). In addition to 25 previously defined virulence genes, such as *hrpG*, *clp*, *rpfG*, and *gumK*, 39 new genes and 11 intergenic regions were identified, providing insight into the genetic basis of *Xcc* pathogenicity (Fig. 3; details in Supplemental Table 3).

Table 2. Comparison of the genomes of *Xanthomonas campestris* pv. *campestris* 8004 and *Xcc* ATCC 33913

	Xcc 8004	*Xcc* ATCC 33913[a]
General features of the chromosome		
Genome size (bp)	5,148,708	5,076,187
GC content	64.94%	65.00%
Total number of predicted CDSs	4273 (87)[b]	4181
CDSs with functional assignment	2671 (1)	2708
Conserved hypothetical CDSs	1523 (27)	1276
Hypothetical CDSs	79 (59)	198
CDSs categorized by sequence variations		
Homologs with identical nucleic acid sequence	3467[c]	3408
Homologs with SNPs and the same length	498[d]	500[c]
Homologs with insertions/deletions	200	211[c]
Strain-specific genes	108	62
Average CDS length (bp)	1023	1030
Insertion sequence elements (IS)	115	109
Transfer RNA	53	53
tmRNA	1	1
Ribosomal RNA operons	2	2

[a]Genomic data of *Xcc* ATCC 33913 are according to da Silva et al. (2002) and GenBank (accession no. AE008922).
[b]The figures in parentheses indicate the number of CDSs that have the homologs in *Xcc* ATCC 33913, but were not annotated by da Silva et al. (2002).
[c]Signifies the existence of duplication CDSs.
[d]These CDSs were used in analyzing the variation of single-nucleotide polymorphisms (SNPs) between two *Xcc* genomes (Details in Table 3).

The distribution of insertions in the library was assessed previously by identifying both the genomic location and the functional distribution of 50 randomly chosen clones (Sun et al. 2003; this study). Three annotated pathogenicity-related genes (*virB8*, *rbfC*, and *bfeA*) were identified from these 50 candidates. This ratio (6%) was consistent with the overall expected ratio (250 annotated pathogenesis-related genes of a total of ~4200 genes, about 5.9%), and it is worth noting that *virB8* was also identified from the later mutant-screening process. We further assessed the distribution of the mutant library by comparing the ratio of the number of genes identified by this mutagenesis experiment for each functional category related to pathogenicity versus the number of genes of the same category annotated in the genome (Table 4). An obvious bias was observed among the different functional categories. Most mutations were identified in genes related to cell-surface component biosynthesis, enzyme-mediated detoxification (particularly catalase), and secretion systems. The extremely low frequencies (even zero) of certain subgroups are interesting topics for further discussion. On the other hand, general metabolism gene mutations that affect pathogenicity/virulence were found in relatively low frequencies (<10%). Mutations of genes involved in fatty-acid degradation (15.4%) and purine biosynthesis (13.6%), however, constituted exceptions to this trend. In addition to the relatively low coverage of the library (4×), one likely cause of the above-mentioned bias would be the pathogenesis judgment criteria of this study, which may have allowed for the selection of only those mutations that cause significant decreases of pathogenicity. The 75 nonredundant genes identified here under such selection criteria would be expected to be nontrivial for pathogenicity. However, extrapolations of these mutagenesis results require caution, with respect to the coverage and distribution bias issues of the analysis.

Table 3. Comparison of SNP frequency and distribution of CDSs between *Xcc* 8004 and *Xcc* ATCC 33913

Groups[a]	No. CDS	Total SNPs[b]	A	S	Average SNP/codon (10^{-3})	A/S	P_I[c]	P_{II}	P_{III}
I	269	1734	381	1353	1.656	0.282			
II	163	1345	315	1030	2.739	0.306	0.3633		
III	47	284	71	213	1.522	0.333	0.2902	0.6226	
IV	19	185	39	146	1.439	0.267	0.8531	0.5390	0.3857
Sum	498	3548	806	2742	1.915	0.294			

[a]Groups defined as (I) genes not involved in pathogenicity; (II) genes with undefined category and unknown-function; (III) genes whose products are virulence factors that interacted directly with host plant cells; (IV) gene that interacted indirectly with host cells, but is required for full bacterial virulence.
[b](SNPs) single-nucleotide polymorphisms; (A) number of nonsynonymous SNPs; (S) number of synonymous SNPs; (average SNP/codon) ratios of total SNPs vs. total number of codons; (A/S) ratios of total nonsynonymous SNPs vs. total synonymous SNPs.
[c]*P*-values from χ^2 test for the difference in A/S ratios in all the pairwise comparisons among the four functional groups.

Surface structure and appendages

Infection is initiated with bacterial attachment to and colonization of host tissues via surface structures and appendages. Specifically, type IV pili may contribute to bacterial pathogenesis by affecting adhesion, twitching mobility on a solid surface, and secretion or interac-

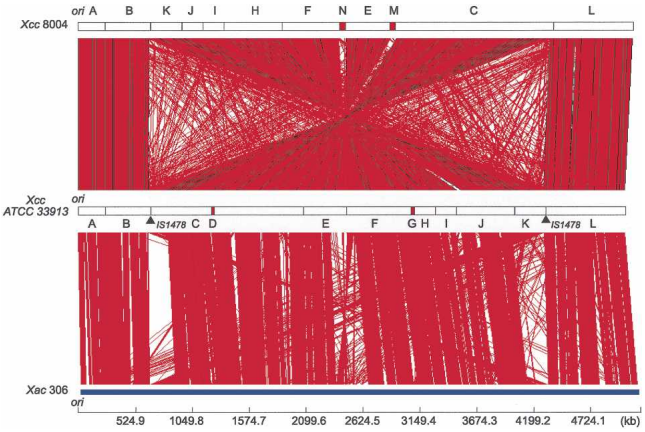

Figure 2. Linear genomic comparison of *Xanthomonas campestris* pv. *campestris* (*Xcc*) 8004, *Xcc* ATCC 33913, and *X. axonopodis* pv. *citri* (*Xac* 306). The red lines represent similar DNA sequences (BLASTN search, *e*-value < 10^{-5}) between genomes. A–N designate the chromosomal segments, with strain-specific segments (D and G in *Xcc* ATCC 33913 and M and N in *Xcc* 8004) colored red. Black triangles indicate IS*1478*, the preferred recombination sites for genomic scale rearrangement. (*ori*) The origin of DNA replication. Genomic sequences of *Xcc* ATCC 33913 and *Xac* 306 are according to da Silva et al. (2002).

tion with host tissues (Craig et al. 2004). Genomic annotation revealed that at least 26 genes are related to type IV pili assembly and are highly conserved in *Xcc* 8004 and *Xcc* ATCC 33913 genomes. Mutations were identified in two of these pili assembly genes, *pilB* and *pilC*, which encode multidomain proteins driving energy-prone polymerization of pilin (Supplemental Table 3).

The majority component of the exopolysaccharides (EPSs) secreted by *Xcc* is xanthan gum, an anionic cellulosic (1–4) β-D-glucose polymer with trisaccharide side chains. EPS was demonstrated to be a virulence determinant in *Xcc*. They may mask the bacterium to prevent host recognition and enable colonization of host tissues (Alvarez 2000). The biosynthesis pathways of xanthan gum have been studied extensively (Becker et al. 1998). From the 16 annotated EPS biosynthesis-related genes (Table 4; Supplemental Table 3), we obtained three mutants, *rmlA*, *xanA*, and *gumK*, responsible for the biosynthesis of L-rhamnose, glucose/mannose 1-phosphate, and glucuronic acid-β-1,2-mannose-Cel-P-P-lipid, intermediates for xanthan formation, respectively (Becker et al. 1998).

Lipopolysaccharides (LPSs), also known as endotoxins, are a group of well-documented virulence determinates for animal pathogens (Lerouge and Vanderleyden 2002). Although LPS was less studied in phytopathogenic bacteria, it might play a wide range of roles during *Xcc* infection (Newman et al. 2002). We identified 10 mutants related to LPS biosynthesis (Table 4; Supplemental Table 3) with various biological functions, such as synthesis of O-antigen (*wxcA*, *wxcB*, *wxcC*, and *wxcD*), LPS core (*gmd*) and N-acetyl-3-amino hexose (*wxcM*) (Vorholter et al. 2001).

Two mutations were localized to a 5.2-kb gene cluster within chromosome segment C, which consists of three poorly characterized genes. XC3815 was annotated in the genome of *Xcc* ATCC 33913 and is thought to encode a conserved hypothetical protein (da Silva et al. 2002). However, its nonpathogenic mutant displayed small rough colonies, and further domain analysis indicated that it likely encodes an O-antigen polymerase. In contrast, the nonpathogenic mutant of XC3814, presumably encoding LPS core biosynthesis glycosyl transferase, forms normal colonies. Although there was no mutation found in XC3813, it was annotated as a gene encoding glycerophosphotransferase by domain analysis. The position of the XC3813–XC3815 gene cluster relative to the *wxc* gene cluster differs between *Xcc* ATCC 33913 and *Xcc* 8004. In *Xcc* ATCC 33913, the XC3813–3815 and *wxc* gene clusters are on opposite sides of the replication origin (*oriC*) (Vorholter et al. 2001), whereas in *Xcc* 8004, these two gene clusters are located on the same side of the *oriC*, and the XC3813–3815 cluster is 190 kb downstream of the *wxc* cluster.

In addition to the eight mutations mentioned above, three independent intergenic insertions (IG7~IG9, Supplemental Table 3) were identified upstream of the function-unknown XC3605 gene. This gene encodes a putative protein with an RgpF domain that is commonly found in rhamnose synthesis proteins. It thus might be related to rhamnose–glucose polysaccharide assembly and is eventually associated with O-antigen biosynthesis (Shibata et al. 2002).

The other identified mutant, *nagA*, which encodes N-acetylglucosamine-6-phosphate deacetylase, was associated with reduced virulence. It may catalyze the hydrolysis of the N-acetyl group of N-acetylglucosamine-6-P to yield glucosamine 6-phosphate and acetate, and therefore is also involved in cell-surface structure formation. It is interesting to note that *nagA* is associated with virulence in the animal mucosal pathogen *Candida albicans* (Kumar et al. 2000)

Detoxification

Once bacteria have successfully invaded host tissues, host-defensive chemicals and an oxidative burst are the major hurdles to be overcome (Van Sluys et al. 2002). The *Xcc* 8004 genome encodes nine glutathione S-transferase (GST), two glutathione peroxidase, and four catalases to detoxify these chemicals. In addition to mutations in genes encoding catalases (*katE* and *catB*) that may detoxify reactive oxygen species (ROS) during plant defensive processes, we found a mutant of *uptB*, which encodes maleylacetoacetate isomerase, an unusual ζ-class of GST that catalyzes the isomerization of C = C double bond in ring-cleavage products of aromatic compounds (Vuilleumier and Pagni 2002). This enzyme is likely to be associated with the inactivation of xenobiotic chemicals.

Unlike the xylem-limited phytopathogen *Xylella fastidiosa* (Simpson et al. 2000), *Xcc* genomes contain intact metabolic pathways for fatty acid oxidation. One mutated gene (*fadB*) encodes a *p*-hydroxycinnamoyl CoA hydratase/lyase, which catalyzes the hydration of 2-trans-enoyl-CoA into 3-hydroxyacyl-CoA, the second step in fatty acid β-oxidation. Another mutated gene (*cypC*) encodes fatty acid α hydroxylase, which contains a cytochrome P450-like domain and catalyzes the initial reaction in α-oxidation of fatty acid in the presence of H_2O_2 (Matsunaga et al. 1997). Mutations in lipid degradation genes leading to attenuation of virulence have been reported previously for animal pathogens, such as *Mycobacterium tuberculosis* and *Salmonella typhimurium* (Mahan et al. 1995). It may be that these pathways provide a protective mechanism for *Xcc* cells by degrading host toxic fatty acids, or alternatively, they may affect basic cellular activities in a manner that reduces in fitness.

Secretion systems

Both animal and plant pathogenic bacteria use secretion systems (SSs) to deliver extracellular enzymes and effectors to host cells. As revealed by whole-genomic sequences, *Xcc* is

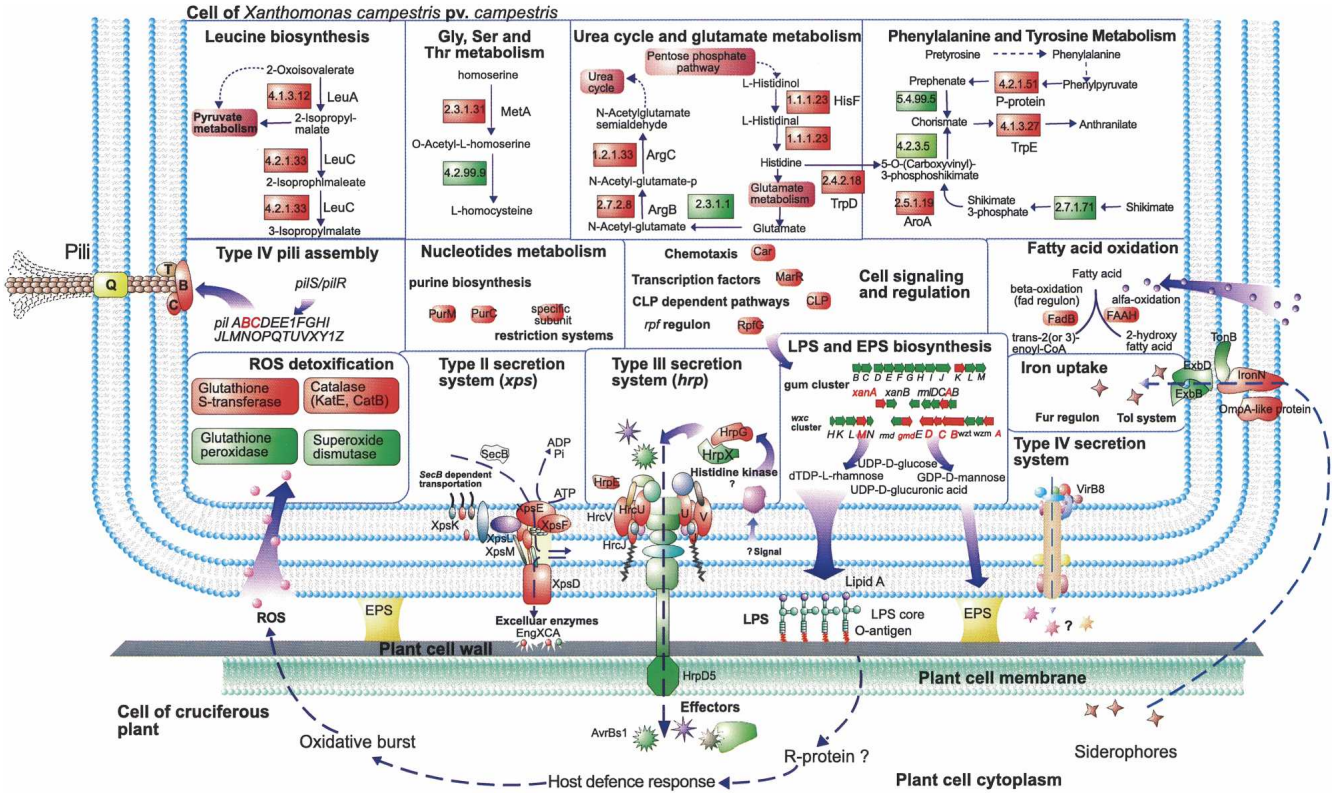

Figure 3. A schematic illustration of the experimentally determined interaction between *Xcc* and host cell. The principle pathways are shown in bold. Putative metabolic pathways were predicted by KEGG. Bacterial genes and gene products associated with pathogenicity are arranged within colored frames. Genes identified in this study are framed in red, while genes that were reported previously in xanthomonads are framed in green. (EPS) Exopolysaccharides; (FAAH) fatty acid α hydroxylase; (Hrp) hypersensitive response and pathogenicity genes; (LPS) lipopolysaccharides; (ROS) reactive oxygen species; (R-protein) resistance gene protein; (Xps) *Xanthomonas* protein secretion.

equipped with two sets of type II SS (*xps* and *xcs*). From 11 annotated genes related to the *xps* system, six pathogenicity-deficient mutants (*xpsD*, *xpsE*, *xpsF*, *xpsK*, *xpsL*, and *xpsM*) and an additional gene (*engXca*, encoding exocellular cellulase) were identified. However, no pathogenicity-deficient mutants were found among the 12 annotated genes related to the *xcs* system (Table 4). The significant difference in mutant identification rates of these two similar systems, even in light of the insufficient coverage and the biased distribution of the library, suggests that the *xcs* system may not play an essential role in *Xcc* pathogenesis.

Type III SS has been studied extensively in *X. campestris* pv. *vesicatoria*, the causative agent of bacterial spot in pepper and tomato (Buttner and Bonas 2003). Four mutants of genes encoding type III SS machinery (*hrcJ*, *hrcU*, *hrcV*, and *hrpE*) and one encoding a regulatory protein (*hrpG*), were associated with complete loss of virulence. Type IV SSs secrete macromolecules such as single-strand DNA in *A. tumefaciens* and pertussis toxin in *Bordetella pertussis* (Christie and Vogel 2000), respectively. We identified a mutated gene encoding the channel-forming protein VirB8, and thus, experimentally confirmed the prediction that type IV SS contributes to *Xcc* virulence (da Silva et al. 2002; Van Sluys et al. 2002).

Cellular metabolism

To cope with the iron-scarce environment of host tissues, *Xcc* evolved sophisticated iron transport systems, including the TonB-ExbB-ExbD, Tol-uptake system, ferric-uptake regulation protein (Fur), and at least 79 TonB-dependent outer-membrane ferric complex receptors. We identified mutants related to four TonB-dependent receptor genes that may participate in the uptake of iron-chelating compounds (siderophores) across the bacterial outer membrane, i.e., *ironN*, XC4223, and two intergenic mutations (IG3 and IG5, which inserted at the upstream regions of *bfeA* and *btuB*, respectively). XC4223 encodes an outer-membrane protein that contains complex domain structures, including TonB-dependent receptor, S-layer protein, collagen-binding, and hemaglutinin domains. The mutant of this gene totally eliminated *Xcc* virulence. However, IroN, BfeA, and ButB only contain a single TonB-dependent receptor domain and their mutants only showed reduced virulence. The variability in protein structure associated with virulence suggests that there are likely subtle functional differentiations among TonB-dependent receptors.

Many mutated genes are involved in general metabolism, including those that participate in the shikimate-chorismate pathway (*aroA*, *pheA*), leucine (*leuA*, *leuC*), histidine (*hisF*), arginine (*argB*, *argC*), methionine (*metA*), and tryptophan (*trpD*, *trpE*) biosynthesis, purine de novo biosynthesis (*purC* and *purD*) and gluconeogenesis (*ppsA*). Except for *ppsA*, these mutants are all auxotrophic as determined by no growth on MMX minimal medium, and thus, they might not reduce pathogenicity phenotype directly, but rather, may simply affect survival due to auxotrophy.

Table 4. A summary of pathogenicity-related genes and mutants identified in this study[a]

Functional category	No. genes annotated in the genome	No. genes identified from mutants	Ratio of genes identified vs. annotated
Cell mobility			
Type IV pili assembly	26	2	0.077
Flagellar assembly	33	0	0
Surface polysaccharides biosynthesis			
Exopolysaccharides	16	3	0.188
Lipopolysaccharides	57	10	0.158
Cell-wall degradation enzymes and secretion system (SS)			
Type II SS (*xps* system)	11	6	0.545
Type II SS (*xcs* system)	12	0	0
Cellulase	9	1	0.111
Other extracellular enzymes	15	0	0
Effectors and secretion system			
Type III SS	26	5	0.192
Effectors	8	0	0
Other secretion systems			
Type IV SS	17	1	0.059
Autotransporters	5	0	0
Detoxification and adaptation			
Glutathione *S*-transferase	9	1	0.111
Catalase	4	2	0.500
Glutathione peroxidase	4	0	0
Fatty acid degradation	13	2	0.154
Chemotaxis	46	1	0.022
Related general metabolisms			
TonB-dependent receptor	79	4[b]	0.051
Amino acid biosynthesis	115	10	0.087
Purine biosynthesis	22	3[b]	0.136
DNA restriction	18	1	0.056
Energy metabolism	210	2[b]	0.014
Regulatory element	ND	5	ND
Others[c]	ND	16	ND
Total		75	

[a]Details of genes identified from virulence-reduced mutants are listed in Supplemental Table 3.
[b]Signifies inclusion of intergenic mutations that might affect the transcription of functional genes.
[c]This category contains function-unknown genes (including *Xcc* 8004 strain-specific genes) and part of the intergenic mutations.

Cell signaling

Several regulatory and responsive genes were found to affect pathogenicity. In addition to the previously reported gene *clp* (de Crecy-Lagard et al. 1990), we identified a gene (*car*) that encodes a transducer car-like protein that may act as a chemosensor for arginine chemotaxis, as reported in *Halobacterium salinarum* (Storch et al. 1999). The XC2827 gene encodes a putative *marR* family transcriptional factor and possesses a helix-turn-helix domain. Insertion in XC2827 erased *Xcc* pathogenicity. A tail-specific protease mutant (*prc*) also had decreased pathogenicity. This protease was predicted to locate within the periplasmic space (PSORT certainty = 0.928) and contains a PDZ domain that may degrade proteins with a nonpolar C terminus and solicit cell signaling (Walsh et al. 2003). Although the *Xcc* 8004 genome encodes 106 two-component signal transduction genes, we identified only two previously reported response regulators (*hrpG* and *rpfG*).

Function-unknown and strain-specific genes

We obtained mutations in nine genes encoding function-unknown proteins (Table 4) that are involved in the pathogenicity of black rot disease. The presence of recognizable domains, such as helix-turn-helix (XC1021) and haemolysin-type calcium-binding domains (XC3922) in several of the gene products provides hints for further biochemical analysis.

In particular, three mutations that resulted in reduced virulence in cabbage were found to be located in two *Xcc* 8004-specific segments (Fig. 1; Table 1). The first gene, XC2068, does not have any recognizable domains, but its product has a significant degree of homology (*e*-value of BLASTP $<10^{-5}$) with proteins found in diverse pathogenic bacteria, including *Pseudomonas putida*, *Ralstonia metallidurans*, *Xylella fastidiosa*, *Burkholderia fungorum*, and *Yersinia pseudotuberculosis*, suggesting that it may be a virulence determinant in these bacteria. The second, XC2055, was matched significantly with two hypothetical proteins (BPP0494 and BPP0499) of the human pathogen *Bordetella bronchiseptica* (Parkhill et al. 2003). This protein contains a cell-signaling-related PadR domain, which suggests that it may be a transcriptional factor. The third, XC2416, encodes a hypothetical protein with a PH domain that has been associated with cell signaling. We inoculated five commercially available crucifers (Table 1) with these mutants and found that their host specificity was different from that of the wild-type strain *Xcc* 8004. These findings suggest that these proteins may be involved in host-specific pathogenesis, and highlight the direct correlation between genomic dynamics and virulence.

Discussion

The comparative genomics analysis was initiated by studying phylogenetically distant organisms; intraspecies comparison of multiple strains is becoming one of the new trends for studying bacterial pathogens (Whittam and Bumbaugh 2002). It can systematically reveal the genetic variations in genome organization, gene complement, and genetic message expression among different bacterial strains. Such an approach will not only provide insight into the molecular nature of virulence and host specificity, but will also advance our understanding of the evolutionary dynamics responsible for the emergence of novel infectious diseases.

Genomic variation between *Xcc* 8004 and *Xcc* ATCC 33913

We sequenced the complete genome of *Xcc* 8004 and compared it with that of *Xcc* ATCC 33913. Our results revealed that the two genomes are highly conserved with respect to gene content. The majority of the CDSs (3467) are actually identical at the nucleic-acid sequence level. Although 3548 SNPs are detected in 498 pairs of genes, we did not find any significant bias in mean response to selection force among four major different functional groups for those genes with SNPs, and thus, no strong evidence of relaxation of purifying selection or adaptive evolution for pathogenicity-associated genes was observed.

Large-scale genome reorganization, including deletion or acquisition of blocks of DNA segment and chromosomal rearrangements, were identified in both strains. One of the most

significant findings in this study is that of a dramatic genome rearrangement that occurred across the replication axis between two *Xcc* genomes (Fig. 2; Supplemental Fig. 2). This intraspecies event was greater in scale than the interspecies chromosomal rearrangements between *Xac* 306 and *Xcc* ATCC 33913. Such a rearrangement pattern, which results in an explicit X-shaped pattern by whole-genomic alignment, has been observed previously between distantly related species, such as between *H. pylori* and *Chlamydia jejuni*, *Vibrio cholera* and *Escherichia coli*, and *M. tuberculosis* and *M. leprae* (Eisen et al. 2000; Tillier and Collins 2000). An intraspecies rearrangement of this type was reported previously in a Gram-positive bacterium (*Streptococcus pyogenes*). In that case, the rearrangement probably occurred at homologous 5.6-kb-long sequences that included 16S RNA, tRNAs, 23S rRNA, and *comX1* homologs (Nakagawa et al. 2003). However, the present case is the first of its type recognized in such closely related strains in *Xanthomonas*. We prefer the replication-directed translocation mechanism (Tillier and Collins 2000) over the whole-genome inverted duplication (Eisen et al. 2000) and alternative transposition (Gray 2000) mechanisms to explain the X-shaped genomic rearrangement in *Xcc*, particularly in light of the recognition of two identical copies of IS*1478* elements in *Xcc* ATCC 33913. The two IS*1478* elements are located symmetrically at each extreme end of the rearrangement sites with two palindromic sequences immediately upstream (CGGCCCCACG CATGGGGCCG, 837 bp upstream of XCC0535 and GCAGAAG GTCCGACGGACCTTCTGC, 882 upstream of XCC3627). This configuration could form DNA hairpins and might facilitate the initiation of RecA-mediated homologous recombination during replication by cleavage with nuclease and introduction of DNA breaks, as suggested in *E. coli* (Leach et al. 1997).

There has been a long-standing and unsolved debate regarding the evolutionary impact of large-scale genomic rearrangements upon the global regulation of gene-expression patterns, which may affect the bacterial fitness or result in phenotypic difference, so that the altered bacteria could escape the attack from a host-immune system (Hughes 2000; Feil 2004). Recently, experimental evidences have accumulated to indicate that large-scale rearrangements could play a positive role during bacterial adaptation. Such rearrangements were found to be associated with strain diversification in *Pseudomonas aeruginosa* causing cystic fibrosis (Kresse et al. 2003), resource specialization in *E. coli* (Zhong et al. 2004), and biofilm formation/antibiotic resistance in *Staphylococcus epidermidis* (Kozitskaya et al. 2004). Because none of these rearrangements in the two *Xcc* strains directly broke up any predicted CDS or similar sets of genes, these events may reflect an early state in genome differentiation without obvious phenotypic alterations.

Under these circumstances, strain-specific genomic segments and strain-specific genes therein (insertion and deletion), which may or may not be associated with the large-scale rearrangement process, are of particular interest. As far as *Xcc* ATCC 33913 was concerned, the majority of strain-specific CDSs in its chromosomal segment G are phage related (Supplemental Table 1). In addition, although four of seven strain-specific CDSs located on the small *Xcc* ATCC 33913 segment D do encode an oxidoreductase, a DNA methylase, and two DNA helicases; no direct relationship with pathogenesis can be inferred. On the other hand, pathogenicity-related genes are found in *Xcc* 8004 strain-specific segments, especially those related to transportation. Later in our mutagenesis experiment, we further identified three genes that encode function-unknown proteins clearly essential for full virulence. Collectively, these results indicate that the genetic variation between two *Xcc* genomes represents a flexible set of "accessory genes" in closely related strains, but with a conserved "core" genome maintained in a species as the backbone (Mushegian and Koonin 1996). This pattern of genetic variation might contribute to rapid development of novel host specificity in pathogenic bacteria by altering virulent ability or cellular viability. Several studies of pathogenic bacteria, such as *Helicobacter pylori* (Alm et al. 1999), *P. aeruginosa* (Wolfgang et al. 2003), *Staphylococcus aureus* (Holden et al. 2004) and *X. fatidiosa* (Bhattacharyya et al. 2002) are consistent with this possibility.

Identification of black rot disease genes with mutagenesis approach

Transposon-based, large-scale mutagenesis approaches have been considered to be more efficient than gene expression-based techniques for high-throughput functional analysis (Winzeler et al. 1999; Geoffroy et al. 2003). These mutagenesis approaches were used to identify virulence factors in several pathogenic bacteria, including *H. pylori* (Kavermann et al. 2003), *Stahylococcus aureus* (Bae et al. 2004), and *Serratia marcescens* (Kurz et al. 2003); we reported here for the first time, using this approach to identify *Xcc* genes essential for black rot disease at the genomic scale.

Because of the relatively low library coverage, the strict screening standards, the specialized inoculation strategy, and the pathosystem used, a certain bias was observed in our data (Table 4). However, when comparing our results with genomic annotations, where about 250 genes in the *Xcc* genomes were predicted to be involved in pathogenicity (da Silva et al. 2002; this study), most of the expected pathways, such as EPS and LPS biosynthesis and detoxifying cascades, were identified by our approach, and data within each of these groups are much less biased than the general data set. Furthermore, we identified a cluster of genes that originally either had no functional assignment or are involved in unpredicted cellular pathways (e.g., fatty acids degradation, cell signaling, and general metabolism) and correlated them to the pathogenesis of the black rot disease. Our results suggest that the mutagenesis approach is advantageous for comprehensive investigation of bacterial pathogenesis.

Based on the available experimental results, we have proposed a molecular genetics model of the essential pathways in *Xcc* pathogenesis (Fig. 3). *Xcc* uses various degradative and regulatory mechanisms to parasitize its host, and most of the identified gene's functions are related to surface compounds or secreted compounds (Chan and Goodwin 1999; Alvarez 2000). For example, the *rpf* gene cluster (*rpf*A-I, for regulation of pathogenicity factors) positively regulates synthesis of extracellular enzymes, EPS or xanthan (Tang et al. 1991; Slater et al. 2000) and an adipose diffusible signal factor that is related to biofilm dispersal (Dow et al. 2003; Wang et al. 2004).

The novel virulence determinants identified in this study have opened a new avenue to further explore *Xcc* pathogenesis. For example, although the primary metabolic functions of most metabolic genes have been well studied with respect to normal bacterial physiology, little is known about the relationship between de novo nutrient biosynthesis and pathogenicity (Smith 1998). The present and previous results together indicate that many metabolic genes are involved in bacterial pathogenesis. For instance, purine biosynthesis pathways were identified in rela-

tion to pathogenesis in this study, as well as studies of *Brucella abortus* (Alcantara et al. 2004), *H. pylori* (Kavermann et al. 2003), and *Neisseria meningitides* (Sun et al. 2000).

Furthermore, certain metabolic intermediates have recently been implicated in important and varied roles in pathogenesis with respect to aggressiveness or cell-to-cell signaling in host tissues. These results indicated that before bacterial metabolic pathways within their host were fully understood, the roles of housekeeping genes remain difficult to explain. For example, host-derived lipoic acid was found to be critical for the growth and virulence of the intracytosolic pathogen *Listeria monocytogenes* (O'Riordan et al. 2003); a cluster of sulfur metabolism genes (e.g., *raxR*, *raxQ*) are required for plant resistance gene *Xa21*-mediated host recognition against *X. oryzae* pv. *oryzae*, the causative agent of rice bacterial blight (Burdman et al. 2004). In the present study, aromatic amino-acid biosynthesis genes, e.g., *trpD*, *trpE*, and *aroA* were found to be associated with *Xcc* pathogenesis (Supplemental Table 3). These genes are involved in the biosynthesis of anthranilate, which is a down-stream product of the shikimate-chorismate pathway (AroA) and is synthesized by anthranilate synthase (TrpD, TrpE, two subunits). It is worth mentioning that anthranilate is the precursor of intercellular signal molecule 2-heptyl-3-hydroxy-4-quinolone, which is involved in quorum sensing in *P. aeruginosa* (Pesci et al. 1999).

Conclusions

In conclusion, we report the complete genome sequence of *Xcc* 8004 and our observations of the various kinds of genetic variation that exist between *Xcc* 8004 and *Xcc* ATCC 33913. Our findings indicate that strain-specific genes, focused on genomic rearrangement-associated strain-specific segments, are likely to be the major driving force for strain differentiation. Furthermore, our high-throughput mutagenesis analysis method identified a wealth of genes that contribute to *Xcc* pathogenicity. These results provide new insight into the relationship between host specificity and *Xcc* genome evolution. Moreover, this work generated a mutant profile as a starting point from which to investigate the molecular pathogenesis of black rot disease.

Methods

Genome sequencing and annotation

Genomic DNA was sheared by sonication. The blunt-ends were created by mung bean nuclease, and 1.5–3.0 kb fragments were cloned into pUC18 digested with SmaI. The 8–10-kb inserts generated by Sau3AI/MboI were ligated into pUC18 digested with BamHI. Approximately 55,000 clones were sequenced from both ends on ABI3700 or MegaBACE DNA autosequencers. Sequence assembly was accomplished by Phred/Phrap/Consed. A total of 107,860 sequences (10-fold genome coverage) served as scaffolding, and gaps were filled by primer walking, subcloning, or multiplex PCR. The sequence assembly was confirmed by pulsed-field gel electrophoresis and Southern blotting analysis with PacI and PmeI. Putative CDSs were identified with GLIMMER. CDS functions were predicted by searching the database with BLAST, FASTA, and Pfam. Metabolic pathways were examined on KEGG. Transfer RNAs were predicted by TRNASCAN-SE.

SNP analysis

SNPs between the two *Xcc* genomes were identified by BLASTN ($e < 10^{-5}$). Synonymous and nonsynonymous sites were determined by ClustalW. The number of nonsynonymous SNPs (A), synonymous SNPs (S), the value of average SNP/codon (ratio of total SNPs vs. total number of codons), and the ratios of total nonsynonymous SNPs vs. total synonymous SNPs (A/S) were calculated. A/S differences among the gene groups were detected by χ^2 tests.

Mutant library construction and plant inoculation

The large-scale transposon-insertional mutant library used in this study was constructed as described previously (Sun et al. 2003). The library contains 16,512 clones and was generated by transforming the EZ::TN <KAN-2> Tnp transposome (Epicentre) into *Xcc* electro-competent cells. The saturation of this library was deduced as 97% (or 4× genome coverage) by its randomicity as determined by Southern blotting and flanking sequence analysis (Sun et al. 2003).

For plant inoculation, transformants were grown in PSA medium (peptone 10 g/L, sucrose 10 g/L, glutamic acid 1 g/L) with kanamycin (25 µg/mL) at 28°C, shaking at 200 rpm until the OD_{600} value reached 0.3–0.5. Four-week-old cabbages (*B. oleracae* cultivar Jingfeng 1) were inoculated by clipping leaf tips with sterile scissors dipped in bacteria cultures. After inoculation, the plants were kept in a greenhouse at 28~35°C, with a relative humidity >95%. Lesion length was scored 10 d after inoculation, and virulence level was scored as follows: 0, no visible effect; 1, limited chlorosis around the cut site; 2, chlorosis extending from the cut site; 3, blackened leaf veins, death, and drying of tissue within the chlorotic area; 4, extensive vein blackening, death, and drying of tissue (Dow et al. 1990). In the initial round of screening, the virulence level of transformants (a total of 16,512) was determined individually from a minimum of 12 innoculation sites distributed over at least three leaves on a minimum of two different plant individuals. Only transformants with significant decrease in virulence were selected as pathogenicity-deficient candidates and were confirmed further in at least four independent rounds of inoculation (at lease 24 innoculation sites for each transformant distributed over at least three leaves on a minimum of four different plant individuals). For host range determination, the confirmed pathogenicity-deficient mutants were inoculated onto different host plants, including two radish cultivars (*R. sativus* cv. Huaye and *R. sativus* cv. Xiaojingzhong), Chinese cabbage (*B. chinensis* cv. Zhongbai 5) and pakchoi (*B. chinensis* cv. Wuyueman; all of these cultivars are available from the Institute of Vegetables and Flowers, Chinese Academy of Agricultural Sciences, Beijing 100081). Wild-type strains of *Xcc* 8004, *Xcc* ATCC 33913, and sterile water were used as controls.

EZ::TN transposon copy number determination

Genomic DNA of *Xcc* was extracted from *Xcc* and digested with PstI, which has no target site on the probe, but has one site at the 3′ extreme end of EZ::TN transposon. The kanamycin resistance gene of the EZ::TN transposon was amplified with primers Kan-1 and Kan-2 (Supplemental Table 4) and labeled with [α-^{32}P]dCTP by using the Prime-a-Gene labeling system (Promega).

Flanking sequence analysis

Genomic sequences flanking the EZ::TN transposon in mutants with reduced pathogenicity were analyzed using thermal asymmetric interlaced polymerase chains reactions (TAIL-PCR) with three nested sequence-specific (SP) primers complementary to the transposon sequence and one arbitrary degenerate (AD) primer (Liu and Whittier 1995). The primer sequences are listed in Supplemental Table 4 and the TAIL-PCR procedure is summa-

rized in Supplemental Table 5. TAIL-PCR products were cloned into the pGEM-T easy vector (Promega), sequenced, and mapped on the genome of *Xcc* 8004.

Acknowledgments

We thank Professors X. Tang (Kansas State University), L Zhang (Institute of Molecular and Cell Biology, Singapore), R.L. Somerville (Purdue University), and C-I. Wu (Chicago University) for comments on the manuscript. Ms. Y.C. Fu helped to create the animated flash. This work was funded by the High-Tech Program (Grant 101-02-07-02) of Ministry of Science and Technology of China, the National Natural Science Foundation of China (Grants 30228002 and 30270027), and the Chinese Academy of Sciences (Grant KSCX2-SW-314 and "Bairen Jihua" to C.H.).

References

Alcantara, R.B., Read, R.D., Valderas, M.W., Brown, T.D., and Roop, R.M. 2004. Intact purine biosynthesis pathways are required for wild-type virulence of *Brucella abortus* 2308 in the BALB/c mouse model. *Infect. Immun.* **72:** 4911–4917.

Alm, R.A., Ling, L.S., Moir, D.T., King, B.L., Brown, E.D., Doig, P.C., Smith, D.R., Noonan, B., Guild, B.C., deJonge, B.L., et al. 1999. Genomic-sequence comparison of two unrelated isolates of the human gastric pathogen *Helicobacter pylori*. *Nature* **397:** 176–180.

Alvarez, A.M. 2000. Black rot of crucifers. In *Mechanisms of resistance to plant diseases* (eds. A.J. Slusarenko et al.), pp. 21–52. Kluwer Academic Publications, Dordrecht, Netherlands.

Bae, T., Banger, A.K., Wallace, A., Glass, E.M., Aslund, F., Schneewind, O., and Missiakas, D.M. 2004. *Staphylococcus aureus* virulence genes identified by *bursa aurealis* mutagenesis and nematode killing. *Proc. Natl. Acad. Sci.* **101:** 12312–12317.

Becker, A., Katzen, F., Puhler, A., and Ielpi, L. 1998. Xanthan gum biosynthesis and application: A biochemical/genetic perspective. *Appl. Microbiol. Biotechnol.* **50:** 145–152.

Bhattacharyya, A., Stilwagen, S., Ivanova, N., D'Souza, M., Bernal, A., Lykidis, A., Kapatral, V., Anderson, I., Larsen, N., Los, T., et al. 2002. Whole-genome comparative analysis of three phytopathogenic *Xylella fastidiosa* strains. *Proc. Natl. Acad. Sci.* **99:** 12403–12408.

Burdman, S., Shen, Y., Lee, S.W., Xue, Q., and Ronald, P. 2004. RaxH/RaxR: A two-component regulatory system in *Xanthomonas oryzae* pv. *oryzae* required for AvrXa21 activity. *Mol. Plant Microbe Interact.* **17:** 602–612.

Buttner, D. and Bonas, U. 2003. Common infection strategies of plant and animal pathogenic bacteria. *Curr. Opin. Plant Biol.* **6:** 312–319.

Cascales, E. and Christie, P.J. 2003. The versatile bacterial type IV secretion systems. *Nat. Rev. Microbiol.* **1:** 137–149.

Chan, J.W. and Goodwin, P.H. 1999. The molecular genetics of virulence of *Xanthomonas campestris*. *Biotechnol. Adv.* **17:** 489–508.

Christie, P.J. and Vogel, J.P. 2000. Bacterial type IV secretion: Conjugation systems adapted to deliver effector molecules to host cells. *Trends Microbiol.* **8:** 354–360.

Craig, L., Pique, M.E., and Tainer, J.A. 2004. Type IV pilus structure and bacterial pathogenicity. *Nat. Rev. Microbiol.* **2:** 363–378.

da Silva, A.C., Ferro, J.A., Reinach, F.C., Farah, C.S., Furlan, L.R., Quaggio, R.B., Monteiro-Vitorello, C.B., Van Sluys, M.A., Almeida, N.F., Alves, L.M., et al. 2002. Comparison of the genomes of two *Xanthomonas pathogens* with differing host specificities. *Nature* **417:** 459–463.

de Crecy-Lagard, V., Glaser, P., Lejeune, P., Sismeiro, O., Barber, C.E., Daniels, M.J., and Danchin, A. 1990. A *Xanthomonas campestris* pv. *campestris* protein similar to catabolite activation factor is involved in regulation of phytopathogenicity. *J. Bacteriol.* **172:** 5877–5883.

Dow, J.M., Clarke, B.R., Milligan, D.E., Tang, J.L., and Daniels, M.J. 1990. Extracellular proteases from *Xanthomonas campestris* pv. *campestris*, the black rot pathogen. *Appl. Environ. Microbiol.* **56:** 2994–2998.

Dow, J.M., Crossman, L., Findlay, K., He, Y.Q., Feng, J.X., and Tang, J.L. 2003. Biofilm dispersal in *Xanthomonas campestris* is controlled by cell-cell signaling and is required for full virulence to plants. *Proc. Natl. Acad. Sci.* **100:** 10995–11000.

Dye, D., Bradbury, W., Goto, J.F., Hayward, M., Lelliott, A.C., and Schroth, M.N. 1980. International standards for naming pathovars of phytopathogenic bacteria and a list of pathovar names and pathotype strains. *Rev. Plant Pathol.* **59:** 153–168.

Eisen, J.A., Heidelberg, J.F., White, O., and Salzberg, S.L. 2000. Evidence for symmetric chromosomal inversions around the replication origin in bacteria. *Genome Biol.* **1:** research0011.

Feil, E.J. 2004. Small change: Keeping pace with microevolution. *Nat. Rev. Microbiol.* **2:** 483–495.

Geoffroy, M.C., Floquet, S., Metais, A., Nassif, X., and Pelicic, V. 2003. Large-scale analysis of the meningococcus genome by gene disruption: Resistance to complement-mediated lysis. *Genome Res.* **13:** 391–398.

Gray, Y.H. 2000. It takes two transposons to tango: Transposable-element-mediated chromosomal rearrangements. *Trends Genet.* **16:** 461–468.

Holden, M.T., Feil, E.J., Lindsay, J.A., Peacock, S.J., Day, N.P., Enright, M.C., Foster, T.J., Moore, C.E., Hurst, L., Atkin, R., et al. 2004. Complete genomes of two clinical *Staphylococcus aureus* strains: Evidence for the rapid evolution of virulence and drug resistance. *Proc. Natl. Acad. Sci.* **101:** 9786–9791.

Hughes, D. 2000. Evaluating genome dynamics: The constraints on rearrangements within bacterial genomes. *Genome Biol.* **1:** reviews0006.

Israel, D.A., Salama, N., Krishna, U., Rieger, U.M., Atherton, J.C., Falkow, S., and Peek, R.M. 2001. *Helicobacter pylori* genetic diversity within the gastric niche of a single human host. *Proc. Natl. Acad. Sci.* **98:** 14625–14630.

Kavermann, H., Burns, B.P., Angermuller, K., Odenbreit, S., Fischer, W., Melchers, K., and Haas, R. 2003. Identification and characterization of *Helicobacter pylori* genes essential for gastric colonization. *J. Exp. Med.* **197:** 813–822.

Kozitskaya, S., Cho, S.H., Dietrich, K., Marre, R., Naber, K., and Ziebuhr, W. 2004. The bacterial insertion sequence element IS*256* occurs preferentially in nosocomial *Staphylococcus epidermidis* isolates: Association with biofilm formation and resistance to aminoglycosides. *Infect. Immun.* **72:** 1210–1215.

Kresse, A.U., Dinesh, S.D., Larbig, K., and Romling, U. 2003. Impact of large chromosomal inversions on the adaptation and evolution of *Pseudomonas aeruginosa* chronically colonizing cystic fibrosis lungs. *Mol. Microbiol.* **47:** 145–158.

Kumar, M.J., Jamaluddin, M.S., Natarajan, K., Kaur, D., and Datta, A. 2000. The inducible N-acetylglucosamine catabolic pathway gene cluster in *Candida albicans*: Discrete N-acetylglucosamine-inducible factors interact at the promoter of NAG1. *Proc. Natl. Acad. Sci.* **97:** 14218–14223.

Kurz, C.L., Chauvet, S., Andres, E., Aurouze, M., Vallet, I., Michel, G.P., Uh, M., Celli, J., Filloux, A., De Bentzmann, S., et al. 2003. Virulence factors of the human opportunistic pathogen *Serratia marcescens* identified by in vivo screening. *EMBO J.* **22:** 1451–1460.

Leach, D.R., Okely, E.A., and Pinder, D.J. 1997. Repair by recombination of DNA containing a palindromic sequence. *Mol. Microbiol.* **26:** 597–606.

Lerouge, I. and Vanderleyden, J. 2002. O-antigen structural variation: Mechanisms and possible roles in animal/plant-microbe interactions. *FEMS Microbiol. Rev.* **26:** 17–47.

Lin, N.T., Wen, F.S., and Tseng, Y.H. 1996. A region of the filamentous phage ϕ *Lf* genome that can support autonomous replication and miniphage production. *Biochem. Biophys. Res. Commun.* **218:** 12–16.

Liu, Y.G. and Whittier, R.F. 1995. Thermal asymmetric interlaced PCR: Automatable amplification and sequencing of insert end fragments from P1 and YAC clones for chromosome walking. *Genomics* **25:** 674–681.

Mahan, M.J., Tobias, J.W., Slauch, J.M., Hanna, P.C., Collier, R.J., and Mekalanos, J.J. 1995. Antibiotic-based selection for bacterial genes that are specifically induced during infection of a host. *Proc. Natl. Acad. Sci.* **92:** 669–673.

Matsunaga, I., Yokotani, N., Gotoh, O., Kusunose, E., Yamada, M., and Ichihara, K. 1997. Molecular cloning and expression of fatty acid α-hydroxylase from *Sphingomonas paucimobilis*. *J. Biol. Chem.* **272:** 23592–23596.

Mushegian, A.R. and Koonin, E.V. 1996. A minimal gene set for cellular life derived by comparison of complete bacterial genomes. *Proc. Natl. Acad. Sci.* **93:** 10268–10273.

Nakagawa, I., Kurokawa, K., Yamashita, A., Nakata, M., Tomiyasu, Y., Okahashi, N., Kawabata, S., Yamazaki, K., Shiba, T., Yasunaga, T., et al. 2003. Genome sequence of an M3 strain of *Streptococcus pyogenes* reveals a large-scale genomic rearrangement in invasive strains and new insights into phage evolution. *Genome Res.* **13:** 1042–1055.

Newman, M.A., von Roepenack-Lahaye, E., Parr, A., Daniels, M.J., and Dow, J.M. 2002. Prior exposure to lipopolysaccharide potentiates expression of plant defenses in response to bacteria. *Plant J.* **29:** 487–495.

O'Riordan, M., Moors, M.A., and Portnoy, D.A. 2003. Listeria intracellular growth and virulence require host-derived lipoic acid.

Science **302:** 462–464.
Parkhill, J., Sebaihia, M., Preston, A., Murphy, L.D., Thomson, N., Harris, D.E., Holden, M.T., Churcher, C.M., Bentley, S.D., Mungall, K.L., et al. 2003. Comparative analysis of the genome sequences of *Bordetella pertussis*, *Bordetella parapertussis* and *Bordetella bronchiseptica*. *Nat. Genet.* **35:** 32–40.
Pesci, E.C., Milbank, J.B., Pearson, J.P., McKnight, S., Kende, A.S., Greenberg, E.P., and Iglewski, B.H. 1999. Quinolone signaling in the cell-to-cell communication system of *Pseudomonas aeruginosa*. *Proc. Natl. Acad. Sci.* **96:** 11229–11234.
Shibata, Y., Yamashita, Y., Ozaki, K., Nakano, Y., and Koga, T. 2002. Expression and characterization of streptococcal *rgp* genes required for rhamnan synthesis in *Escherichia coli*. *Infect. Immun.* **70:** 2891–2898.
Simpson, A.J., Reinach, F.C., Arruda, P., Abreu, F.A., Acencio, M., Alvarenga, R., Alves, L.M., Araya, J.E., Baia, G.S., Baptista, C.S., et al. 2000. The genome sequence of the plant pathogen *Xylella fastidiosa*. *Nature* **406:** 151–157.
Slater, H., Alvarez-Morales, A., Barber, C.E., Daniels, M.J., and Dow, J.M. 2000. A two-component system involving an HD-GYP domain protein links cell-cell signalling to pathogenicity gene expression in *Xanthomonas campestris*. *Mol. Microbiol.* **38:** 986–1003.
Smith, H. 1998. What happens to bacterial pathogens in vivo? *Trends Microbiol.* **6:** 239–243.
Storch, K.F., Rudolph, J., and Oesterhelt, D. 1999. Car: A cytoplasmic sensor responsible for arginine chemotaxis in the archaeon *Halobacterium salinarum*. *EMBO J.* **18:** 1146–1158.
Sun, Y.H., Bakshi, S., Chalmers, R., and Tang, C.M. 2000. Functional genomics of *Neisseria meningitidis* pathogenesis. *Nat. Med.* **6:** 1269–1273.
Sun, Q., Wu, W., Qian, W., Hu, J., Fang, R., and He, C. 2003. High-quality mutant libraries of *Xanthomonas oryzae* pv. *oryzae* and *X. campestris* pv. *campestris* generated by an efficient transposon mutagenesis system. *FEMS Microbiol. Lett.* **226:** 145–150.
Swings, J.G. and Civerolo, E.L. 1993. *Xanthomonas*. Chapman & Hall, London.
Tang, J.L., Liu, Y.N., Barber, C.E., Dow, J.M., Wootton, J.C., and Daniels, M.J. 1991. Genetic and molecular analysis of a cluster of *rpf* genes involved in positive regulation of synthesis of extracellular enzymes and polysaccharide in *Xanthomonas campestris* pathovar *campestris*. *Mol. Gen. Genet.* **226:** 409–417.
Tillier, E.R. and Collins, R.A. 2000. Genome rearrangement by replication-directed translocation. *Nat. Genet.* **26:** 195–197.
Van Sluys, M.A., Monteiro-Vitorello, C.B., Camargo, L.E., Menck, C.F., da Silva, A.C., Ferro, J.A., Oliveira, M.C., Setubal, J.C., Kitajima, J.P., and Simpson, A.J. 2002. Comparative genomic analysis of plant-associated bacteria. *Annu. Rev. Phytopathol.* **40:** 169–189.
Vorholter, F.J., Niehaus, K., and Puhler, A. 2001. Lipopolysaccharide biosynthesis in *Xanthomonas campestris* pv. *campestris*: A cluster of 15 genes is involved in the biosynthesis of the LPS O-antigen and the LPS core. *Mol. Genet. Genomics* **266:** 79–95.
Vuilleumier, S. and Pagni, M. 2002. The elusive roles of bacterial glutathione S-transferases: New lessons from genomes. *Appl. Microbiol. Biotechnol.* **58:** 138–146.
Walsh, N.P., Alba, B.M., Bose, B., Gross, C.A., and Sauer, R.T. 2003. OMP peptide signals initiate the envelope-stress response by activating DegS protease via relief of inhibition mediated by its PDZ domain. *Cell* **113:** 61–71.
Wang, L.H., He, Y., Gao, Y., Wu, J.E., Dong, Y.H., He, C., Wang, S.X., Weng, L.X., Xu, J.L., Tay, L., et al. 2004. A bacterial cell-cell communication signal with cross-kingdom structural analogues. *Mol. Microbiol.* **51:** 903–912.
Whittam, T.S. and Bumbaugh, A.C. 2002. Inferences from whole-genome sequences of bacterial pathogens. *Curr. Opin. Genet. Dev.* **12:** 719–725.
Wilson, T.J., Bertrand, N., Tang, J.L., Feng, J.X., Pan, M.Q., Barber, C.E., Dow, J.M., and Daniels, M.J. 1998. The *rpfA* gene of *Xanthomonas campestris* pathovar *campestris*, which is involved in the regulation of pathogenicity factor production, encodes an aconitase. *Mol. Microbiol.* **28:** 961–970.
Winzeler, E.A., Shoemaker, D.D., Astromoff, A., Liang, H., Anderson, K., Andre, B., Bangham, R., Benito, R., Boeke, J.D., Bussey, H., et al. 1999. Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. *Science* **285:** 901–906.
Wolfgang, M.C., Kulasekara, B.R., Liang, X., Boyd, D., Wu, K., Yang, Q., Miyada, C.G., and Lory, S. 2003. Conservation of genome content and virulence determinants among clinical and environmental isolates of *Pseudomonas aeruginosa*. *Proc. Natl. Acad. Sci.* **100:** 8484–8489.
Wren, B.W. 2000. Microbial genome analysis: Insights into virulence, host adaptation and evolution. *Nat. Rev. Genet.* **1:** 30–39.
Zhong, S., Khodursky, A., Dykhuizen, D.E., and Dean, A.M. 2004. Evolutionary genomics of ecological specialization. *Proc. Natl. Acad. Sci.* **101:** 11719–11724.

Received October 26, 2004; accepted in revised form April 15, 2005.