Population genomics and pathotypic evaluation of the bacterial leaf blight pathogen of rice

Jinshui Zheng[a,1], Zhiwei Song[b,1], Dehong Zheng[a], Huifeng Hu[a], Haoquan Yu[a], Hongxia Liu[c], Yancun Zhao[b], Ming Sun[a], Lifang Ruan[a,2], Fengquan Liu[b,2]

[a] State Key Laboratory of Agricultural Microbiology, Hubei Key Laboratory of Agricultural Bioinformatics, Huazhong Agricultural University, Wuhan, 430070, China.

[b] Institute of Plant Protection, Jiangsu Academy of Agricultural Sciences, Nanjing 210014, China.

[c] Department of Plant Pathology, College of Plant Protection, Nanjing Agricultural University, Nanjing 210095, China

[1] These authors contributed equally to this work.

[2] Corresponding authors：Fengquan Liu (fqliu20011@sina.com) and Lifang Ruan (ruanlifang@mail.hzau.edu.cn).

Keywords: Plant pathogen, Genomic epidemiology, Large-scale virulence test, Rapid virulence dynamics

Abstract

Knowledge about population structure and virulence dynamics is critical to prevent outbreaks of bacterial diseases. *Xanthomonas oryzae* pv. *oryzae* (*Xoo*) causes bacterial blight disease, leading to substantial losses in rice production during the last century in Asia. For this concern, we used whole-genome sequences and large-scale virulence tests to explore the diversity and evolution of *Xoo* population during the past 30 years in the main rice-planting areas of China. Six separate lineages were revealed by phylogenomic analysis, two of which (CX-5 and CX-6) predominated in the population. Four and 5 sub-lineages from these two lineages respectively persisted in different areas for decades. Many recent sporadic outbreaks were caused by *Xoo* derived from the local sub-populations of these sub-lineages. The lineage and sub-lineage distribution of isolates strongly correlated to their geographical origin, which was found to be determined by the planting of the two major rice subspecies. Large-scale virulence tests indicated rapid dynamics of pathogenicity for *Xoo*. The *Xoo* isolates from lineages and sub-lineages with stricter geographical distribution showed less compatibility with resistance genes of rice than the wider distributed ones. Genetic background of *Xoo*, rice resistance genes and the planting environment of rice contributed to the rapid dynamics of virulence. This study provided a good model to understand the evolution and dynamics of plant pathogens in the context of interaction with their hosts which are influenced by both geographical conditions and farming practices.

Significance

Xanthomonas oryzae pv. *oryzae* (*Xoo*) which causes bacterial blight disease, was one of the major serious threats to food security in East Asian countries during last century. The introduction of several resistance genes into rice tremendously reduced the damage. Through population genomics and large-scale pathotypic evaluation of more than 240 *Xoo* strains isolated in China during the past 30 years, we found that the genome evolution and virulence dynamics of *Xoo* are extraordinary rapid, indicating its ability to overcome the resistance conferred by the current resistance genes and the potential for large-scale outbreaks in the future. It is therefore prudent to continue surveillance of disease outbreaks caused by this pathogen, and to develop novel strategies for its control.

53 **Introduction**

54 Population genomics is a powerful tool for understanding the formation and evolution
55 of bacterial pathogens of human and some important domestic animals (1-6). It has also
56 provided new strategies for the detection, prevention and control of important human
57 pathogens (7, 8). Although plant pathogens are serious threat to food security
58 worldwide, only a few studies have used genome-based methods to study their
59 population structure, evolution and transmission in the context of interaction between
60 pathogens and plant hosts (9-13).

61 *Xanthomonas oryzae* pv. *oryzae* (*Xoo*) causes bacterial blight (BB) disease of rice
62 (*Oryza sativa*) and has been considered one of the top 10 plant bacterial pathogens
63 based on scientific/economic importance (14). It is a notorious destructive pathogen
64 which can cause a considerable reduction in rice production in both temperate and
65 tropical regions, especially in Asia (15). On average, BB can lead to 20%~30% overall
66 yield reduction with some severe cases even causing up to 50% loss (16). As the world's
67 largest producer and consumer of rice, China has suffered serious BB outbreaks and
68 huge food loss since the 1930s (17), with the most serious damage occurring between
69 1950s and 1980s (18). Breeding resistant cultivars has been proved to be the most
70 effective method in controlling this pathogen. Since 1980s, BB-resistant rice varieties
71 have been widely planted in China and have significantly reduced the loss caused by
72 BB (19-21). However, sporadic outbreaks have occurred in different areas of China
73 after the year of 2000 (22, 23). It is unknown whether the recent outbreaks relate to
74 changes in the genetics and virulence of the past *Xoo* populations in China, allowing it

to overcome host defenses. Moreover, very little is known about the evolution, spread and dynamics of this important plant pathogen involved in different outbreaks during the past decades around the whole country in the context of rapid development of modern agricultural science and technology.

The virulence differentiation among different strains of *Xoo* was studied by defining different races or pathotypes. Strains of the same race share a common pathogenic phenotype in a set of tested host cultivars. Rice lines carrying different resistance genes (R genes) determine the race of strains of *Xoo*. Up to now, more than 100 races of *Xoo* were identified (24-29). Although the pathogenic diversity of *Xoo* in China have been studied previously (22, 30, 31), these studies classified *Xoo* isolates only based on their pathogenic patterns on several rice lines, while the genetic basis for the pathogenicity of these isolates is still unknown.

In this study, a comprehensive population genomic study was combined with a large-scale determination of virulence of *Xoo* that have been isolated in the past 30 years in China, to obtain a framework of genetic dynamics and virulence diversity over time and space. Results from this work shed new light on the evolution of this important plant pathogen.

Results

Population structure *Xoo* from China

To understand the population structure of *Xoo* in China, genomes of the 247 strains were sequenced, including 237 isolated from China, 4 and 6 representative ones from

97 Japan and Philippines, respectively (Table S1). Reads were mapped to the genome of
98 PXO99^{A}, SNPs were called by Samtools. After the prophage region, repeat sequences
99 and recombinant regions were filtered; the core SNPs contained 5,271 variable sites.
100 Two methods were used to outline the population structure based on the core SNPs,
101 Maximum likelihood (ML) phylogenetic analysis and a tree-independent hierarchical
102 Bayesian clustering (BAPS).
103 All *Xoo* from China were clustered into 6 lineages (CX-1 to CX-6), with more than 70%
104 strains belonging to CX-5 and CX-6 (Figure 1). The three major rice production areas,
105 including South China, Yangtze Valley area and North China, were displayed in Figure
106 2a. We annotated the lineage-specific tree with the time and space information of the
107 isolated *Xoo* (Figure 1, 2b and 2c). CX-1 and CX-2 were most frequently represented
108 in South China at the year of 2003, of which most isolates were from Yunnan province.
109 The *Xoo* belonging to CX-3 were isolated from North China in 1984, 2003 and 2014,
110 with most of which from Northeast China in 2014. CX-5 and CX-6 were nationally
111 distributed during the past 30 years; while most members in CX-5 were dominant in
112 South China and Yangtze Valley, and CX-6 were more frequently from North China
113 and Yangtze Valley.

114 **Phylogeography of CX-5 and CX-6**

115 To investigate more details about evolution and dynamics of *Xoo* in China, we focused
116 on the two major lineages, CX-5 and CX-6. Phylogenetic analysis based on core SNP
117 alignment (Figures 3a and 3b) and pairwise SNP distance between isolates was
118 performed for both lineages (Figure 3c). Though no correlation between root-to-tip

branch lengths and the known years of isolation of the sequenced was observed for both lineages, it's worth mentioning that some important points about the dynamics of these population were illuminated, especially those for the recent outbreaks.

Four sub-lineages were identified in CX-5, CX-5.1 to CX-5.5 (Figure 3a). CX-5.1, CX-5.2, and CX-5.5 were restricted in South China, while CX-5.3 and CX-5.4 can be found in some places of both South China and Yangtze Valley. When time was considered, we found that all these sub-lineages persisted at these places at least for decades. Then we focused on the recent sporadic outbreaks being attributed to this lineage. Besides clustered together on the tree, we considered epidemiologically and genomically linked outbreak with an average SNP pairwise distance less than 5. In 2003, the outbreak in Hunan province of Yangtze Valley was caused by CX-5.4 (HN-2003), while the outbreak in Guangdong province of South China ascribed to two sub-population of CX-5.5 (GD-2003-1 and GD-2003-2). One cluster of CX-5.3 caused outbreak in Guangxi province of South China in 2014 (GX-2014).

CX-6 was divided into 5 sub-lineages, CX-6.1 to CX-6.5 (Figure 3b). Seven out of 10 isolates in CX-6.1 were from different places of South China. Actually, this sub-lineage was the most diverse one in CX-6 with long branches. In CX-6.2, *Xoo* from North China and Yangtze Valley were dominant on different branches, respectively. CX-6.3 and CX-6.4 were most frequently found in Yangtze Valley, while *Xoo* from North China predominated in CX-6.5. Similar as CX-5, all these sub-lineages have contributed to BB on rice for a long time in China. These sub-lineages totally caused at less 4 local outbreaks in 2003 and 2014. In 2003, the local outbreaks in two provinces of Yangtze

141 Valley (Henan, and Jiangsu) were caused by CX-6.2 (HeN-2003) and CX-6.4 (JS-2003),
142 respectively, while the one in North China (Jilin province) was ascribed to CX-6.5 (JL-
143 2003). Outbreak in Jiangsu province (Yangtze Valley) in 2014 was caused by one sub-
144 population of CX-6.4 (JS-2014).

145 **Chinese *Xoo* diversity and dynamics in the context of rice host**

146 There are two major domesticated rice subspecies *Oryza sativa japonica* and *indica*
147 existing in different areas of China; and most of rice varieties were derived from these
148 two subspecies (32, 33). Rice planted in high-altitude areas of Yunnan province from
149 South China and North China belong to subspecies *japonica* (http://www.ricedata.cn).
150 *Xoo* isolated from these areas were respectively clustered into 3 lineages, CX-1, CX-2,
151 CX-3 and one sub-lineage of CX-6 (CX-6.5) (Figure 1 and Figure 3b). Lineage CX-5
152 mainly contained *Xoo* isolated from rice in South China where subspecies *indica* is
153 planted (Figure 3a). CX-6 also contained *Xoo* isolated from rice of *japonica* from
154 Yangtze Valley (CX-6.4), *indica* from Yangtze Valley (CX-6.3) and some *indica* from
155 South China (CX-6.1), indicating the widest adaptation to different rice varieties of *Xoo*
156 in this lineage (Figure 3b). In CX-6.2, old *Xoo* were dominant on indica, while recent
157 isolated ones were almost from *japonica*, suggesting a putative host switch in this sub-
158 lineage.

159 Taken together, the distribution of Chinese *Xoo* lineages appeared to be impaired by
160 both biogeography and rapid dynamics over isolated time, which may be mainly due to
161 the distribution and dynamics of the two major subspecies of rice, *japonica* and *indica*.

162 **Diversity of *Xoo* from different counties of Asia**

To place *Xoo* from China into a global context, phylogenomic analysis was performed on the core genome SNPs of 109 *Xoo* genomes available in Genbank (including 100 from India, 8 from Philippines and 1 from Japan), and the 247 sequenced ones in this study. Phylogenetic analysis revealed a similar topological structure of Asian *Xoo* to that of Chinese *Xoo*, with some small differences in some branches (Figure S1). Eleven genetic lineages were identified for the *Xoo* population from Asia. *Xoo* from India and Philippines were respectively assigned into PX-A to PX-C and IX-I to IX-V according to previous studies (34, 35). Eight lineages displayed strict geographic distribution features, most of them being specific to one or at least a limited amount of rice-planting areas. Besides the three Chinese lineages, CX-1, CX-2 and CX-3, PX-B and PX-C were mainly represented in Philippines, and IX-I to IX-III were from India. CX-4 contained old *Xoo* from China and some recently isolated ones from India belonging to IX-V. For CX-5 and CX-6, besides both represented the major *Xoo* from China, they also contained a few *Xoo* from Japan, India and Philippines, with PX-A of Philippines being nested within the former one and IX-IV from India being encompassed among the members of the last one, indicating frequently transmission of these 2 lineages not only among different places of China but also among different Asian countries.

Rapid virulence dynamics of *Xoo* strains isolated in China during the past 30 years

To determine the pathogenic diversity of *Xoo* strains, virulence assays were conducted on six rice lines. Based on the interactions between *Xoo* strains and rice lines, most of the tested *Xoo* were classified into 9 pathogenic races (Table S1and Figure 4a). The top two races with most members, R5 and R8, contain 26% and 18% of the isolates,

respectively.

The three major rice-planting regions had different race compositions (Figure 4b). Yangtze River Valley mainly contained *Xoo* from races R5 and R8, with the respective percentage of the total population being 50% and 30%, respectively. *Xoo* isolates from North China mostly belongs to R7 and R5, together accounts for about two third of the total population. Five of the nine *Xoo* races were well represented in the *Xoo* population from South China, with each of them account for more than 10% of the total population. The virulence dynamics of *Xoo* during the last 30 years was hypothesized by analyzing race distribution in China at different period (Figure 4c). Strains isolated before the year of 2000 were mainly of the races of R5, R4, R8 and R1, while most *Xoo* isolated in 2003–2004 were R8, R7, R5 and R6; and about 45% of the isolates in 2014 were R5. We also found different lineages have different race compositions, indicating genetic background playing important roles in the interaction between *Xoo* and rice hosts (Figure 4d). Compared to other lineages, the CX-6 *Xoo* isolates fall mostly within the races of R5, R8 and R4, while have reduced representation in R1, R2 and R6. Moreover, the rapid dynamics of pathogenic features can be inferred by focusing on the races of *Xoo* isolated from the same places at the same time and belonging to the same lineage (Figure 1). For example, most *Xoo* from Yunnan, one province of South China, in 2003 were clustered in CX-1 and CX-2. The 19 isolates in CX-1 represented 8 of the 9 races described above, and the 14 strains in CX-2 belonged to 6 races. Even *Xoo* isolates from the same outbreak were allocated into different races, with 16 *Xoo* in JL-2003 belonging to 3 different races and 10 isolates in GD-2003-2 representing 2 races.

207 **Resistance gene of rice contributes to the interaction between *Xoo* and rice**

208 As each of the near-isogenic rice lines used in this study contains one known resistance
209 (R) gene, we studied the interaction between *Xoo* and rice host by focusing on the
210 capacity of *Xoo* to overcome the resistances conferred by these genes (Figure 1, Tables
211 1 and 2). When the time of isolation was considered, we found that *Xoo* from different
212 times have different capacity to overcome different R genes (Table 1). R gene *xa5*
213 showed the most effective resistance against *Xoo* all the time. At different periods, more
214 than 90% *Xoo* could not infect the rice line IRBB5, which contained gene *xa5*. The
215 resistance mediated by R gene *Xa3* was effective during the period of 1990 to 2004.
216 However, rice containing this R gene only showed resistance against 5% of *Xoo* in 2014.
217 Rice cultivars carrying R gene *Xa4* are resistant to more than 60% of *Xoo* isolated from
218 1970s to 2000 and 2014. However, they are only resistant to about 30% of the isolates
219 recovered in 2003 and 2004.

220 To determine whether the *Xoo* genetic background contributes to the interaction
221 between *Xoo* and rice lines, we studied the resistance capacity of R genes against *Xoo*
222 from different lineages (Table 2). We found that resistance provided by R genes against
223 *Xoo* are indeed depended on the genetic background of *Xoo*. R genes *Xa2* and *Xa14*
224 effectively confer resistance against *Xoo* from Lineages CX-1, CX-2 and CX-3,
225 preventing infection by more than 70% of *Xoo* strains from these lineages. Similarly, R
226 gene *Xa3* and *Xa4* showed potential resistance against *Xoo* from CX-1 and CX-2, and
227 CX-2 and CX-3, respectively. No R gene except *xa5* conferred resistance to more than
228 50% of *Xoo* strains from CX-5 and CX-6, which is consistent with the situation that

229 CX-5 and CX-6 had much more extensive distribution across different areas of China
230 during the past decades, comparing to the other Chinese *Xoo* lineages. Among the sub-
231 lineages of CX-5 and CX-6, *Xa2* was only effective against *Xoo* of CX-6.5, with
232 effectiveness of 90%, while *Xa4* could confer the resistance against more than 60% *Xoo*
233 from these two lineages except for CX-5.1 and CX-6.5; *Xa14* showed potential medium
234 level of resistance against *Xoo* from CX-6.1 and CX-6.5 (Table S2).
235

236 **Discussion**

237 RFLP (Restriction Fragment Length Polymorphism) was one of the successfully used
238 molecular techniques for diversity studies of plant pathogens before genomic methods
239 emerging (36-38). It was frequently used for genetic diversity of *Xoo* in several
240 important rice-planting countries (28, 39-44). These studies had provided some new
241 insights into the evolution and dynamics of *Xoo* population; however, many crucial
242 details were overlooked due to the intrinsic low sensitivity and stability of the typing
243 method. As the *Xoo* population structure were studied overly dependent on the marker
244 genes chosen, the definite genetic relationships among *Xoo* isolates from the same and
245 different outbreaks, the particular evolutionary history of the different clusters and
246 different isolates, and the sophisticated connection between genetic diversity and
247 virulence dynamics cannot be will well elaborated through these probe-dependent
248 methods. Genome-based population genetics methods have been proved more powerful
249 to study the diversity, evolution and dynamics of many pathogens of animals and plants
250 (1-6, 9-13). Previous report focusing on *Xoo* from India through population genomics

251 defined 5 lineages, with each of them showing a restricted-region distribution (35).
252 Here, we focused on genomic epidemiology of *Xoo* and provided some novel insights
253 into the diversity, evolution and dynamics of this notorious plant pathogen during the
254 past 30 year in the main rice-planting areas of China. We also found that the diversity
255 and dynamics of virulence features of this pathogen population are affected by the
256 population structure, genetic background, environmental factors and human activities.
257 These findings will benefit for development of new strategies and methods for
258 surveillance and controlling this destructive rice pathogen.
259 Some *Xoo* clusters found by traditional molecular typing methods from several Asian
260 countries showed strict geographical distribution, but others were widely distributed in
261 these countries (28, 39, 42). In this study, based on population genomics, we found that
262 there is a general correspondence between the areas of *Xoo* isolated and most Asian
263 lineages and some sub-lineages in CX-5 and CX-6 (Figures 1, 3 and S2). Moreover,
264 Asian *Xoo* was reported to be very different from those of Africa and North America
265 regions (45, 46). This suggests that the geographical factors as affected by climate, soil
266 types, or the choice of different rice cultivars and varieties may influence on the
267 diversity of *Xoo*. Information of the *Xoo* host subspecies in China indicates that variety
268 of rice was one of the major factors forces on the formation and dynamics of Chinese
269 *Xoo* population. Rice varieties planting in different counties and different areas of a
270 country are determined by geographical environment, climate factors and local
271 agricultural policy.
272 Though resistance breeding of rice had effectively prevented the unprecedented

273 outbreak of BB since 1990s, sporadic outbreaks occurred in different areas of China
274 after the year 2000 (22, 23). Genomic epidemiology analysis indicates that the recent
275 outbreaks in different places were contributed by different lineages or sub-lineages,
276 even some outbreaks from the same places caused by different lineages. There was no
277 special lineage or sub-lineage spreading nationwide, partially because of the great
278 diversity of rice varieties planting in China (http://www.ricedata.cn). All *Xoo* strains
279 causing the recent epidemic outbreaks except those in CX-1 were derived from the ones
280 leading to heavy losses in the history. On the other hand, only the minor lineage CX-4
281 was not represented by the recent outbreaks, so the diversity of the *Xoo* population was
282 not reduced due to wide planting of BB-resistant rice varieties. For the current practice
283 of rice production, the major BB threat in the three rice-planting areas of China is still
284 mainly from CX-5 and CX-6 as they were present in history, which call for different
285 detection and prevention strategies for different areas and rice varieties.
286 The virulence diversity of *Xoo* in China against rice was well investigated during the
287 past 30 years, however, the relationships between population genetic structure and
288 pathotype of *Xoo* were not explored well (22, 30, 31, 44, 47). All the tested *Xoo* strains
289 in this study were classified into nine major pathogenic races (Figure 4a). Races of *Xoo*
290 from South China were more diverse than those from North China and Yangtze Valley
291 areas. In line with this, phylogenetic analysis revealed that *Xoo* from South China are
292 more diverse compared to other places of China (Figure 1). Some rice growing regions
293 in South China, particularly in the Yunnan province, have three harvests per year and
294 use different rice cultivars for each harvest (48, 49). The greater genetic and virulence

295 diversity of *Xoo* thus could be caused by a greater diversity of rice cultivars in use. Our
296 result also showed that there were diverse races in the same lineage, for example, 85%
297 *Xoo* from lineage CX-6 could be assigned to 8 races (Figure 4d), indicating rapid
298 virulence dynamics in the same lineage. Moreover, the isolates from the same local
299 outbreak even belonged to different races. Although *Xoo* strains isolated in China were
300 consistently assigned to one of the 6 Chinese lineages, the dominating population of
301 *Xoo* races in specific areas displayed rapid dynamics and high level of diversity during
302 the past 30 years (Tables 1 and 2). Rapid virulence dynamics with relative stability of
303 the standing genetic variability of *Xoo* population indicates fast adaptation to host
304 during interaction for this plant pathogen.
305 Earlier isolates were more susceptible to resistances conferred by rice resistance genes
306 than strains isolated in later years after the resistance genes had been introduced to
307 control BB (39, 43). In China, since the 1980s, *indica* hybrid rice (especially Shanyou
308 63) had been planted on a large scale (50). Later, bacterial blight resistant rice varieties,
309 mainly the *india* rice carrying *Xa4* and *japonica* rice carrying *Xa3*, were grown on a
310 large scale for a long time (51). Soon after these rice varieties widely used for rice
311 production, the effectiveness of these R genes conferring resistance against infection
312 causing by *Xoo* was significantly compromised (Table 1). Actually, the resistance
313 abilities of these two R genes were also overcome frequently by *Xoo* population in other
314 Asian countries (40, 42, 43). Long-term cultivation of rice varieties carrying a single
315 resistance gene can be overcome by *Xoo*, and breeding rice varieties with multiple
316 resistance loci can provide prolonged resistance (52-54). R gene *xa5* showed very good

317 resistance (above 95%) to *Xoo* from all lineages. Previously, isolates of *Xoo* having
318 compatibility with *xa5* were only reported from Indian Lineages IX-I and IX-III, and
319 some Philipps and Korea isolates (35, 55-57). The only 2 Chinese *Xoo* from Yunnan
320 province (YN24 and YN04-5) belonging to IX-I and IX-III showed good infection
321 ability to rice with *xa5*, suggesting putative transmission of *Xoo* from India to South
322 China. Besides, several isolates from CX-5 and CX-6, especially those obtained
323 recently, can overcome the resistance contributed by *xa5*, indicating different origins of
324 these resistance compatibilities (Tables 1 and 2). Since rice variants with gene *xa5* were
325 less used in China, the development of its compatibility could not be due to a direct
326 selection pressure of this R gene.
327 The genetic background of *Xoo* is crucial for the interaction between *Xoo* and its host.
328 Isolates from different lineages and sub-lineages showed distinct capabilities to fight
329 against the same and different R genes. Overall, lineages and sub-lineages with stricter
330 geographical distribution can be more easily defeated by R genes than the wider
331 distributed ones (Tables 2 and S2). The underlying genetic determinations of these
332 differences call for advanced comparative genomics and wet experiments. However,
333 there is no lineage or sub-lineage with all isolates incompatible with one of the tested
334 R genes, which means that *Xoo* with different genetic backgrounds can adapt to
335 different rice hosts with variable resistances. This suggests that the introduction of new
336 R genes could reduce the loss causing by *Xoo*, but can't change the genetic diversity of
337 the *Xoo* population, as no lineage or sub-lineage can be eliminated thoroughly. This
338 partially because R genes usually restrict the pathogens to the initial infection site, but

339 have no killing abilities (58).

340

341 **Materials and Methods**

342 **Genome sequencing**

343 *Xoo* strains from 3 countries were used in this study (Table S1). The major strains were
344 isolated from China. And the represented *Xoo* strains for 4 different races from Japan
345 and 7 different races from Philippines were also sequenced.

346 The bacterial DNA of overnight cultures (30 °C) was extracted and purified using the
347 Easy-DNA kit (Invitrogen, USA) following the manufacturer's protocol. Total DNA
348 was sequenced by with Illumina HiSeq 2000, 2500, or 4000 to produce pair-end reads
349 with lengths of 100, 125 and 150 bp. For each *Xoo*, raw reads were assessed with the
350 FastQC tool (https://github.com/s-andrews/FastQC) and quality filtered using
351 Trimmomatic (https://github.com/timflutre/trimmomatic). The filtered reads were
352 error-corrected by library with Quake to produce clean reads (59).

353 **Mapping and SNP calling**

354 To study the population structure of *Xoo* from China, reads from each strain sequenced
355 was mapped onto the reference genome PXO99^{A}, which is used as a model strain for
356 studying the interaction between *Xoo* and rice host and is the first *Xoo* with complete
357 genome sequence (60), using BWA (61). Variant detection was performed using
358 Samtools mpileup (https://github.com/samtools/samtools) and filtered with a minimum
359 mapping quality of 30. SNP was excluded if its coverage was less than 10% or more
360 than 200% of the average coverage, if it was not supported by at least 5 reads on each

361 strand. Phage regions and repetitive sequences of the PXO99^{A} genome were predicted
362 by PHASTER (http://phaster.ca) and RepeatScout (https://bix.ucsd.edu/repeatscout),
363 respectively. Snippy (https://github.com/tseemann/snippy) was used to generate the
364 whole genome alignment of all studied *Xoo*, and then the recombination was detected
365 by Gubbins (https://sanger-pathogens.github.io/gubbins) on the core genome alignment
366 after prophage and repeat regions were filtered. SNPs located within phage regions,
367 repetitive sequences or recombinant regions were excluded.

368 **Assembly and whole genome alignment**

369 To study the population structure of Asian *Xoo*, genomes from different locations were
370 downloaded from the Genbank as follows: 100 from India, 9 from Philippines, 1 from
371 South Korea and 1 from Japan (34, 35). We performed genome assembly from all the
372 strains sequenced in this study. For each strain, the pair-end clean reads were assembled
373 using Spades (62). After few evaluations, different sets of k-mer values were chosen
374 for Spades according to different read lengths. The final assemblies were obtained by
375 filtering out contigs with few reads supported or with lengths lower than 200 bp (Table
376 S1). Whole genome alignment was carried out on all the genomes obtained from public
377 database and assembled here by Parsnp (63) using the genome of PXO99^{A} as reference.
378 Core SNPs were extracted from the core genome alignment by snp-sites
379 (https://github.com/sanger-pathogens/snp-sites). Recombination of the core genome
380 alignment was inferred by Gubbins. The final SNPs were obtained by filtering out the
381 prophage region and repeat sequences mentioned above, and the recombined regions
382 predicted here.

383 **Phylogenetic analysis**

384 The SNPs-based maximum likelihood (ML) phylogenies were built by RAxML with
385 the generalized time-reversible model and a Gamma distribution to model site-specific
386 rate variation (64). Bootstrap support values were calculated from 500 replicates. *Xoo*
387 population structure of China was defined with the hierBAPS module of the BAPS
388 software, which delineates the population structure by nested clustering (65). Three
389 independent iterations with upper population sizes of 5, 10, and 15 were used to obtain
390 optimal clustering of the population. Phylogenetic trees were annotated and visualized
391 by iTOL (https://itol.embl.de/).

392 **Virulence evaluation of *Xoo* strains**

393 Virulence of *Xoo* was assesses by inoculating 6 near-isogenic rice lines, each carrying
394 a specific R gene, IRBB2 (*Xa2*), IRBB3 (*Xa3*), IRBB4 (*Xa4*), IRBB5 (*xa5*), IRBB14
395 (*Xa14*) and IR24 (*Xa18*). IR24 was used as a susceptible check. Seeds of all lines in
396 this study were obtained from the China National Rice Research Institute (CNRRI).
397 Three weeks after sowing on a seedbed, seedlings were uprooted and transplanted into
398 the field. The management of the rice plants in the field was proceeded as usual.
399 Bacterial strain was cultivated on NA plates at 28 °C for 48 h. Then the bacterial
400 colonies were suspended in sterile ddH_2O with concentration adjusted to 3×10^8 cfu/ml
401 before inoculation. The leaf-clipping method was used to test the virulence of all the
402 *Xoo* strains (66). Fifteen top fully expanded leaves per rice plant were inoculated with
403 each strain. The lesions length was measured after 21 days post inoculation. The ratio
404 of the lesion length compared to the whole leaf length was calculated as described

405 previously (31). When the ratio of the rice line was lower than 1/4, the rice line was
406 classified as resistant (R); when the ratios were between 1/4 and 1, the rice line was
407 rated as susceptible (S). The lesion length 14 days following inoculation was also
408 measured and used to compare with the data obtained 7 days later.

409 **Data access**

410 Sequence data generated for this study have been submitted to the NCBI Sequence Read
411 Archive under BioProject accession PRJNA350904.

412

413 **Acknowledgments**

414 This work was supported by National Key Research and Development Program of
415 China (2017YFD0201100), National Natural Science Foundation of China (NSFC)
416 (31770003, 31571974 and 31801697), Fundamental Research Funds for the Central
417 University (2662018JC015 and 2662016QD039), Science and Technology Department
418 of Hubei Province (2017CFB438) and Special Fund for Agro-scientific Research in the
419 Public Interest (201303015). Michael Gänzle and Jens Walter are acknowledged for
420 critical feedback on the manuscript.

421

422 **Conflict of Interest**

423 The authors declare no conflict of interest.

424

425 **Reference**

426 1. Didelot X, Walker AS, Peto TE, Crook DW, & Wilson DJ (2016) Within-host evolution of
427 bacterial pathogens. *Nat. Rev. Microbiol.* 14(3):150-162.

2. Baker S, Thomson N, Weill FX, & Holt KE (2018) Genomic insights into the emergence and spread of antimicrobial-resistant bacterial pathogens. *Science* 360(6390):733-738.
3. Feasey NA*, et al.* (2016) Distinct *Salmonella Enteritidis* lineages associated with enterocolitis in high-income settings and invasive disease in low-income settings. *Nat. Genet.* 48(10):1211-1217.
4. Okoro CK*, et al.* (2012) Intracontinental spread of human invasive *Salmonella Typhimurium* pathovariants in sub-Saharan Africa. *Nat. Genet.* 44(11):1215-1221.
5. Weill FX*, et al.* (2017) Genomic history of the seventh pandemic of cholera in Africa. *Science* 358(6364): 785-789.
6. Coll F*, et al.* (2018) Genome-wide analysis of multi- and extensively drug-resistant *Mycobacterium tuberculosis*. *Nat. Genet.* 50(2):307-316.
7. Gardy JL & Loman NJ (2018) Towards a genomics-informed, real-time, global pathogen surveillance system. *Nat. Rev. Genet.* 19(1):9-20.
8. Quainoo S*, et al.* (2017) Whole-Genome Sequencing of Bacterial Pathogens: the Future of Nosocomial Outbreak Analysis. *Clin. Microbiol. Rev.* 30(4):1015-1063.
9. Zhong Z*, et al.* (2018) Population genomic analysis of the rice blast fungus reveals specific events associated with expansion of three main clades. *ISME J* 12(8):1867-1878.
10. Baltrus DA, McCann HC, & Guttman DS (2017) Evolution, genomics and epidemiology of *Pseudomonas syringae*: Challenges in Bacterial Molecular Plant Pathology. *Mol. Plant Pathol.* 18(1):152-168.
11. Brader G*, et al.* (2017) Ecology and Genomic Insights into Plant-Pathogenic and Plant-Nonpathogenic Endophytes. *Annu. Rev. Phytopathol.* 55:61-83.
12. Hartmann FE & Croll D (2017) Distinct Trajectories of Massive Recent Gene Gains and Losses in Populations of a Microbial Eukaryotic Pathogen. *Mol. Biol. Evol.* 34(11):2808-2822.
13. Hartmann FE, Sanchez-Vallet A, McDonald BA, & Croll D (2017) A fungal wheat pathogen evolved host specialization by extensive chromosomal rearrangements. *ISME J* 11(5):1189-1204.
14. Mansfield J*, et al.* (2012) Top 10 plant pathogenic bacteria in molecular plant pathology. *Mol. Plant Pathol.* 13(6):614-629.
15. Mew T, Alvarez A, Leach J, & Swings J (1993) Focus on bacterial blight of rice. *Plant Dis.* 77:5-12.
16. Mew T (1987) Current status and future prospects of research on bacterial blight of rice. *Annu. Rev. Phytopathol.* 25:359-382.
17. Fang Z (1963) *Rice Bacterial Blight* (Jiangsu People Press).
18. Zhang Q (2009) Genetics and Improvement of Bacterial Blight Resistance of Hybrid Rice in China. *Rice Science* 16(2):83-92.
19. Pradhan SK*, et al.* (2015) Pyramiding of three bacterial blight resistance genes for broad-spectrum resistance in deepwater rice variety, Jalmagna. *Rice (N Y)* 8(1):51.
20. Kottapalli KR, Lakshmi Narasu M, & Jena KK (2010) Effective strategy for pyramiding three bacterial blight resistance genes into fine grain rice cultivar, Samba Mahsuri, using sequence tagged site markers. *Biotechnol. Lett.* 32(7):989-996.
21. Huang N*, et al.* (1999) Pyramiding of bacterial blight resistance genes in rice: marker-assisted selection using RFLP and PCR. *Theor. Appl. Genet.* 95(3):313-320.
22. Liu H, Yang W, Hu B, & Liu F (2007) Virulence analysis and race classification of *Xanthomonas*

oryzae pv. *oryzae* in China. *J. Phytopathol.* 155(3):129-135.

23. Chen XL*, et al.* (2012) Elevational Variation in Diversity of *Xanthomonas oryzae* pv. *oryzae* in South-West China. *J. Phytopathol.* 160(6):261-268.

24. NODA T, YAMAMOTO T, KAKU H, & HORINO O (1996) Geographical Distribution of Pathogenic Races of *Xanthomonas oryzae* pv. *oryzae* in Japan in 1991 and 1993. *Ann Phytopathol Soc Jpn* 62:549–553.

25. Liu H, Yang W, Hu B, & Liu F (2007) Virulence Analysis and Race Classification of *Xanthomonas oryzae* pv. *oryzae* in China. *J Phytopathol* 155(3):129-135.

26. Li G, Song C-F, Pang X-M, Yang Y, & Wang J-S (2008) Analysis of Pathotypic and Genotypic Diversity of *Xanthomonas oryzae* pv. *oryzae* in China. *J Phytopathol* 154(4):208-218.

27. Jeung JU, Heu SG, Shin MS, Vera Cruz CM, & Jena KK (2006) Dynamics of *Xanthomonas oryzae* pv. *oryzae* Populations in Korea and Their Relationship to Known Bacterial Blight Resistance Genes. *Phytopathol.* 96(8):867-875.

28. Adhikari TB, Mew TW, & Leach JE (1999) Genotypic and pathotypic diversity in *Xanthomonas oryzae* pv. *oryzae* in Nepal. *Phytopathol.* 89(8):687-694.

29. Mishra D*, et al.* (2013) Pathotype and genetic diversity amongst Indian isolates of *Xanthomonas oryzae* pv. *oryzae*. *PLoS One* 8(11):e81996.

30. Li G, Song CF, Pang XM, Yang Y, & Wang JS (2009) Analysis of Pathotypic and Genotypic Diversity of *Xanthomonas oryzae* pv. *oryzae* in China. *J. Phytopathol.* 157(4):208-218.

31. Fang Z*, et al.* (1990) Studies on pathotypes of *Xanthomonas campestris* pv. *oryzae* in China. *Acta Phytopathol Sin* 20:81-88.

32. Huang X*, et al.* (2012) A map of rice genome variation reveals the origin of cultivated rice. *Nature* 490(7421):497-501.

33. Kovach MJ, Sweeney MT, & McCouch SR (2007) New insights into the history of rice domestication. *Trends Genet.* 23(11):578-587.

34. Quibod IL*, et al.* (2016) Effector Diversification Contributes to *Xanthomonas oryzae* pv. *oryzae* Phenotypic Adaptation in a Semi-Isolated Environment. *Sci. Rep.* 6:34137..

35. Midha S*, et al.* (2017) Population genomic insights into variation and evolution of *Xanthomonas oryzae* pv. *oryzae*. *Sci. Rep.* 7:40694.

36. Tzeng TH, Lyngholm LK, Ford CF, & Bronson CR (1992) A Restriction-Fragment-Length-Polymorphism Map and Electrophoretic Karyotype of the Fungal Maize Pathogen Cochliobolus-Heterostrophus. *Genetics* 130(1):81-96.

37. Hulbert SH*, et al.* (1988) Genetic-Analysis of the Fungus, Bremia-Lactucae, Using Restriction Fragment Length Polymorphisms. *Genetics* 120(4):947-958.

38. Lazo GR, Roffey R, & Gabriel DW (1987) Pathovars of *Xanthomonas-Campestris* Are Distinguishable by Restriction-Fragment-Length-Polymorphism. *Int. J. Syst. Bacteriol.* 37(3):214-221.

39. Mishra D*, et al.* (2013) Pathotype and Genetic Diversity amongst Indian Isolates of *Xanthomonas oryzae* pv. *oryzae*. *PLoS One* 8(11).

40. Leach JE*, et al.* (1992) Assessment of Genetic Diversity and Population-Structure of *Xanthomonas-Oryzae* Pv *Oryzae* with a Repetitive DNA Element. *Appl. Environ. Microbiol.* 58(7):2188-2195.

41. George MLC, Bustamam M, Cruz WT, Leach JE, & Nelson RJ (1997) Movement of *Xanthomonas oryzae* pv *oryzae* in southeast Asia detected using PCR-based DNA fingerprinting.

Phytopathol. 87(3):302-309.

42. Adhikari TB, *et al.* (1995) Genetic Diversity of *Xanthomonas-Oryzae* Pv *Oryzae* in Asia. *Appl. Environ. Microbiol.* 61(3):966-971.

43. Nelson RJ, *et al.* (1994) Relationship between Phylogeny and Pathotype for the Bacterial-Blight Pathogen of Rice. *Appl. Environ. Microbiol.* 60(9):3275-3283.

44. Zhang Q, Leach JE, Nelson RJ, & Mew TW (1997) Genetic Structure of Rice Bacterial Blight Pathogen Population in China. *Acta Agronomica Sinica* 23(02):150-158.

45. Triplett LR, *et al.* (2011) Genomic analysis of *Xanthomonas oryzae* isolates from rice grown in the United States reveals substantial divergence from known *X. oryzae* pathovars. *Appl. Environ. Microbiol.* 77(12):3930-3937.

46. Gonzalez C, *et al.* (2007) Molecular and pathotypic characterization of new *Xanthomonas oryzae* strains from West Africa. *Mol. Plant. Microbe Interact.* 20(5):534-546.

47. Yang SQ, *et al.* (2013) Molecular and pathogenic characterization of new *Xanthomonas oryzae* pv. *oryzae* strains from the coastline region of Fangchenggang city in China. *World J. Microbiol. Biotechnol.* 29(4):713-720.

48. Tu M, Lu BR, Zhu Y, & Wang Y (2007) Abundant within-varietal genetic diversity in rice germplasm from Yunnan Province of China revealed by SSR fingerprints. *Biochem. Genet.* 45(11-12):789-801.

49. Xiong ZY, *et al.* (2010) Differentiation and distribution of indica and japonica rice varieties along the altitude gradients in Yunnan Province of China as revealed by InDel molecular markers. *Genet. Resour. Crop Evol.* 57(6):891-902.

50. Chen S, Lin XH, Xu CG, & Zhang QF (2000) Improvement of bacterial blight resistance of 'Minghui 63', an elite restorer line of hybrid rice, by molecular marker-assisted selection. *Crop Sci.* 40(1):239-244.

51. Xu Z, *et al.* (2004) Race monitoring of rice bacterial blight (*Xanthomonas oryzae* pv. *oryzae*) in China. *Chin J Rice Sci* 18:469-472.

52. Joseph M, *et al.* (2004) Combining bacterial blight resistance and Basmati quality characteristics by phenotypic and molecular marker-assisted selection in rice. *Mol. Breed.* 13(4):377-387.

53. Sundaram RM, *et al.* (2008) Marker assisted introgression of bacterial blight resistance in Samba Mahsuri, an elite indica rice variety. *Euphytica* 160(3):411-422.

54. Basavaraj SH, *et al.* (2010) Marker-assisted improvement of bacterial blight resistance in parental lines of Pusa RH10, a superfine grain aromatic rice hybrid. *Mol. Breed.* 26(2):293-305.

55. Carpenter SCD, *et al.* (2018) A Strain of an Emerging Indian *Xanthomonas oryzae* pv. *oryzae* Pathotype Defeats the Rice Bacterial Blight Resistance Gene xa13 Without Inducing a Clade III SWEET Gene and Is Nearly Identical to a Recent Thai Isolate. *Front. Microbiol.* 9:2703.

56. Huang S, *et al.* (2016) The broadly effective recessive resistance gene *xa5* of rice is a virulence effector-dependent quantitative trait for bacterial blight. *Plant J.* 86(2):186-194.

57. Jeung JU, Heu SG, Shin MS, Cruz CMV, & Jena KK (2006) Dynamics of *Xanthomonas oryzae* pv. *oryzae* populations in Korea and their relationship to known bacterial blight resistance genes. *Phytopathol.* 96(8):867-875.

58. Shen YW & Ronald P (2002) Molecular determinants of disease and resistance in interactions of *Xanthomonas oryzae* pv. *oryzae* and rice. *Microb. Infect.* 4(13):1361-1367.

59. Kelley DR, Schatz MC, & Salzberg SL (2010) Quake: quality-aware detection and correction

of sequencing errors. *Genome Biol.* 11(11):R116.

60. Salzberg SL*, et al.* (2008) Genome sequence and rapid evolution of the rice pathogen *Xanthomonas oryzae* pv. *oryzae* PXO99A. *BMC Genomics* 9:204.

61. Li H & Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25(14):1754-1760.

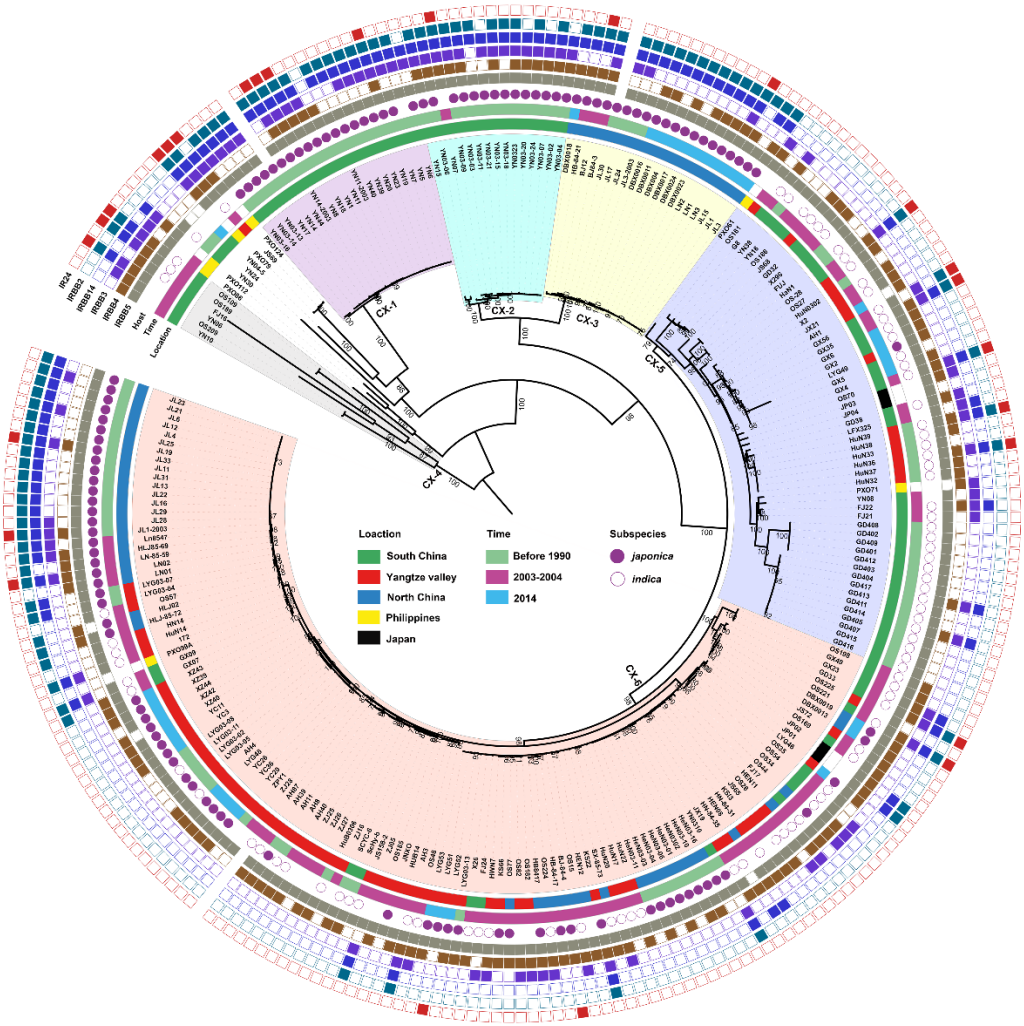
62. Bankevich A*, et al.* (2012) SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *J. Comput. Biol.* 19(5):455-477.

63. Treangen TJ, Ondov BD, Koren S, & Phillippy AM (2014) The Harvest suite for rapid core-genome alignment and visualization of thousands of intraspecific microbial genomes. *Genome Biol.* 15(11):524.

64. Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30(9):1312-1313.

65. Cheng L, Connor TR, Siren J, Aanensen DM, & Corander J (2013) Hierarchical and Spatially Explicit Clustering of DNA Sequences with BAPS Software. *Mol. Biol. Evol.* 30(5):1224-1228.

66. Kauffman H, Reddy A, Hsieh S, & Merca S (1973) An improved technique for evaluating resistance of rice varieties to *Xanthomonas oryzae*. *Plant Dis Rep* 57:737–741.



578

579 **Figure 1 Population structure of *X. oryzae* pv. *oryzae*.** The maximum likelihood tree was inferred

580 by RAxML with the generalized time-reversible model and a Gamma distribution to model site-

581 specific rate variation based on the core SNPs of all the *Xoo* genome sequences. The six lineages

582 were decorated with different colors. From inner to outer, the first circle refers to major rice planting

583 areas, the second one describes isolated times of these strains and the third one represents the rice

584 subspecies which the *Xoo* strain isolates. The outmost 6 squares represent the virulence analysis of

585 these *Xoo* against 6 different rice near-isogenic lines with one known resistance gene in each line,

586 with solid square for resistance and hollow one for sensitivity. Bootstrap support values were

587 calculated from 500 replicates, and only values of > 70% were labelled.

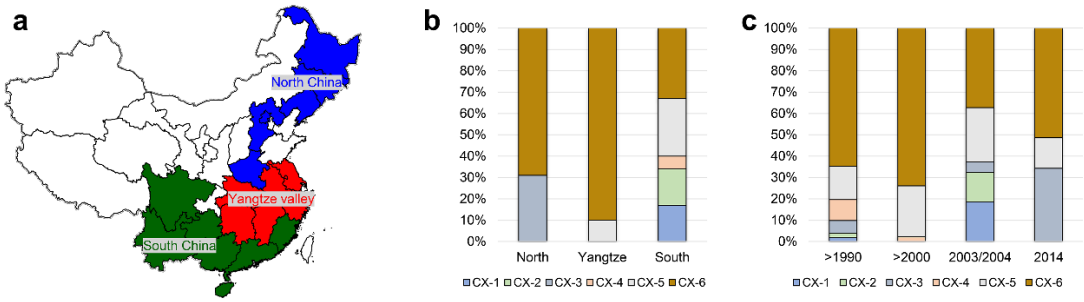


Figure 2 Distribution of the 6 major Chinese *X. oryzae* pv. *oryzae* lineages. Pathotypes of *Xoo* were determined on the following 13 near-isogenic lines, each carrying a specific R gene, IRBB2 (*Xa2*), IRBB3 (*Xa3*), IRBB4 (*Xa4*), IRBB5 (*xa5*), IRBB14 (*Xa14*). IR24 was used as a susceptible check. (a) The map of the 3 major rice producing areas in China. (b) Distribution of the 6 *Xoo* lineages among the 3 rice planting areas. (c) Dynamics of the 6 *Xoo* lineages during the past 30 years.

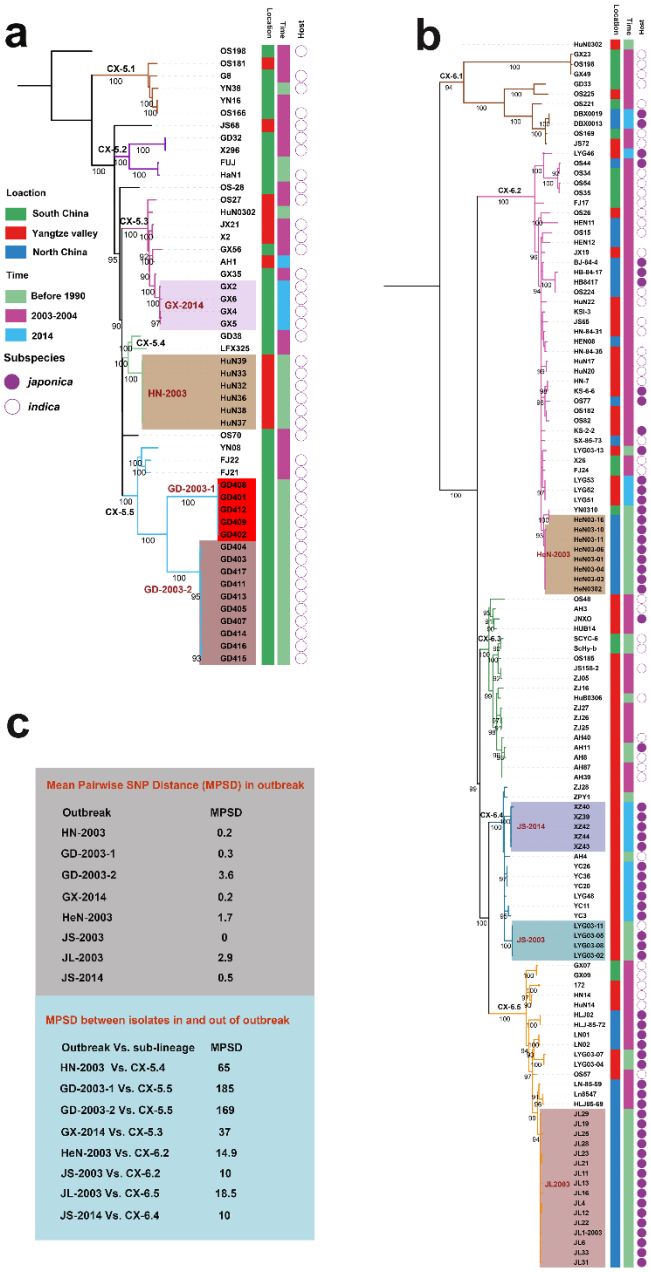


Figure 3 Phylogeography of CX-5 and CX-6. Maximum likelihood phylogeny of Lineage CX-5 (a) and CX-6 (b). Different sub-lineages were displayed by different branch colors. Bootstrap support values were calculated from 500 replicates. The recent outbreaks are represented by different backgrounds. Isolated information of *Xoo* including location, time and rice subspecies were shown in three colored strips. (c) Mean Pairwise SNP Distances (MPSDs) of *Xoo* from the outbreak, and between one in an outbreak and the other one out of the outbreak but in the same sub-lineage were displayed by two different boxes beside the tree.

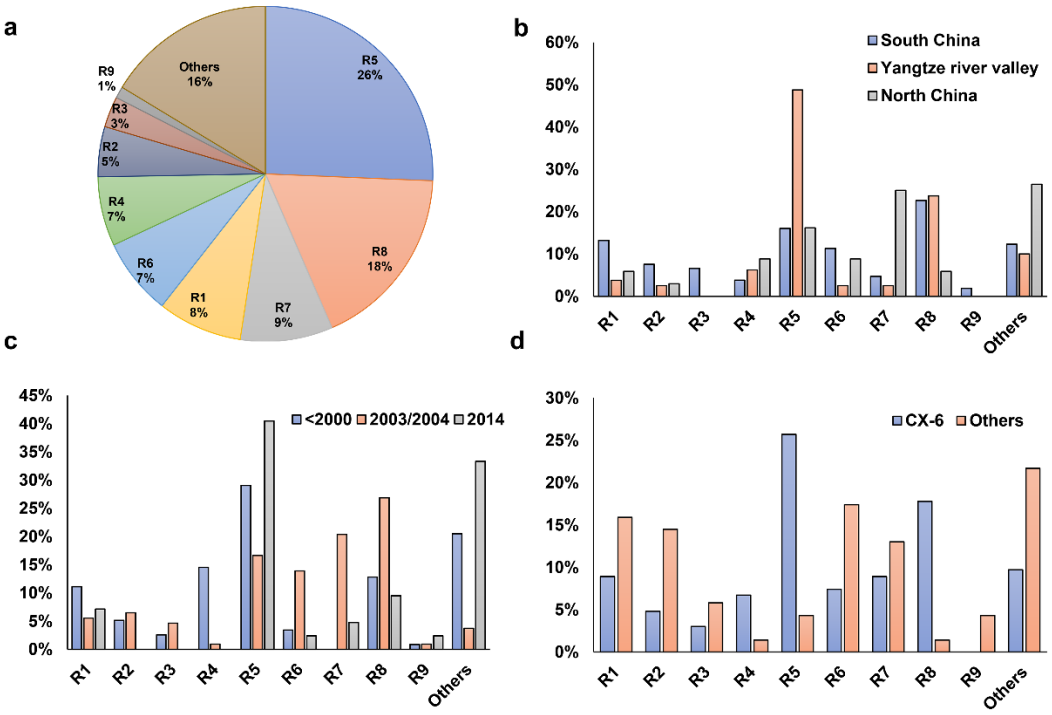


Figure 4 Virulence analysis and race classification of *X. oryzae* pv. *oryzae* China. Pathogenicity profile of each race corresponding to was shown in Table S1. (a) Proportions of the 9 *Xoo* races. (b) Distribution of the 9 races among the 3 major rice planting areas in China. (c) Dynamics of the 9 *Xoo* races during the past 40 years. (d) Distribution of the 9 races in Lineage CX-6 and others.

Table 1 Resistance of different rice lines to *Xoo* strains isolated at different times. Shown is the percentage of *Xoo* strains isolated at different periods that are unable to cause disease.

Time (# of strains)	**IRBB2 (*Xa2*)**	**IRBB3 (*Xa3*)**	**IRBB4 (*Xa4*)**	**IRBB5 (*xa5*)**	**IRBB14 (*Xa14*)**	**IR24 (*Xa18**)**
<1990 (35)	37	0	97	94	25	20
<2000 (69)	25	58	64	100	28	12
2003/2004 (97)	49	35	34	100	56	7
2014 (42)	36	5	79	90	33	5

*: Recurrent parent *Xa18*

620 **Table 2 Resistance of different rice lines to *Xoo* strains of different lineages.** Shown

621 is the percentage of *Xoo* strains of different lineages that are unable to cause disease.

Lineage (# of strains)	**IRBB2 (*Xa2*)**	**IRBB3 (*Xa3*)**	**IRBB4 (*Xa4*)**	**IRBB5 (*xa5*)**	**IRBB14 (*Xa14*)**	**IR24 (*Xa18**)**
CX-1 (19)	89	79	42	100	95	21
CX-2 (15)	73	93	80	100	100	7
CX-3 (20)	100	20	70	100	99	10
CX-5 (53)	11	32	47	96	17	11
CX-6 (123)	29	25	62	98	28	6

622 *: Recurrent parent *Xa18.* CX-4 was not included because of few isolates it contained.

623

Supplementary Information

Figure S1

Tables S1 and S2

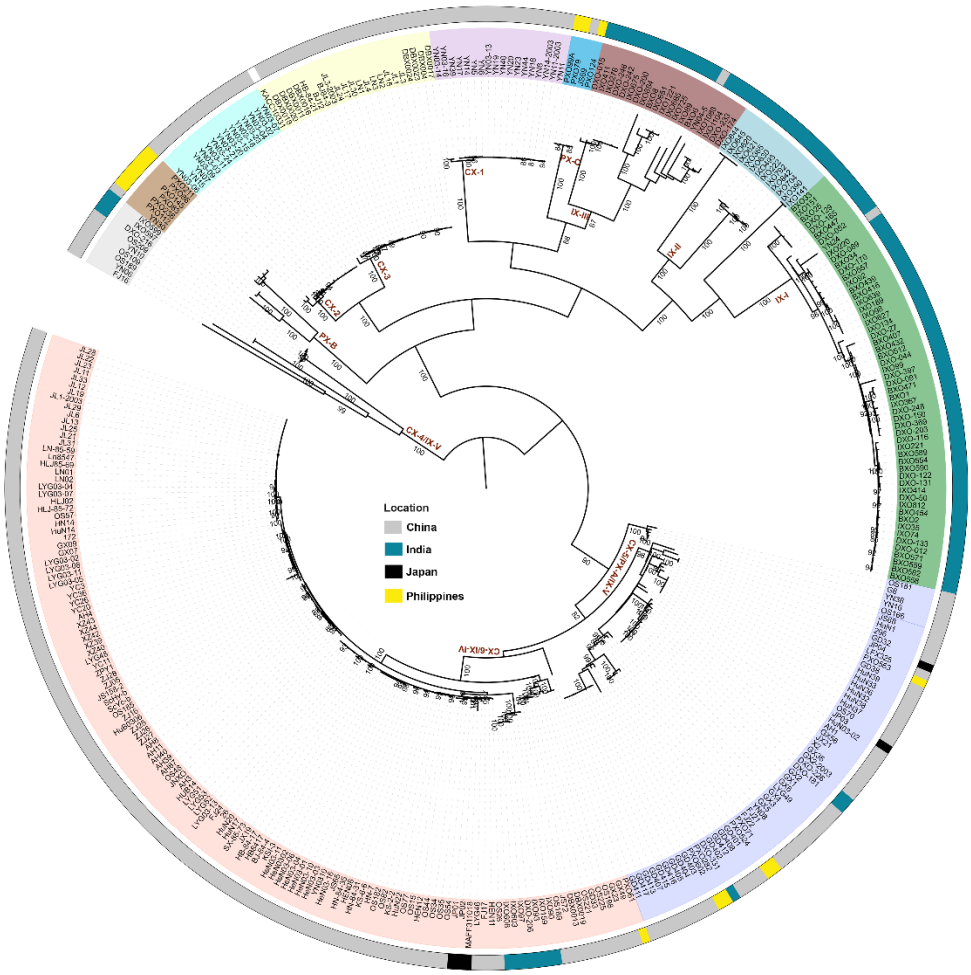


Figure S1 Population structure of Asian *Xanthomonads oryzae* pv. *oryzae*. For phylogenetic analysis, parsnp was used to call SNPs, and RAxML was used to build the ML tree. Background colors of nodes represent different lineages. The outmost circle refers to different countries where *Xoo* were isolated.

Table S2 Resistance of different rice lines to *Xoo* strains of different sub-lineages of CX-5 and CX-6. Shown is the percentage of *Xoo* strains of different lineages that are unable to cause disease.

Lineage (# of strains)	**IRBB2 (*Xa2*)**	**IRBB3 (*Xa3*)**	**IRBB4 (*Xa4*)**	**IRBB5 (*xa5*)**	**IRBB14 (*Xa14*)**	**IR24 (*Xa18**)**
CX-5.1 (5)	0	20	20	100	0	0
CX-5.3 (12)	25	25	67	92	17	17
CX-5.4 (8)	25	50	75	100	28	25
CX-5.5 (18)	0	28	28	100	28	0
CX-6.1(10)	30	30	90	100	50	20
CX-6.2 (47)	13	38	79	98	23	4
CX-6.3 (18)	10	17	68	100	11	0
CX-6.4 (18)	0	0	61	94	0	0
CX-6.5(31)	90	32	29	97	74	10

*: Recurrent parent *Xa18.* CX-5.2 was not included because of few isolates it contained.