

# SCIENTIFIC REPORTS



OPEN

## Analysis of genetic diversity of *Xanthomonas oryzae* pv. *oryzae* populations in Taiwan

Chih-Cheng Chien<sup>1,2,3</sup>, Mei-Yi Chou<sup>2</sup>, Chun-Yi Chen<sup>2,4,5</sup> & Ming-Che Shih<sup>1,2,3</sup>

Rice bacterial blight caused by *Xanthomonas oryzae* pv. *oryzae* (*Xoo*) is a major rice disease. In Taiwan, the tropical indica type of *Oryza sativa* originally grown in this area is mix-cultivated with the temperate japonica type of *O. sativa*, and this might have led to adaptive changes of both rice host and *Xoo* isolates. In order to better understand how *Xoo* adapts to this unique environment, we collected and analyzed fifty-one *Xoo* isolates in Taiwan. Three different genetic marker systems consistently identified five groups. Among these groups, two of them had unique sequences in the last acquired ten spacers in the clustered regularly interspaced short palindromic repeats (CRISPR) region, and the other two had sequences that were similar to the Japanese isolate MAFF311018 and the Philippines isolate PXO563, respectively. The genomes of two Taiwanese isolates with unique CRISPR sequence features, XF89b and XM9, were further completely sequenced. Comparison of the genome sequences suggested that XF89b is phylogenetically close to MAFF311018, and XM9 is close to PXO563. Here, documentation of the diversity of groups of *Xoo* in Taiwan provides evidence of the populations from different sources and hitherto missing information regarding distribution of *Xoo* populations in East Asia.

*Xanthomonas* is a large genus in  $\gamma$ -proteobacteria and causes diseases on more than 400 plant species. Two pathogens within the same species, *Xanthomonas oryzae* pv. *oryzae* (*Xoo*) and *Xanthomonas oryzae* pv. *oryzicola* (*Xoc*), are the most prevalent pathogens that cause diseases in rice around the world. They have a high degree of tissue specificity as well as the host-plant specificity<sup>1</sup>. *Xoo* and *Xoc* are distinguished by the tissue types they colonize in rice leaves. *Xoo* colonizes vascular tissues and causes leaf blight disease, whereas *Xoc* invades the mesophyll parenchyma tissues to cause leaf streak disease. *Xoo* can cause disease in both major rice sub-species, *O. sativa* subsp. *japonica* and *O. sativa* subsp. *indica*. However, *Xoc* cannot form severe disease symptoms in *japonica* rice<sup>2</sup>. *Xoo* has a larger host range and causes more serious problems in rice than *Xoc* in Asia, especially in East and East-Northern Asia<sup>2</sup>. *Xoo* can be further classified into races based on ability to infect different rice cultivars. To date, over 30 races of *Xoo* have been discovered all over the world<sup>3-5</sup>, and the rice blight disease can cause the yield losses up to 70% under environmental conditions favorable to *Xoo* infection<sup>6</sup>. Cultivation of *Xoo*-resistant rice cultivars is the most effective way of dealing with the blight disease, which would require the establishment of the distribution patterns of clonal *Xoo* populations in the rice growing regions.

The rice blight disease was first reported in Taiwan in 1951<sup>7</sup>, but occurred infrequently in the fields before 1970s<sup>8</sup>. However, after an outburst of rice bacterial blight impacted the local rice cultivation system in 1980s, this disease became a serious cultivation issue in Taiwan<sup>8</sup>. Although the race identification of local *Xoo* isolates has been reported since then, the isolates and rice varieties used for the classification are limited<sup>8-11</sup> and a better understanding of *Xoo* population dynamics in Taiwan is needed. Moreover, the *Xoo* populations in Taiwan cannot be compared with populations in the neighboring regions, because (i) the rice hosts selected for race identification are inconsistent in the local studies<sup>8-11</sup>, (ii) there were insufficient genetic markers for the identification of populations<sup>12</sup>, and (iii) the population analyses of Asian *Xoo* in prior studies did not include any samples collected in Taiwan<sup>13,14</sup>.

<sup>1</sup>Molecular and Biological Agricultural Sciences Program, Taiwan International Graduate Program, Academia Sinica, Taipei, 115, Taiwan. <sup>2</sup>Agricultural Biotechnology Research Center, Academia Sinica, Taipei, 115, Taiwan. <sup>3</sup>Graduate Institute of Biotechnology, National Chung Hsing University, Taichung, 402, Taiwan. <sup>4</sup>Bioinformatics Program, Taiwan International Graduate Program, Academia Sinica, Taipei, 115, Taiwan. <sup>5</sup>Institute of Biomedical Informatics, National Yang-Ming University, Taipei, Taiwan. Correspondence and requests for materials should be addressed to M.-C.S. (email: [mcshih@gate.sinica.edu.tw](mailto:mcshih@gate.sinica.edu.tw))

Rice grown in Taiwan is composed of many different varieties<sup>15</sup>, with the temperate japonica *O. sativa* being the major cultivar. This type of rice in Taiwan was originally adopted from Japan in the early 20<sup>th</sup> Century and naturalized through breeding programs. However, the major variety of *O. sativa* in Taiwan before the 20<sup>th</sup> century was the indica subspecies from China. With rice varieties from two different origins being cultivated together, *Xoo* in Taiwan may have undergone a unique evolutionary path. In addition to the different pathogenicity of *Xoo* races toward rice cultivars, the major race of *Xoo* may also be different both in geographic regions and time periods. Since *Xoo* is a rapid evolving pathogen<sup>16</sup>, the selection of cultivated rice varieties may facilitate the race shift or result in the emergence of new races. Indeed, a prior study showed that the shifting of the major race over time in the Philippines might be caused by a dramatic change in the host genotypes<sup>17</sup>. Accordingly, the local rice cultivar-driven changes of *Xoo* populations increases the difficulties in preventing this rice disease, and further emphasizes the necessity for local surveys of *Xoo* populations. By comparing local populations to *Xoo* populations in neighboring regions, it would be possible to delineate the relationships among them, and further focus on the fast-evolving pathogenicity-related genes for breeding of the blight-resistant rice.

Among the available genetic analyses, the insertion sequence *IS1112*, a relatively high copy-number repetitive element in the *Xoo* genome<sup>18</sup>, and the repetitive extragenic palindromic (REP) element have been used to characterize the isolates in Taiwan<sup>12</sup>. However, the groupings based on these two methods are not sufficient to classify the local isolates, and these studies do not provide comparative information among local populations and populations in the neighboring regions<sup>12</sup>. To address these issues, we used a robust genetic classification platform to establish the distribution patterns of *Xoo* clonal populations in Taiwan. We further used a genetically heritable sequence, clustered regularly interspaced short palindromic repeats (CRISPR) and genomic sequencing of representative *Xoo* isolates to study population dynamics of *Xoo* in Taiwan and neighboring regions. Our analyses reveal the complexity of rice cultivation history and indicate that *Xoo* populations in Taiwan may have evolved from multi-sources and undergone a unique evolutionary path.

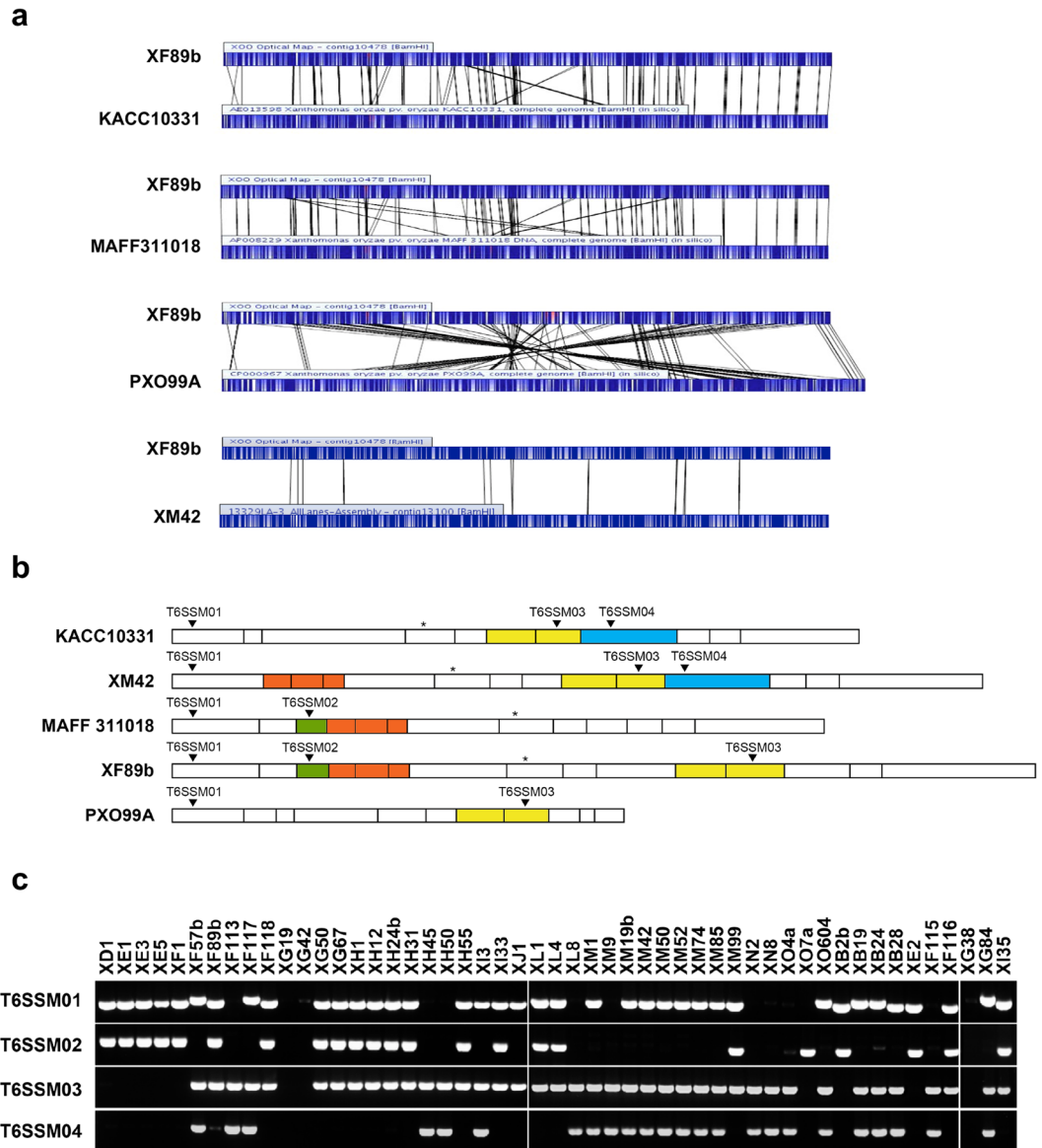
## Results and Discussion

***Xoo* isolates in Taiwan have five distinct clonal populations.** *Xoo* isolates in Asia have been suggested to have evolved into five modern genetic populations by RFLP or variable-number tandem-repeat (VNTR) analysis<sup>13,14</sup>. Nonetheless, isolates in Taiwan were not included in those studies. We used RAPD primers employed in prior studies<sup>19</sup> to analyze 51 Taiwanese isolates collected from 1986 to 2010. However, 38 of these primers had low resolutions in the isolates, and only 11 of them could be used for grouping. These results indicate that *Xoo* isolates in Taiwan are genetically close to each other. However, from these 11 RAPD markers, we could still separate the 51 isolates into five different clonal populations named R1 to R5 (Supplementary Fig. S1).

We next used RFLP to reconfirm the grouping of *Xoo* isolates in Taiwan by RAPD. This method has been widely used to assist race identification as well as the molecular classification<sup>20,21</sup>. Here, we used the *avrXa7* (one of the TALE) repeat region as a probe to classify the pattern of *Xoo* isolates generated from RFLP (Supplementary Fig. S2a). The grouping of TALE patterns resulted in four major groups (T1 to T4) and two outgroup isolates, XO4a (T5) and XM9 (T6) (Supplementary Fig. S2b). The RAPD and RFLP analyses generated similar groupings, except the T2 group of the RFLP analysis was split into R2 and R5 groups in the RAPD classification.

In order to generate novel and reliable markers for fast identification and confirmation of the *Xoo* isolates in Taiwan, we introduced a third method, i.e., searching for indel regions of the reference isolates and different signatures of restriction maps between two local isolates, XF89b and XM42. Due to their high virulence, these two isolates were used to select for *Xoo*-resistant rice cultivars in Taiwan over 30 years. We generated genome restriction maps of XF89b and XM42 using an optical mapping technique (Fig. 1a) and compared them to the maps of KACC10331, MAFF311018, and PXO99A, which were generated by *in silico* genome restriction mapping from their respective genomic sequences<sup>16,22,23</sup>. We found that the region within and near T6SS-II is highly variable among these isolates (Fig. 1b), which is similar to the study on *Pantoea ananatis*<sup>24</sup>. Hence, we designed four molecular markers for the grouping. These four markers could generate eight different patterns (S1–S8) in all the Taiwanese isolates and thus increased the classification accuracy (Figs 1c and S3). In this marker identification assay, we observed that most of the isolates showed consistent classification as in the RAPD analysis. Only five isolates had different marker patterns and thus were grouped into additional classes (S6–S8) (Supplementary Table S1). By analyzing the binary data generated from these three grouping strategies (Supplementary Figs S1a and S2b, Table S1), these 51 isolates were classified into five major groups. Two groups were evolutionarily closer to each other by the neutral marker RAPD analysis, so we designated these five groups as G1, G2a, G2b, G3, and G4 (see Supplementary Fig. S1, Table S1). Population surveys showed that from 1981 to 2010 the G1 population increased and the G4 population decreased, whereas G2a and G3 populations remained relatively steady (Fig. 2). G2a and G3 constitute the major clonal populations in Taiwan, and they accounted for 54.9% (28/51) of the collected samples (Fig. 2). All five groups showed a geographically related distribution (Supplementary Fig. S4). G1 and G2b were mainly collected in the northern part of Taiwan, and G4 was mostly discovered in the southern part, whereas G2a and G3 could be found from central to southern Taiwan.

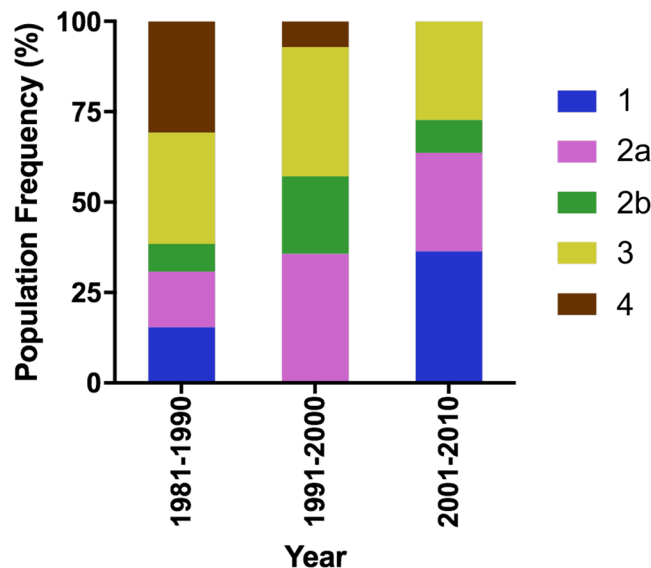
**CRISPR spacer sequences reveal the relationship among *Xoo* clonal populations from different geographic regions.** To further characterize these clonal populations, we used the CRISPR region to establish the relationship among isolates from Taiwan, the Philippines (PXOs), Japan (MAFF311018) and Korea (KACC10331). The CRISPR region contains sequences derived from lytic or filamentous bacteriophages, or conjugative plasmids (also known as spacers) adjacent to its leading sequence region (LDR)<sup>25</sup>. Therefore, the earlier acquired sequences become more distanced from LDR and provide evolutionary information through which the relationship among bacterial isolates can be identified<sup>16,25</sup>. The CRISPR regions of nineteen isolates from each Taiwanese clonal population were randomly chosen, and the 10 spacers near the LDR were sequenced or obtained from a prior study<sup>26</sup>. As expected, the spacer sequences and order among Taiwanese isolates were similar



**Figure 1.** T6SS-II region and marker polymorphism in Taiwanese isolates. **(a)** Genome restriction maps of XF89b, XM42, KACC10331, MAFF311018, and PXO99A. The maps were generated using the optical mapping technique with a restriction enzyme, BamHI. Black lines indicate similar restriction patterns between two isolates. **(b)** Restriction maps in the T6SS-II region among XM42, XF89b, MAFF311018, KACC10331, and PXO99A. The interval lines indicate the cutting site of BamHI. The star indicates the location of one of the core genes, *Hcp2*, in the T6SS-II island. Colored regions in the maps indicate the relative similar patterns between each isolate. **(c)** Four T6SS-II marker primers were designed based on the presence or absence of particular sequence regions between isolates in **(a)**. The predicted amplification region of the markers named T6SSM01 to T6SSM04 were marked in **(b)** in each isolate, and the amplification results by PCR were performed in 51 Taiwanese isolates.

within the same group but varied between groups (Fig. 3a, Supplementary Table S2). Interestingly, we found that G1 isolates had spacer sequences and order that is almost identical to the Japanese isolate MAFF311018, suggesting that G1 has a close phylogenetic relationship with this Japanese isolate. In contrast, G2b shared similar spacer sequences and order with PXO563 (Philippine race 10) with only two differences being found within these 10 spacers. However, G2a and G3 had sequences and order within this region that were distinct from both the Japanese and Philippine isolates, as well as other known CRISPR sequences in *Xoo*. Two G4 isolates XI3 and XM42 had different spacers, with XI3 being similar to G2b and XM42 similar to G3, suggesting that G4 group is a hybrid group, with its CRISPR regions inherited from either G2b or G3.

Interestingly, phylogenetic analysis using spacer sequences of the entire CRISPR region revealed that among the Philippine *Xoo* subgroups<sup>17</sup>, with the exception of PXO99A, PXO71 and PXO563, all the isolates were grouped together and separated from *Xoo* collected from Japan (MAFF311018), Korea (KACC10331), and Taiwan (XO21, XO604, XF89b and XM9) (Fig. 3b, Supplementary Table S3), which is consistent with prior studies<sup>17</sup>. The



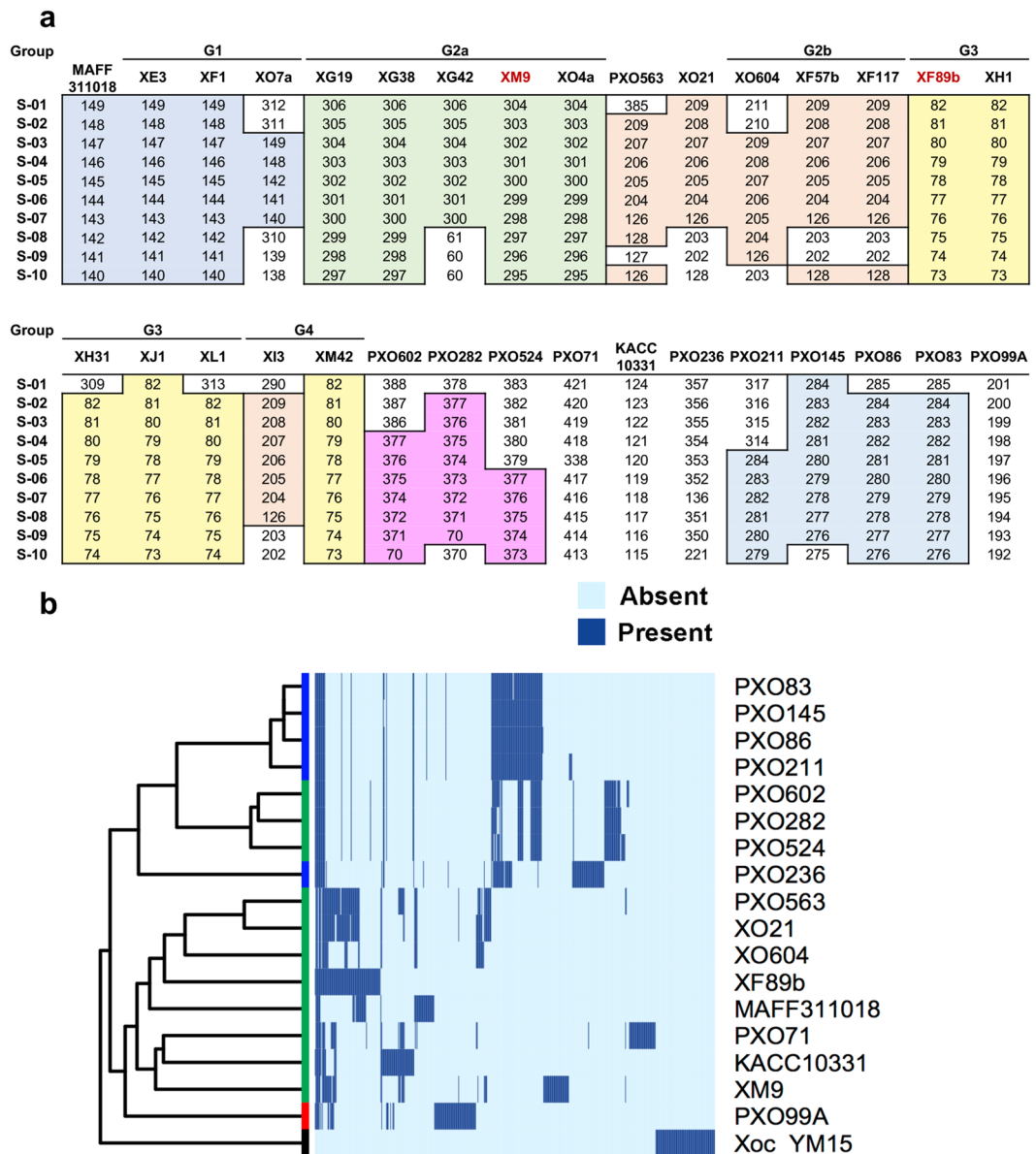
**Figure 2.** *Xoo* clonal populations in Taiwan over a 30-year collection period (1981 to 2010). Frequency of discovery rate in each clonal population during three survey periods. Each color represents a distinct clonal population in Taiwan. The number of *Xoo* entries in the following periods are: 1981–1990 = 26, 1991–2000 = 14, 2001–2010 = 11.

Philippine race 6 (PXO99A) has been considered as a foreign clonal population in the Philippines based on the molecular features, and the Philippine race 10 (PXO563) is an emerging clonal population in that area<sup>17</sup>. In our analysis, the G2b group of *Xoo* isolates, which were highly similar to PXO563 in CRISPR spacers, were first collected during the 1981 to 1990 period and continuously observed in the next two survey periods (Fig. 2), whereas the Philippine race 10 (PXO563) had only been collected in 1998 to 2002 period and therefore only accounted for a small proportion of total samples collected in that period in the Philippines<sup>17</sup>. Considering the appearance time and proportion of the Philippine race 10 (PXO563) in the Philippines and its close relationship with the G2b group in Taiwan, the Philippine race 10 (PXO563) might have spread from Taiwan or nearby countries to the Philippines. The Philippine race 4 (PXO71) might also be a foreign clonal population, because it lacked the typical patterns of the Philippine isolates in CRISPR analysis (Fig. 3b). However, the Philippine race 4 (PXO71) accounts for only a small portion of the collected samples<sup>17</sup> and does not share any last 10 spacer sequences with the Taiwanese populations (Fig. 3a). Hence, the origin of this race needs to be further investigated.

Our CRISPR spacer analysis also revealed that XF89b, XO604 and XO21<sup>26</sup> were clustered together in the same clade with the Japanese isolate MAFF311018, but each isolate also had unique spacers (Fig. 3a,b). However, RAPD, RFLP, and T6SS-II marker assays put XF89b in the G3 group and XO604 in the G2b group, and they are distinct from G1 group. In contrast, the XM9 (G2a) population has a closer relationship with KACC10331 and PXO71 than other populations. Taken together, these results showed that the distribution of clonal populations could be reflected by the CRISPR spacer sequence, and the grouping results from CRISPR spacers were in general similar to our other molecular analyses.

Due to intense breeding of pathogen-resistant rice cultivars, *Xoo* has to evolve fast to overcome new resistant traits, and the dominant population may shift from one race to another locally and temporally because of the different virulence toward the rice cultivars. In the Philippines, population shift was observed across a 40-year-period, and three dominant populations, PXO602(3c), PXO282(1) and PXO524(9b) accounted for over 90% of *Xoo* populations across the latest two survey periods from 2003 to 2012<sup>17</sup>. The genomic analysis<sup>17</sup> and CRISPR analyses suggest that these three major races share a highly similar genetic background and were recently separated from each other (Fig. 3b). Unlike the situation in the Philippines, the population shift in Taiwan is not obvious by the molecular classification. At least two populations, G2a and G3 have dominated the *Xoo* populations in Taiwan over a 30-year-period (Figs 2 and 3a). In addition, the phylogenetic analyses indicated that *Xoo* populations in Taiwan evolved independently from the major Philippine populations and were closer to the Japanese and foreign Philippine isolates, MAFF311018 and PXO563 (Fig. 3b).

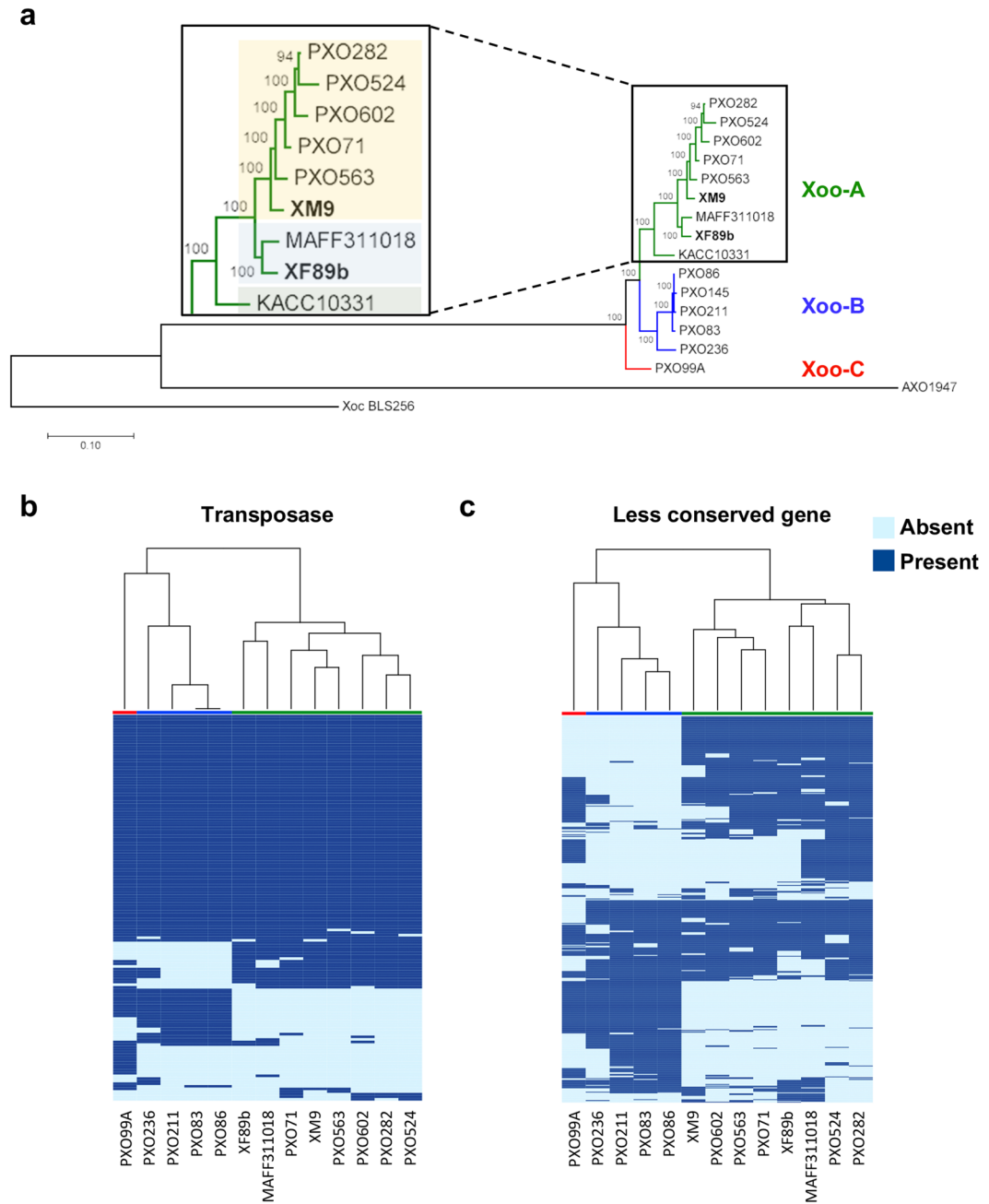
**XF89b and XM9 belong to two different groups within the Xoo-A clade.** Since G2a and G3 have unique CRISPR spacers that have not been reported before, we chose the isolate XM9 from G2a, which has a unique TALE signature, and XF89b from G3, which has a typical TALE pattern (Supplementary Fig. S2a), for genome sequencing. The whole genome sequences indicate that XM9 and XF89b had similar genome structures with KACC10331, MAFF311018, and those of the Philippine *Xoo* group PX-A<sup>17</sup> (Supplementary Fig. S5). A SNP phylogenetic tree was further created by *Xoo* genome sequences, and these *Xoo* isolates were categorized into three major clades, which were named as Xoo-A, Xoo-B, and Xoo-C (Fig. 4a). The tree showed that KACC10331, MAFF311018, XF89b and XM9 were clustered into Xoo-A clade, while the isolates from the Philippines<sup>17</sup> were grouped into Xoo-A, Xoo-B and Xoo-C, respectively (Fig. 4a). Unlike the CRISPR region analysis (Fig. 3b),



**Figure 3.** Alignment of CRISPR spacers from the *Xanthomonas oryzae* pv. *oryzae* (*Xoo*) genomic sequences. **(a)** Comparison of the last 10 spacer sequences beginning from LDR. The same color block indicates the same or similar (<2 nucleotides mismatch) spacers shared in the group. The first column represents the order of the 1<sup>st</sup> (S-01) to the 10<sup>th</sup> (S-10) of the last 10 spacers in the CRISPR region. The number indicates the sequence ID in the spacer library (Supplementary Table S2). **(b)** Alignment of entire CRISPR spacers in all fully sequenced *Xoo* isolates. *Xoc* YM15 is the only *Xoc* isolate containing CRISPR system, and this isolate is used as an outgroup in the analysis. The colored bar at the end of the phylogenetic tree represents the groups defined from genome SNP analysis (Fig. 4a).

PXO282, PXO524, and PXO602 were separated from other Philippine isolates and clustered to the Xoo-A clade (Fig. 4a). The CRISPR spacers evolve by either stepwise acquisition or recombinational loss or duplication of several spacer-repeat units<sup>25</sup>. Although the stepwise acquisition is the major evolutionary event of CRISPR elicited by environmental challenges, the rare events of spacer losses or duplications contribute to the footprint of spacers in the CRISPR region. They also affect the distance of the phylogenetic tree analysis. Most of the Philippine *Xoo* isolates contain a similar spacer footprint that is different from isolates collected from other areas and thus grouped together. Also, CRISPR in bacteria serves as an immune system<sup>27</sup>. Under frequent environmental challenges, the CRISPR region of *Xoo* might be under a stronger selection force than most of the other functional regions and obtain or retain the same spacers in the genome. It could further explain the different grouping results between CRISPR spacer analysis and genome SNP analysis.

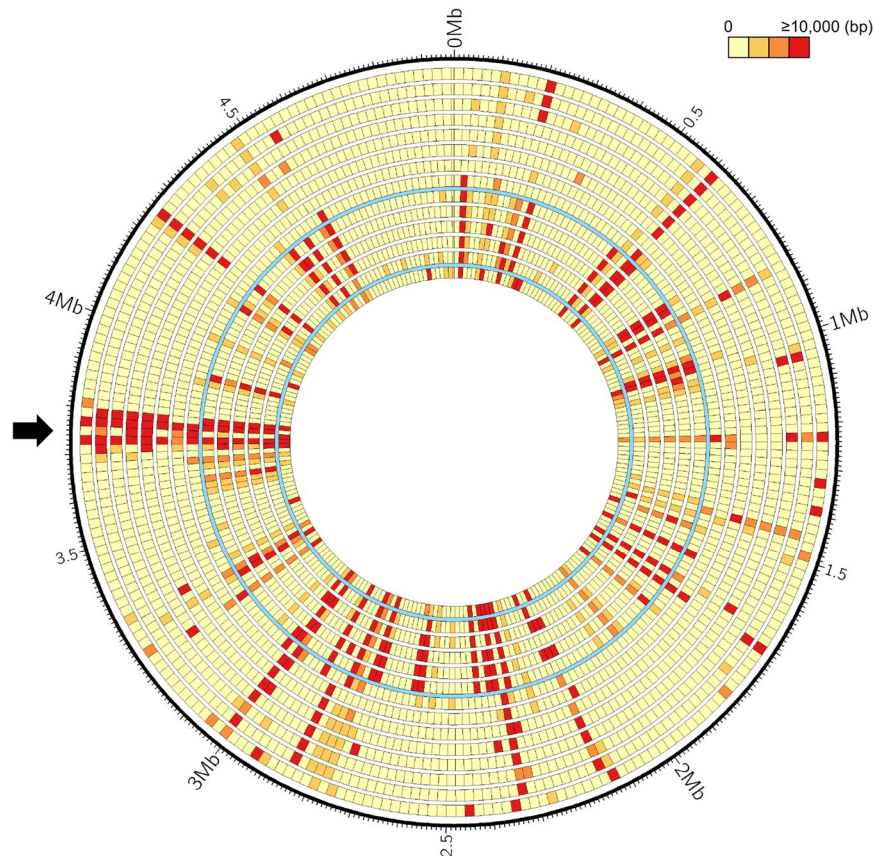
Within Xoo-A group, the isolates collected from different countries could be further divided into three sub-clades, which revealed that XF89b was clustered closely with MAFF311018 (Fig. 4a). Linkage distance analysis with bootstrap test showed that the XF89b branch had longer individual distance in the phylogenetic tree and



**Figure 4.** Phylogenetic relationships of *Xanthomonas oryzae* pv. *oryzae* (*Xoo*) isolates. **(a)** Maximum likelihood phylogenetic tree of sequenced *Xoo* isolates is based on 86,919 concatenated core SNPs. Xoo-A, Xoo-B, and Xoo-C groups are indicated by the green line, blue line and red line, respectively. In Xoo-A, three different background colors represent distinct subgroups. African *Xoo* isolate, AXO1947, and *Xoc* BLS256 were used as outgroups. **(b,c)** A heatmap of hierarchical clustering is based on 148 transposases between *Xoo* isolates **(b)** and 751 less conserved genes between *Xoo* isolates **(c)**, and the colored bar represents the groups defined from genome SNP analysis.

distinct evolutionary direction from other isolates in Xoo-A, including XM9, suggesting that it was under adaptation process (Fig. 4a). Besides, in transposase and less conserved gene analyses, XF89b and XM9 belonged to different subclades of Xoo-A (Fig. 4b,c). These results suggest that there are at least two distinct major clonal populations in Taiwan. XM9 might share a more recent ancestor with Philippine Xoo-A isolates and XF89b evolved from a later common ancestor of the Japanese isolate MAFF311018.

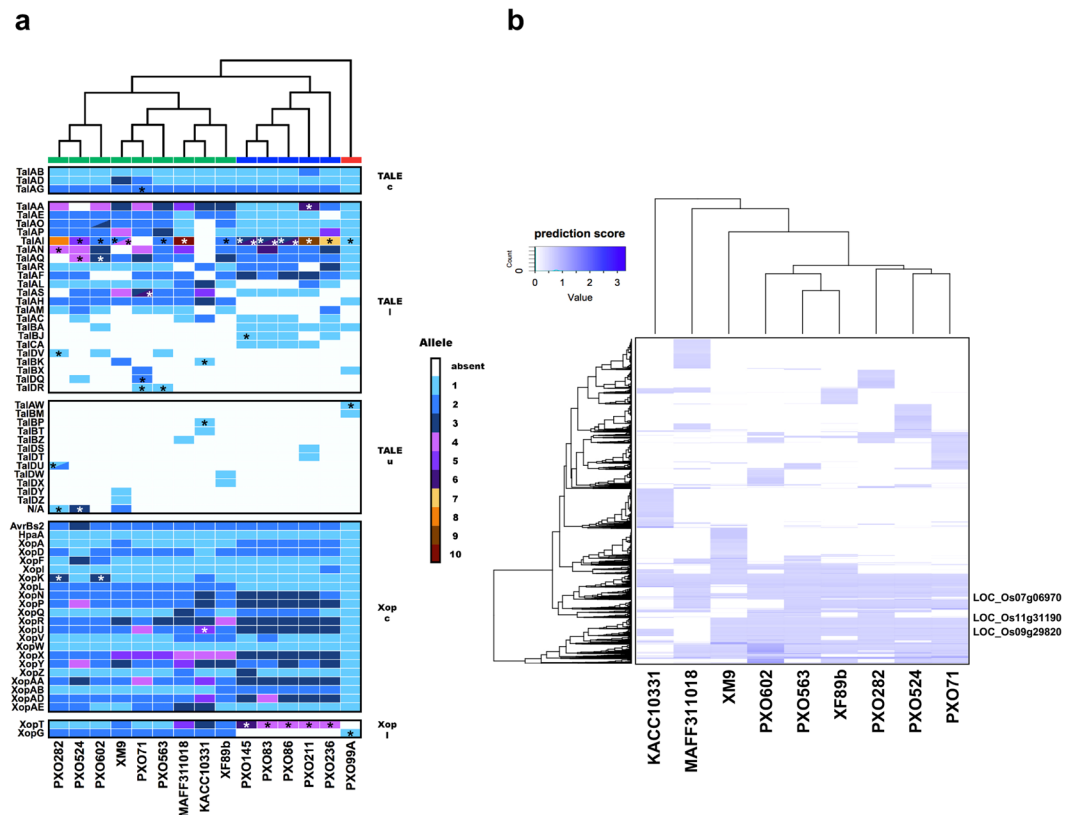
**T6SS-II evolved rapidly in *Xoo*.** In bacterial species, recombination events drive evolutionary fitness and have long-term effects. However, the isolates in the Xoo-A clade share a highly similar genome structure and might have undergone a bottleneck event in their ancestry<sup>17</sup>. Although the recombination events were not strong



**Figure 5.** A heatmap of indel comparison in *Xanthomonas oryzae* pv. *oryzae* (*Xoo*) isolates. The window of indel map was fixed to 10 kb for each block, and the number of base-pair differences was generated from the gap detection function in MAUVE<sup>52</sup>. The cyan lines separate Xoo-A, Xoo-B, and Xoo-C clades of *Xoo* isolates and the arrow indicates T6SS-II region of *Xoo* isolates. The circles from outside to inside: XF89b versus PXO282, PXO524, PXO602, PXO71, PXO563, XM9, MAFF311018, KACC10331, PXO86, PXO145, PXO211, PXO83, PXO236, and PXO99A.

enough to disrupt the linkage signal in the Xoo-A group due to strong host selection, it might still be an important source of variability. Thus, we focused on searching for the unique genes and recombination events. We found that a region containing T6SS-II and putative T6SS-related genes were variable in their sequences in the indel map (Fig. 5). T6SS-II of *Xoo* was considered to contain a large number of T6SS-unrelated ORFs interspersed with T6SS-related genes<sup>28</sup>. After annotating the genes in this region, many hypothetical genes were found after the *VgrG*-like gene, and each *VgrG*-like gene was considered as the first gene in the putative operons without intergenic non-coding region between genes (Supplementary Fig. S6). *VgrG* operon has been shown to contain toxin and anti-toxin pairs for bacterial competition and preventing suicide<sup>29,30</sup>. Therefore, these hypothetical genes might be T6SS-related genes and under T6SS-II regulation. Notably, none of the genes annotated as a unique gene were found in this region, and the genome structure is different from each other among Xoo-A isolates. Taken together, these results indicate that T6SS-II region is not only variable between the Xoo-A, Xoo-B and Xoo-C groups, but frequently rearranged even within the Xoo-A group, in which the isolates had similar genomic backgrounds. This finding implies that the evolutionary rate might be high in this region. These results also suggest that T6SS-II plays an important role in either host-pathogen interaction or the survival of *Xoo* in the environment.

**XF89b and XM9 carry different sets of T3SS effectors.** T3SS effectors, including TALEs and Xops, are one of the most important classes of effectors in *Xanthomonas* pathogens that are related to the overall pathogenicity of each *Xoo* isolate<sup>2,17,31</sup>. The TALEs has a conserved repetitive region in the middle of their coding sequences, and the variations at 12<sup>th</sup> and 13<sup>th</sup> amino acid of each repeat, so called repeat variable diresidue (RVD), determine the binding specificity of the nucleotide<sup>32</sup>. The binding targets of TALEs play important roles in rice resistance towards *Xoo*<sup>2</sup>. Also, Xops can determine the virulence of *Xoo*<sup>31</sup>. Therefore, the classification of these effectors might provide additional information to clarify the differences between each *Xoo* isolate. Here, we used all the TALEs including previously classified TALEs<sup>33</sup> and Xops<sup>17</sup> to group all the sequenced *Xoo* isolates (Fig. 6a, Supplementary Tables S4 and S5). The grouping pattern was similar to that of the genome SNP classification (Fig. 4a). As expected, XF89b shared large numbers of T3SS genes with MAFF311018, as they are the closest clones based on the genome sequences. However, XM9 has quite different T3SS gene sequences, from its closest



**Figure 6.** Genome comparison of *Xanthomonas oryzae* pv. *oryzae* isolates. **(a)** Comparison of type III secretion system effectors in analyzed *Xoo* isolates with genome sequences. Different allele types of effectors were grouped and colored. The color scale indicates types of alleles present in each class of effectors. N/A represent TALEs that were not assigned to any class in AnnoTALE<sup>33</sup>. Different colors in the N/A group do not share similar RVD compositions. Stars represent truncated TALEs/interfering TALEs or pseudo-Xop genes. The color-bar at the top of the heatmap indicates the classification based on genome SNPs. On the right y axis: c; core gene, l; less conserved gene, and u; unique gene. **(b)** TALE target prediction of Xoo-A group of *Xoo* isolates. A recognition possibility data of 200 predicted target genes of the rice host in each TALE were generated using TALVEZ<sup>55</sup>, and an *Oryza sativa* genome, MSU7, was used as the template for target prediction. The top one target of the list was normalized to 1, and values for different TALEs that targeted to the same gene were calculated as a sum of the normalized values in each TALE. The known TALE-targeted gene IDs were shown at the right side of the heatmap. The gene IDs and recognition scores are shown in Supplementary Tables S6 and S7.

reference isolates PXO71, PXO563 and local isolate XF89b. This is consistent with the observation that XM9 contains a unique TALE pattern, which is distinct from other *Xoo* isolates in RFLP analysis (Supplementary Fig. S2). In the genome TALE distribution analysis, we found that TALE islands in the genome are highly conserved within each *Xoo* group (Supplementary Fig. S7), and rearrangements of TALEs between each island are usually accompanied by genome retro-rearrangements in the nearby regions (Supplementary Fig. S5).

XF89b had a similar TALE distribution to its closest isolate MAFF311018. The fifth TALE island of both XF89b and MAFF311018 contains a single pseudo-TALE with a similar repeat region, but XM9 and its closest isolates, PXO563 and PXO71, have two pseudo-TALEs in the region, one of which is similar to the pseudo-TALE in XF89b and MAFF311018 (Supplementary Fig. S7, Table S5). Moreover, XF89b and MAFF311018 shared the same type of TALE gene, *Tal2b*<sub>XF89b</sub> (allele type 2 of TalAB), whereas XM9, PXO563 and PXO71 shared the other type of TALE gene, *Tal2b*<sub>XM9</sub> (allele type 1 of TalAB) in the corresponding position (Supplementary Fig. S7). In order to investigate whether there are some TALEs that are uniquely evolved in Taiwan, we used TALE sequences of *Xoo* isolates from other regions and Taiwanese isolates to generate a DisTAL tree using each repeat of TALEs as a unit and transform it into a coded repeat<sup>34</sup>. The pair-wise alignment was then applied to construct the tree. From the DisTAL tree, we found that XF89b and XM9 shared five identical TALE repeat arrays (TalAA, TalAE, TalAH, TalAO and TalAR). Nevertheless, other TALEs were in different sub-roots (TalAB, TalAD, TalAP and TalaQ). XF89b and XM9 also share two identical RVDs of the TALEs in the AnnoTALE<sup>33</sup> classes (TalAG and TalAL) but have short distance in DisTAL tree. This is caused by the amino acid variations in the repeat arrays other than RVDs. Also, these two isolates had eight TALEs classified into different AnnoTALE classes and sub-roots (see Supplementary Fig. S8, Table S4). These results suggest an early divergence occurred between XF89b and XM9. Interestingly, the DisTAL tree showed that one of the TALEs in XM9, *Tal3a*<sub>XM9</sub>, was an outgroup of all the analyzed TALEs. It has a remarkably short sequence with a 3.5 repeat array and a complete TALE sequence without



disruption. Indeed, this short repeat array was also shown in RFLP analysis (Supplementary Fig. S2). This small fragment was not unique in XM9, 13 (27%) other Taiwanese isolates also contained this or similar-sized fragment. Over a quarter of the collected samples have the short central repeats, implying that it is not a recent mutated form. This TALE is believed to be too short to bind DNA<sup>32</sup>. Recently, a study on Xo1 suggests that TALEs with short central repeats (at least 3.5 repeats) are required to trigger the rice resistance. However, studies on truncated TALEs (truncTALEs) and/or interfering TALEs (iTALEs) reveal that truncTALEs/iTALEs can mimic the TALE structure and overcome the Xo1- and Xa1-mediated resistance<sup>35,36</sup>. Accordingly, it remains to be clarified whether these small TALEs function as iTALEs and are recognized by Xa1 and Xo1.

Two R genes in rice, *Xa27* and *Xa7* that are effective against specific *Xoo* isolates, have been well studied<sup>37–40</sup>. Although XF89b and MAFF311018 are the closest isolates in genome SNP analysis, only XF89b, but not MAFF311018, seems to have escaped the *Xa27* gene activation trap. This is probably because XF89b has a TALE in the allele type 2 of TalAO class, which is distinct from *AvrXa27* (allele type 1 of TalAO) in MAFF311018<sup>17</sup> (see Supplementary Table S5) and that the activation of the *Xa27* gene in rice is initiated by the binding of *avrXa27* on its promoter region<sup>41</sup>. This phenomenon was also observed in XM9, suggesting that *Xa27* is not an efficient resistance gene against Taiwanese isolates. In addition, XF89b and XM9 do not have an *AvrXa7* (TalAC class) or an *AvrXa7* homolog (TalDV class) (see Supplementary Table S5). The TALEs in XF89b and XM9 that are closest to *AvrXa7* are *Tal6c<sub>XF89b</sub>* (TalDW1) and *Tal6c<sub>XM9</sub>* (TalDZ1) (Supplementary Fig. S8, Table S5). However, these two TALEs have longer TALE repeat arrays and different dipeptide compositions from other reference TALEs. They are similar to each other but still have minor variations in dipeptide compositions. This observation suggested that *Tal6c<sub>XF89b</sub>* and its related TALEs in Taiwanese isolates might have evolved from *AvrXa7* under higher selection pressure.

**XM9 and XF89b differentially regulate TALE-targeted rice genes in TNG67.** Prior studies showed that many TALEs have the ability to cause disease lesion by up-regulating expression of certain host genes<sup>17</sup>. Thus, TALE target prediction also provides insight on how rice cultivars are affected by *Xoo* at the molecular level. Here, we compared the predicted rice gene targets potentially recognized by TALEs in the Xoo-A group (Fig. 6b, Supplementary Tables S6 and S7).

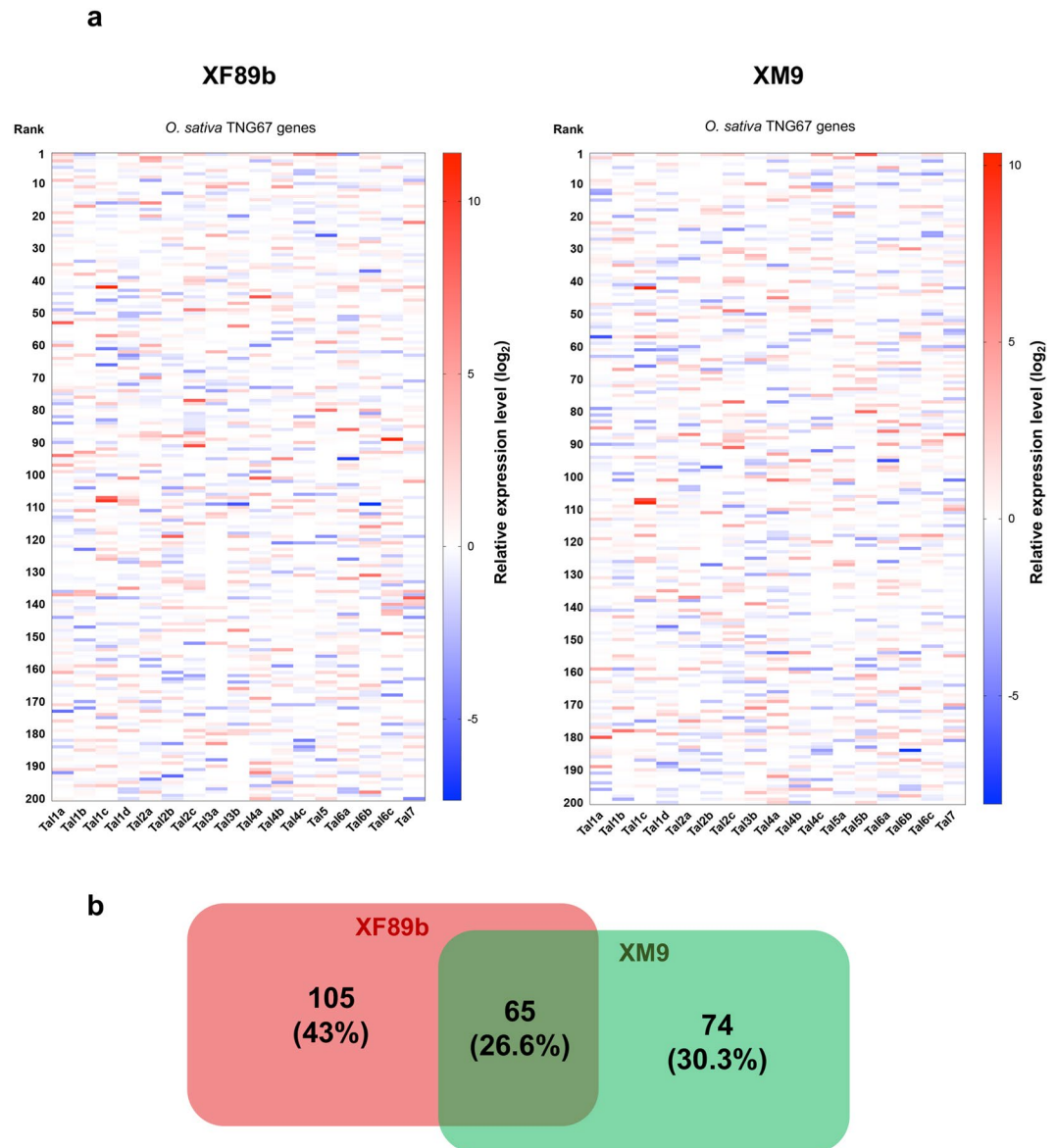
XM9 and XF89b had distinct lists of predicted TALE-targeted genes (Fig. 6b, Supplementary Table S7). Therefore, they might regulate different sets of rice genes during colonization. To evaluate the up-regulated genes at 7 day-post-inoculation (dpi) with *Xoo* isolates compared to mock-inoculated rice leaves in the predicted TALE-targeting lists, we used RNA-seq to examine the gene expression profiles of the TALE-targeted genes after inoculation of XM9 and XF89b (Supplementary Table S8). The heatmap of expression levels of the putative TALE-targeted genes showed that some of the genes were highly expressed after *Xoo* infection and ranked as the top 10 genes in the lists, but some of them did not have high rankings (Fig. 7a). This indicates that the predicted lists combined with transcriptome data may enhance the chance of finding the real targets of TALEs. Moreover, when we compared the putative TALE-targeted gene lists in XM9 and XF89b, only 26.6% of genes were simultaneously up-regulated. Many other genes were either unique in the strain-specific putative TALE-targeted gene lists or only up-regulated by one of these two isolates (Fig. 7b).

Next, we focused on the two TALEs, *Tal6c<sub>XF89b</sub>* (TalDW1) and *Tal6c<sub>XM9</sub>* (TalDZ1), because these two TALEs were close but different from the *AvrXa7* (TalAC) or the *AvrXa7* homolog (TalDV). The predicted targets for these two TALEs were different. *Tal6c<sub>XM9</sub>* had the highest prediction score to target *Os11N3* (Supplementary Table S6), which is also a direct target of *AvrXa7*<sup>42</sup>. Nevertheless, this rice gene was only up-regulated 3.2-fold (Supplementary Tables S8 and S9). In addition, several other rice genes predicted as *Tal6c<sub>XM9</sub>* targets had higher expression levels than *Os11N3*, but most of them have unknown functions. In contrast, *Tal6c<sub>XF89b</sub>* is predicted to target *LOC\_Os04g58860* (Supplementary Table S6). However, this rice gene was not induced under *Xoo* XF89b infection (Supplementary Table S9). Also, *Os11N3* was only up-regulated 5.7-fold (Supplementary Table S9). Instead, several other unknown proteins, such as *LOC\_Os05g48840* and *LOC\_Os10g06000*, were up-regulated to 2708-fold and 107-fold, respectively (Supplementary Table S9). We hypothesize that the variations in *AvrXa7* and *AvrXa7* homologs were caused by the selection pressure of *Xa7*, since the TALE and its homologs showed a high correlation with IRBB7 rice isogenic line containing a single R gene, *Xa7*<sup>17</sup>.

In the other cases, *Tal1a<sub>XF89b</sub>* and *Tal1a<sub>XM9</sub>* were in the TalAP class (Supplementary Table S5). *Tal1a<sub>XF89b</sub>* presented as the major form (allele type 1 of TalAP) in the class, and *Tal1a<sub>XM9</sub>* had some variations that were not identical to other TalAP members. They share the same target gene, *OsHEN1*, in the prediction list (Supplementary Table S6). *OsHEN1* is probably a target of allele type 1 of TalAP TALEs<sup>43</sup>. However, the variation form of TalAP, *Tal1a<sub>XM9</sub>*, failed to activate the transcription of this gene, since it was only up-regulated by 1.7-fold under XM9 infection (Supplementary Table S9). On the contrary, *OsHEN1* was up-regulated 6-fold under XF89b infection (Supplementary Table S9).

Moreover, XM9 also showed many variations in TALE repeat arrays and a different predicted target list as well as the transcriptional levels of the targeted rice genes from other TALEs within the AnnoTALE classes. This suggests that XM9 might be under a strong selection pressure to escape the gene activation traps, or it has evolved a new ability to target other host genes. Taken together, XF89b and XM9 have similar TALE island distribution in the genome as the isolates of Xoo-A clade, but they only share a few types of TALEs. In addition, they only share about one-fourth of TALE targets that can be induced by the inoculation of both isolates. Therefore, further investigation in recognition of these TALEs will be required to better characterize the effects of changes in the RVDs toward rice hosts.

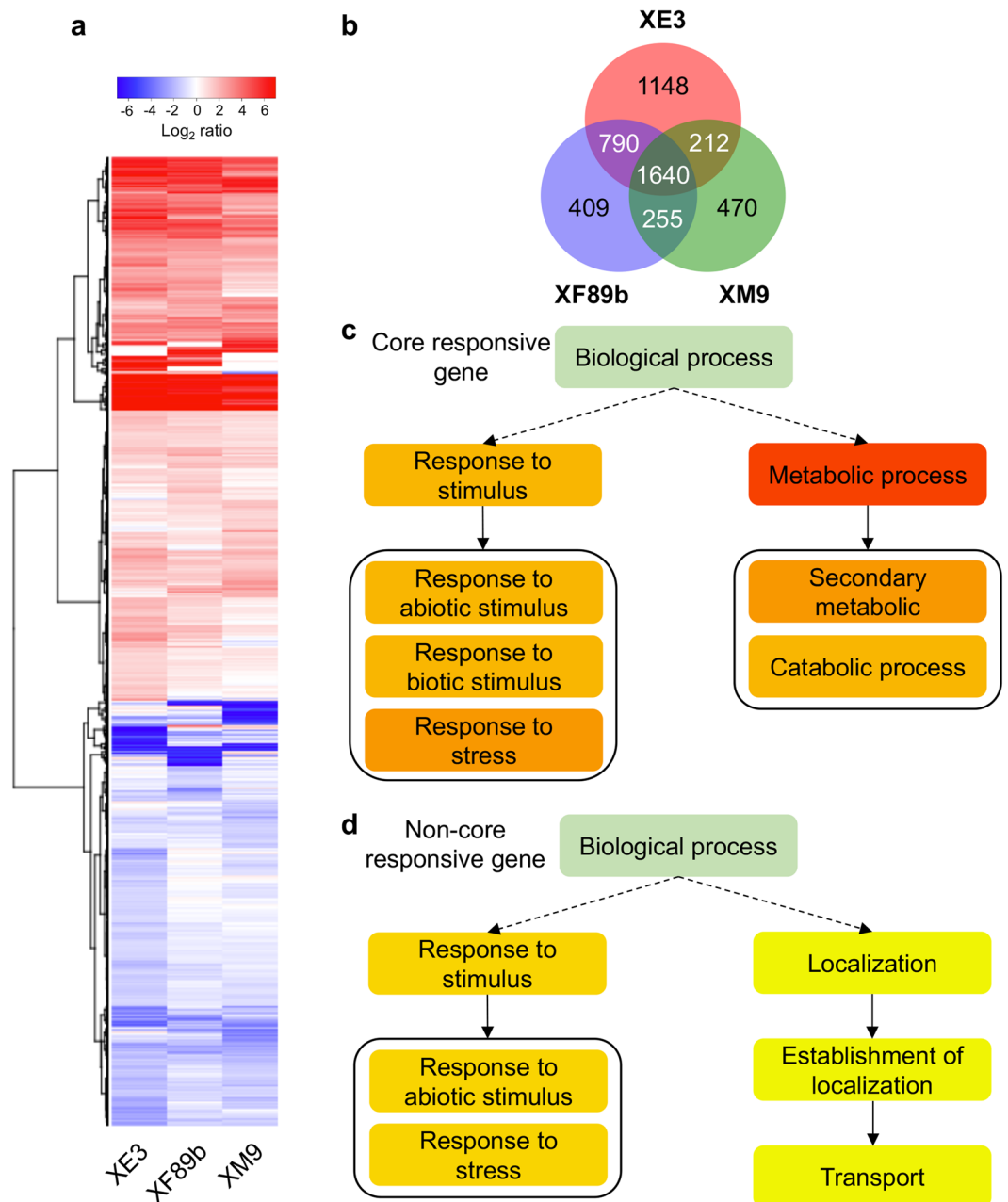
**Different *Xoo* clonal populations regulate rice genes distinctively.** Besides the regulations of TALE-targeted host genes, different clonal populations may also regulate different sets of host genes to overcome the obstacles from various host genetic backgrounds during colonization. From the molecular classification of



**Figure 7.** Transcriptome profiles of predicted TALE-targeted rice genes on *O. sativa* TNG67. (a) The up-regulated and down-regulated genes of *O. sativa* TNG67 were profiled after 7 days of *Xoo* XF89b and XM9 inoculation compared to mock-inoculated rice samples, and the relative expression levels of the top 200 TALE-targeted genes predicted by TALVEZ<sup>55</sup> are listed. The relative expression levels were calculated from infected samples divided by mock-treated samples. Every column represents a different gene list which was predicted as the putative targets of individual TALEs. TAL3a<sub>XM9</sub> only had three DNA binding residues, and its targets were excluded from the prediction lists. (b) The Venn diagram of up-regulated genes ( $\geq 2$ -fold changes) in the TALE-targeted prediction lists between XF89b and XM9. A total of 244 putative TALE targets in XF89b and XM9 were included in the comparison.

Taiwanese *Xoo* isolates, we observed that G2a, and G3 dominate the populations, and G1 was increasing in the last survey period. To understand whether these three populations regulate different sets of the host genes, we chose representative isolates from these three populations, XE3 (G1), XM9 (G2a) and XF89b (G3), for transcriptomic analyses after the inoculation of local temperate japonica rice, TNG67. XE3 had exactly the same sequence and order in the last 10 CRISPR spacers with MAFF311018. Also, it showed similar TALE patterns in Southern blot assay as well as the indel marker assay with MAFF311018 (Fig. 1a,b).

After the inoculation of TNG67 with these three isolates, we found that 4924 rice genes were up-regulated and 3856 genes were down-regulated by XE3, XM9 or XF89b (Fig. 8a, Supplementary Table S9). In the genome and molecular marker analyses, we observed that XF89b (G3) was genetically closer to MAFF311018, and XE3 (G1) was considered to belong to the same group as MAFF311018. On the other hand, XM9 (G2a) phylogenetically separated from the other four groups at the earlier time point. Hence, XE3 and XF89b may induce a more similar proportion of host genes than either XM9 and XE3 or XF89b and XE3. As expected, among the 4924 up-regulated



**Figure 8.** Transcriptome profiles of differentially expressed genes in TNG67 after inoculation of *Xoo* isolates. (a) A heatmap of differentially expressed rice genes after inoculation of *Xoo* XE3, XF89b, and XM9. The rice samples were collected at 7 days-post-inoculation (dpi). A total of 4924 up-regulated ( $\geq 2$ -fold change in at least one sample) genes, and 3856 down-regulated genes ( $\geq 2$ -fold change) (Supplementary Table S9) were included in the map. (b) The Venn diagram of up-regulated genes in all three samples. A total of 3790, 3094, and 2577 up-regulated genes ( $\geq 2$ -fold changes) in XE3, XM9, and XF89b were used for the diagram. (c) The GO enrichment of up-regulated genes that altered ( $\geq 2$ -fold change) in all three samples as core responsive genes. (d) The GO enrichment of up-regulated genes ( $\geq 2$ -fold changes) that were only altered in one or two samples as non-core responsive genes.

genes in TNG67 after inoculation of *Xoo* isolates, 2430 (49.3%) of these genes were up-regulated by both XE3 and XF89b, which is more than genes up-regulated by XE3/XM9 (37.6%) or XE3/XF89b (38.5%) (Fig. 8b). Among these 4924 genes, one-third of genes were up-regulated by all inoculated isolates, and these genes might be core genes responding to *Xoo* infections. Indeed, GO enrichments showed that these genes were mainly involved in the responses to stresses and metabolic processes (Fig. 8c). Interestingly, there were still many genes that were specifically up-regulated under the infections of one or two isolates. These isolate-specific *Xoo*-induced genes were enriched in response to the stimuli and were also involved in localization and transport (Fig. 8d). As the OsSWEET11/OsSWEET14 are known as sugar transporters, and many TALEs target these transporters to create

a more suitable environment for *Xoo*<sup>44,45</sup>, the transcript levels of transporters are crucial for the fitness of *Xoo* races. Thus, the enrichment of genes involved in transport in the non-core responsive gene list might result in different fitness of these three *Xoo* isolates toward different rice cultivars.

## Conclusions

We showed here that *Xoo* isolates in Taiwan consisted of 5 clonal populations, G1, G2a, G2b, G3 and G4. Genome SNPs and CRISPR spacer analyses suggested that G2a and G4 were evolutionarily closer to the Philippine *Xoo*-A group of isolates, whereas G3 and G1 were closer to the Japanese isolate MAFF311018. This phenomenon could be caused by the mixed cultivation of both indica and japonica rice originally from China and Japan, respectively. Comparison of genome indel regions and mutational analyses also revealed the possible important role of T6SS-II in *Xoo* pathogenicity. TALE analysis also revealed that XF89b and XM9 may target different genes in rice, and some of TALE genes, such as *Tal6c<sub>XM9</sub>* and *Tal6c<sub>XF89b</sub>*, evolved faster in Taiwanese isolates than isolates in other countries. Furthermore, by genomic comparison of two local isolates with isolates from neighboring regions, we are able to depict the distribution patterns of East Asian *Xoo* populations and how they spread and adapt to local environments. These results contribute to our understanding of the population dynamics of this important rice pathogen and also provide a novel means through which *Xoo*-rice interactions may be studied in the future.

## Methods

**Bacterial isolates, plasmids, primers and isolation of genomic DNA.** Fifty-one isolates of *Xanthomonas oryzae* pv. *oryzae* were collected from a randomly chosen areas in Taiwan by the Taiwan Agricultural Research Institute (Supplementary Table S10). *X. oryzae* pv. *oryzae* and *E. coli* were grown in LB medium at 28 °C for 3 days and 37 °C for 18 hours, respectively. Genomic DNA was extracted by Easy Tissue & Cell Genomic DNA Purification Kit (GMBiolab Co, Taiwan) and stored at 4 °C. Plasmids were purified by Mini Plus Plasmid DNA Extraction Kit (Viogene BioTek Co., Taiwan) and stored at 4 °C. Primers used in this study are listed in Supplementary Table S11.

**Random amplified polymorphic DNA analyses.** Thirty-eight random amplified polymorphic DNA (RAPD) primers selected from Hu *et al.*<sup>19</sup> were used for DNA fingerprints, and 11 of them were used for phylogenetic tree analyses. RAPD analyses was carried out by PCR amplification using 50 ng gDNA, 100 pmole RAPD primer, and 7.5 µl DNA polymerase master mix in 15 µl final volume. The 11 primers for phylogenetic tree analyses were P09, S01, Y02, AB07, AJ01, AM13, AS13, AT20, AZ16, BF04, and BG02 (Supplementary Table S11). The PCR reactions were performed in an MJ Research Thermal Cycle PCR machine. The PCR products were analyzed by 1.5% agar gel. The assays were repeated three times and only repeatable bands were selected for further analysis. The banding patterns were then coded in binary form, and the phylogenetic tree was generated using the unweighted pair-group method with arithmetic averages (UPGMA) method. Jaccard similarity coefficient was calculated by the SIMQUAL routine and the significant bootstrap probability (>70%) was obtained for 1000 repetitions.

**Southern hybridization analysis.** The procedures of Southern blot were similar to a previously described method with some modifications<sup>46</sup>. Briefly, 3.1-kb *SphI* fragment of plasmid pZWavrXa7 were gel purified after digestion. This fragment was labeled with DIG DNA Labeling Mix (Roche) or AP Direct Labeling and Detection kit (GE) as a probe. Twenty micrograms (for DIG-labeling probe) or 500 ng (for AP-labeling probe) of genomic DNAs from *Xoo* isolates were digested with *SphI* and used for hybridization. Size markers were visualized by adding DIG/AP labels. Three independent repeats were performed in each sample and the phylogenetic tree was generated as described in RAPD analyses.

**CRISPR sequencing.** The last 10 spacers of the CRISPR region were sequenced as described by Semenova *et al.*<sup>26</sup>. Briefly, leader fragments were amplified from genomic DNA with primers specific to *Cas2* gene, RP: CAGGCTCGCGAAATTTCCAAGTGAT, and CRISPR repeat region, FR: CTTGACGGTGTGATGGCC. Then, the fragments were sequenced by a primer-walking method. For XF89b CRISPR region, a clone from a 10 kb fragment library was selected and sequenced by Sanger sequencing. For XM9 CRISPR region, the sequence was obtained from Pacbio Sequel single molecule real-time sequencing and assembly result. For XO21 and XO604 isolates, CRISPR spacer sequence was obtained from a prior study<sup>26</sup>. XO21 and XO604 are Taiwanese isolates reported in prior studies<sup>47,48</sup>. XO604 is also in our sample collection. The last 10 spacers in XO604 CRISPR region were sequenced by Sanger sequencing and showed identical sequence as the prior study<sup>26</sup>. For the other isolates, the sequences were obtained from Quibod *et al.*<sup>17</sup>.

**Optical mapping.** Genomic DNA of *Xoo* XF89b and XM42 were isolated and digested with BamHI in optical mapping procedures as described by Kotewicz *et al.*<sup>49</sup>. The detailed procedure followed the manufacturer's instruction established by Yourgene Bioscience, Taiwan. After the assembly of the restriction map, the comparison of *Xoo* MAFF311018, KACC10331, PXO99A, XF89b, XM42, and *Xoc* BLS256 were analyzed with MapSolver software. Also, the *Xoo* XF89b contigs from the next-generation sequencing result was aligned to optical mapping results with the same software.

**Genome sequencing.** *Xoo* XF89b genomic DNA was extracted and randomly sheared. The size of 0.45 kb and 3 kb of sheared DNA were selected on gel and ligated with adaptors. The DNA libraries were sequenced by Illumina GA IIX and MiSeq sequence protocols established and modified by the High Throughput Genomics Core, Academia Sinica. The GA IIX sequencer yielded 33,974,650 reads of the 80 bp mate pair dataset (~1000x), and the MiSeq sequencer yielded 9,509,144 reads of the 250 bp paired end dataset (~480x). The datasets were then assembled by ALLPATHS-LG software. The sheared gDNA was also constructed and sequenced by Roche 454

sequencer. The results yielded 323,730 trimmed reads with an average length of 741.72 bp (~50x). The 454 dataset was assembled by Newbler 2.7.0. The contigs generated from ALLPATHS-LG and Newbler were further aligned and merged by Minimus2 in the AMOS package. The combination of datasets resulted in 224 non-repeated contigs. The order of the contigs was defined with the optical map of *Xoo* XF89b *in silico*. For the highly repeated regions, TAL gene clusters and a CRISPR region, gDNA were partially digested by *SphI*, and 10 kb gDNA fragments were selected on gel. The selected fragments were purified by GenepFlow Gel/PCR Kit (Geneaid, Taiwan). The purified DNA fragments were then repaired by mung bean nuclease. The blunt-end DNA fragments were constructed into Lucigen BigEasy v2.0 Linear Cloning System (pJAZZ-OK blunt vector) according to the manufacturer's instruction. *E. coli* colonies on kanamycin selection plates containing plasmids of TAL genes or a CRISPR region were selected by a PCR screen using locus-specific primers (Supplementary Table S11). The plasmids and gaps of the genome were then filled by PCR and the Sanger sequencing method. The genome sequence is available at NCBI, accession: NZ\_CP011532.1. For *Xoo* XM9, gDNA were sequenced by PacBio Sequel sequencing protocol at Genomics, Taiwan. The raw reads from Sequel sequencing were then assembled by pbsmrtpipe v.0.51.2 and Falcon v.0.3<sup>50</sup>. The assembly generates a single contig. This contig was further corrected by raw reads using Arrow. The annotation of the genome is performed with NCBI prokaryotic genome annotation pipeline v.2.10. The schematic representation of the indel heat map was generated by Circos<sup>51</sup>. In-del regions and a cladogram were generated by Mauve v.2.4.0<sup>52</sup>. The XM9 genome sequence is available at NCBI, accession: CP020334.1. The TALE analysis in this study was mainly based on AnnoTALE classification<sup>33</sup>. RVDs were extracted by the software or derived from a prior study<sup>17</sup> with manual checking. DisTALE analysis was using QueTAL software<sup>34</sup>.

**RNA-seq and transcriptome analyses.** Fourteen-day-old rice cultivar TNG67 was inoculated with *Xoo* XE3, XF89b, and XM9 using the leaf clipping method. Briefly, *Xoo* isolates were grown in 1/2 TSB medium overnight, and the culture of *Xoo* was diluted to OD<sub>600</sub> = 1.0 with the same medium. The cultured mediums containing *Xoo* isolates and the 1/2 TSB buffer as a control group were used in the assay. Leaf blades from the clipping site to 5 cm below the clipping site were collected at 7-days post-inoculation. Leaf RNA of the infected rice was extracted by TRIzol RNA isolation reagent (Invitrogen) followed by DNase I digestion. Purified RNAs were further processed by Novogene Illumina Hiseq 4000, PE 150 standard protocol. RNA-seq raw data were further analyzed using CLC workbench software v.10.1.2 (Qiagen). All the configurations for transcriptome analyses by CLC workbench were set to default values. For sequencing data mapping, a rice cDNA library (MSU7) was used as a reference dataset. For the comparison between sample sets, 75<sup>th</sup> percentile of total mapped reads was used for normalization. The statistics of the transcriptomic data from two biological replicates was performed using the empirical analysis of DGE in CLC workbench. The list of differentially expressed genes was selected by the genes with fold change  $\geq 2$  and  $P < 0.05$  in either one of the three control and *Xoo*-inoculated sample sets (XE3 vs. mock, XF89b vs. mock or XM9 vs. mock). The classification of genes in the heatmap was grouped using R v. 3.3.3 with ggplot2 v. 2.2.1<sup>53</sup>. GO enrichment analysis of differentially expressed genes was performed by agriGO v. 2.0<sup>54</sup> with the Singular Enrichment Analysis (SEA) tool and the *Oryza sativa* MSU7.0 nonTE dataset.

## Availability of Data and Materials

All complete genome assemblies and corresponding genome annotations are available at the National Center for Biotechnology Information (NCBI). BioProject: PRJNA284661 (XF89b) and PRJNA378144 (XM9). BioSample: SAMN03729481 (XF89b) and SAMN06480557 (XM9). SRA: XF89b: SRR6510602, SRR6510603, and SRR6510604. XM9: SRR5319797. Accession: XF89b: NZ\_CP011532.1, XM9: CP020334.1. Rice RNA-seq data: BioProject: PRJNA378144. BioSample: SAMN08384324. SRA: SRR6513657.

## References

- Leyns, F., De Cleene, M., Swings, J.-G. & De Ley, J. The host range of the genus *Xanthomonas*. *The Botanical Review* **50**, 308–356, <https://doi.org/10.1007/BF02862635> (1984).
- Nino-Liu, D. O., Ronald, P. C. & Bogdanove, A. J. *Xanthomonas oryzae* pathovars: model pathogens of a model crop. *Molecular plant pathology* **7**, 303–324, <https://doi.org/10.1111/j.1364-3703.2006.00344.x> (2006).
- Mishra, D. *et al.* Pathotype and genetic diversity amongst Indian isolates of *Xanthomonas oryzae* pv. *oryzae*. *PLoS One* **8**, e81996, <https://doi.org/10.1371/journal.pone.0081996> (2013).
- Shanti, M. L. *et al.* Understanding the Bacterial Blight Pathogen-Combining Pathotyping and Molecular Marker Studies. *International Journal of Plant Pathology* **1**, 58–68, <https://doi.org/10.3923/ijpp.2010.58.68> (2010).
- Adhikari, T. B., Basnyat, R. C. & Mew, T. W. Virulence of *Xanthomonas oryzae* pv. *oryzae* on Rice Lines Containing Single Resistance Genes and Gene Combinations. *Plant Disease* **83**, 46–50, <https://doi.org/10.1094/pdis.1999.83.1.46> (1999).
- Reddy, A. P. K. Relationship of Bacterial Leaf Blight Severity to Grain Yield of Rice. *Phytopathology* **69**, 967, <https://doi.org/10.1094/Phyto-69-967> (1979).
- Hashioka, Y. Bacterial leaf blight of rice and its control [in Japanese]. *Agriculture and Horticulture* **26**, 644–648 (1951).
- Chang, Y. C. & Hsieh, L. J. Screen of Rice Varieties for Resistance to *Xanthomonas oryzae* pv. *oryzae* in Taiwan [in Chinese]. *Journal of Agricultural Research of China* **48**, 101–109 (1999).
- Hsieh, L. J., Chang, Y. C. & Hsieh, T. F. Susceptibility of Rice Varieties to Pale-Yellow-Leaf Type of Bacterial Leaf Blight Caused by *Xanthomonas oryzae* pv. *oryzae* [in Chinese]. *Journal of Taiwan Agricultural Research* **56**, 1–10 (2007).
- Chien, C. C. & Shieh, L. J. Pathogenic Groups of *Xanthomonas campestris* pv. *oryzae* in Taiwan [in Chinese]. *Journal of Agricultural Research of China* **38**, 216–228 (1989).
- Liao, Y. M. & Chien, C. C. Pathotypes of *Xanthomonas campestris* pv. *oryzae* in Taiwan [in Chinese]. *Journal of Agricultural Research of China* **31**, 321–333 (1982).
- Deng, W.-L. *et al.* Genotypic and Pathotypic Diversity of *Xanthomonas oryzae* pv. *oryzae* Strains in Taiwan. *Journal of Phytopathology* **164**, 745–759, <https://doi.org/10.1111/jph.12495> (2016).
- Adhikari, T. B. *et al.* Genetic Diversity of *Xanthomonas oryzae* pv. *oryzae* in Asia. *Appl Environ Microbiol* **61**, 966–971 (1995).
- Poulin, L. *et al.* New multilocus variable-number tandem-repeat analysis tool for surveillance and local epidemiology of bacterial leaf blight and bacterial leaf streak of rice caused by *Xanthomonas oryzae*. *Appl Environ Microbiol* **81**, 688–698, <https://doi.org/10.1128/AEM.02768-14> (2015).

15. Zhao, K. *et al.* Genome-wide association mapping reveals a rich genetic architecture of complex traits in *Oryza sativa*. *Nat Commun* **2**, 467, <https://doi.org/10.1038/ncomms1467> (2011).
16. Salzberg, S. L. *et al.* Genome sequence and rapid evolution of the rice pathogen *Xanthomonas oryzae* pv. *oryzae* PXO99A. *BMC genomics* **9**, 204, <https://doi.org/10.1186/1471-2164-9-204> (2008).
17. Quibod, I. L. *et al.* Effector Diversification Contributes to *Xanthomonas oryzae* pv. *oryzae* Phenotypic Adaptation in a Semi-Isolated Environment. *Sci Rep* **6**, 34137, <https://doi.org/10.1038/srep34137> (2016).
18. Ryba-White, M., Sakthivel, N., Yun, C., White, F. & Leach, J. E. Identification and characterization of IS1112 and IS1113 insertion element sequences in *Xanthomonas oryzae* pv. *oryzae*. *DNA Sequence* **16**, 75–79, <https://doi.org/10.1080/10425170500050932> (2009).
19. Hu, J., Zhang, Y., Qian, W. & He, C. Avirulence gene and insertion element-based RFLP as well as RAPD markers reveal high levels of genomic polymorphism in the rice pathogen *Xanthomonas oryzae* pv. *oryzae*. *Syst Appl Microbiol* **30**, 587–600, <https://doi.org/10.1016/j.syapm.2007.08.003> (2007).
20. Yashitola, J., Krishnaveni, D., Reddy, A. P. K. & Sonti, R. V. Genetic Diversity within the Population of *Xanthomonas oryzae* pv. *oryzae* in India. *Phytopathology* **87**, 760–765, <https://doi.org/10.1094/phyto.1997.87.7.760> (1997).
21. Yang, B. & White, F. F. Diverse members of the AvrBs3/PthA family of type III effectors are major virulence determinants in bacterial blight disease of rice. *Molecular plant-microbe interactions: MPMI* **17**, 1192–1200, <https://doi.org/10.1094/MPMI.2004.17.11.1192> (2004).
22. Lee, B. M. The genome sequence of *Xanthomonas oryzae* pathovar *oryzae* KACC10331, the bacterial blight pathogen of rice. *Nucleic Acids Research* **33**, 577–586, <https://doi.org/10.1093/nar/gki206> (2005).
23. Ochiai, H., Inoue, Y., Takeya, M., Sasaki, A. & Kaku, H. Genome Sequence of *Xanthomonas oryzae* pv. *oryzae* Suggests Contribution of Large Numbers of Effector Genes and Insertion Sequences to Its Race Diversity. *Japan Agricultural Research Quarterly: JARQ* **39**, 275–287, <https://doi.org/10.6090/jarq.39.275> (2005).
24. Shyntum, D., Venter, S., Moleleki, L., Toth, I. & Coutinho, T. Comparative genomics of type VI secretion systems in strains of *Pantoea ananatis* from different environments. *BMC genomics* **15**, 163, <https://doi.org/10.1186/1471-2164-15-163> (2014).
25. Makarova, K. S. *et al.* Evolution and classification of the CRISPR-Cas systems. *Nat Rev Microbiol* **9**, 467–477, <https://doi.org/10.1038/nrmicro2577> (2011).
26. Semenova, E., Nagornykh, M., Pyatnitskiy, M., Artamonova, I. I. & Severinov, K. Analysis of CRISPR system function in plant pathogen *Xanthomonas oryzae*. *FEMS Microbiol Lett* **296**, 110–116, <https://doi.org/10.1111/j.1574-6968.2009.01626.x> (2009).
27. Horvath, P. & Barrangou, R. CRISPR/Cas, the Immune System of Bacteria and Archaea. *Science* **327**, 167–170, <https://doi.org/10.1126/science.1179555> (2010).
28. Ryan, R. P. *et al.* Pathogenomics of *Xanthomonas*: understanding bacterium-plant interactions. *Nat Rev Microbiol* **9**, 344–355, <https://doi.org/10.1038/nrmicro2558> (2011).
29. Dong, T. G., Ho, B. T., Yoder-Himes, D. R. & Mekalanos, J. J. Identification of T6SS-dependent effector and immunity proteins by Tn-seq in *Vibrio cholerae*. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 2623–2628, <https://doi.org/10.1073/pnas.1222783110> (2013).
30. Bondage, D. D., Lin, J. S., Ma, L. S., Kuo, C. H. & Lai, E. M. VgrG C terminus confers the type VI effector transport specificity and is required for binding with PAAR and adaptor-effector complex. *Proceedings of the National Academy of Sciences of the United States of America* **113**, E3931–3940, <https://doi.org/10.1073/pnas.1600428113> (2016).
31. Song, C. & Yang, B. Mutagenesis of 18 type III effectors reveals virulence function of XopZ(PXO99) in *Xanthomonas oryzae* pv. *oryzae*. *Mol Plant Microbe Interact* **23**, 893–902, <https://doi.org/10.1094/MPMI-23-7-0893> (2010).
32. Boch, J. *et al.* Breaking the Code of DNA Binding Specificity of TAL-Type III Effectors. *Science* **326**, 1509–1512, <https://doi.org/10.1126/science.1178811> (2009).
33. Grau, J. *et al.* AnnoTAL: bioinformatics tools for identification, annotation, and nomenclature of TALEs from *Xanthomonas* genomic sequences. *Sci Rep* **6**, 21077, <https://doi.org/10.1038/srep21077> (2016).
34. Perez-Quintero, A. L. *et al.* QueTAL: a suite of tools to classify and compare TAL effectors functionally and phylogenetically. *Frontiers in plant science* **6**, 545, <https://doi.org/10.3389/fpls.2015.00545> (2015).
35. Read, A. C. *et al.* Suppression of Xo1-Mediated Disease Resistance in Rice by a Truncated, Non-DNA-Binding TAL Effector of *Xanthomonas oryzae*. *Frontiers in Plant Science* **7**, <https://doi.org/10.3389/fpls.2016.01516> (2016).
36. Ji, Z. *et al.* Interfering TAL effectors of *Xanthomonas oryzae* neutralize R-gene-mediated plant disease resistance. *Nature Communications* **7**, <https://doi.org/10.1038/ncomms13435> (2016).
37. Gu, K. *et al.* R gene expression induced by a type-III effector triggers disease resistance in rice. *Nature* **435**, 1122–1125, <https://doi.org/10.1038/nature03630> (2005).
38. Sidhu, G. S., Khush, G. S. & Mew, T. W. Genetic analysis of bacterial blight resistance in seventy-four cultivars of rice, *Oryza sativa* L. *Theoretical and Applied Genetics* **53**, 105–111, <https://doi.org/10.1007/bf00272687> (1978).
39. Ogawa, T., Yamamoto, T., Khush, G. S. & Mew, T.-W. Breeding of Near-Isogenic Lines of Rice with Single Genes for Resistance to Bacterial Blight Pathogen (*Xanthomonas campestris* pv. *oryzae*). *Ikushugaku zasshi* **41**, 523–529, <https://doi.org/10.1270/jsbbs1951.41.523> (1991).
40. Vera Cruz, C. M. *et al.* Predicting durability of a disease resistance gene based on an assessment of the fitness loss and epidemiological consequences of avirulence gene mutation. *Proceedings of the National Academy of Sciences* **97**, 13500–13505, <https://doi.org/10.1073/pnas.250271997> (2000).
41. Bogdanove, A. J., Schornack, S. & Lahaye, T. TAL effectors: finding plant genes for disease and defense. *Current opinion in plant biology* **13**, 394–401, <https://doi.org/10.1016/j.pbi.2010.04.010> (2010).
42. Streubel, J. *et al.* Five phylogenetically close rice SWEET genes confer TAL effector-mediated susceptibility to *Xanthomonas oryzae* pv. *oryzae*. *New Phytol* **200**, 808–819, <https://doi.org/10.1111/nph.12411> (2013).
43. Yu, Y. H., Lu, Y., He, Y. Q., Huang, S. & Tang, J. L. Rapid and efficient genome-wide characterization of *Xanthomonas* TAL effector genes. *Sci Rep* **5**, 13162, <https://doi.org/10.1038/srep13162> (2015).
44. Blainvillain-Baufume, S. *et al.* Targeted promoter editing for rice resistance to *Xanthomonas oryzae* pv. *oryzae* reveals differential activities for SWEET14-inducing TAL effectors. *Plant Biotechnol J* **15**, 306–317, <https://doi.org/10.1111/pbi.12613> (2017).
45. Chen, L. Q. *et al.* Sugar transporters for intercellular exchange and nutrition of pathogens. *Nature* **468**, 527–532, <https://doi.org/10.1038/nature09606> (2010).
46. Yang, B., Zhu, W., Johnson, L. B. & White, F. F. The virulence factor AvrXa7 of *Xanthomonas oryzae* pv. *oryzae* is a type III secretion pathway-dependent nuclear-localized double-stranded DNA-binding protein. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 9807–9812, <https://doi.org/10.1073/pnas.170286897> (2000).
47. Lin, N. T. *et al.* Characterization of two novel filamentous phages of *Xanthomonas*. *Journal of General Virology* **75**, 2543–2547, <https://doi.org/10.1099/0022-1317-75-9-2543> (1994).
48. Cheng, C.-M., Tu, J., Yang, C.-C. & Kuo, T.-T. Specific protein phosphorylation induced in *Xanthomonas campestris* pv. *oryzae* by bacteriophage Xp12. *Archives of Microbiology* **161**, 281–285, <https://doi.org/10.1007/bf00303581> (1994).
49. Kotewicz, M. L., Jackson, S. A., LeClerc, J. E. & Cebula, T. A. Optical maps distinguish individual strains of *Escherichia coli* O157:H7. *Microbiology* **153**, 1720–1733, <https://doi.org/10.1099/mic.0.2006/004507-0> (2007).
50. Chin, C.-S. *et al.* Phased diploid genome assembly with single-molecule real-time sequencing. *Nature Methods* **13**, 1050–1054, <https://doi.org/10.1038/nmeth.4035> (2016).

51. Krzywinski, M. *et al.* Circos: an information aesthetic for comparative genomics. *Genome research* **19**, 1639–1645, <https://doi.org/10.1101/gr.092759.109> (2009).
52. Darling, A. E., Mau, B. & Perna, N. T. Progressive Mauve: multiple genome alignment with gene gain, loss and rearrangement. *PloS one* **5**, e111147, <https://doi.org/10.1371/journal.pone.0011147> (2010).
53. Ihaka, R. & Gentleman, R. R. A Language for Data Analysis and Graphics. *Journal of Computational and Graphical Statistics* **5**, 299, <https://doi.org/10.2307/1390807> (1996).
54. Tian, T. *et al.* agriGOv2.0: a GO analysis toolkit for the agricultural community, 2017 update. *Nucleic Acids Research* **45**, W122–W129, <https://doi.org/10.1093/nar/gkx382> (2017).
55. Perez-Quintero, A. L. *et al.* An improved method for TAL effectors DNA-binding sites prediction reveals functional convergence in TAL repertoires of *Xanthomonas oryzae* strains. *PloS one* **8**, e68464, <https://doi.org/10.1371/journal.pone.0068464> (2013).

## Acknowledgements

We thank the Taiwan Agricultural Research Institute, Council of Agriculture, Executive Yuan, for providing all the *Xoo* isolates in this study. Dr. Frank F. White kindly provided a pZWavrXa7 plasmid for probing TALE genes in Southern blot assays. We also thank Dr. Chan-Sen Wang for discussion. This work was supported by a grant (AS-103-SS-A03) from the Center for Sustainability Science, Academia Sinica, Taiwan.

## Author Contributions

C.-C.C. and M.-C.S. conducted all the experiments, results, discussion, and manuscript preparation. M.-Y.C. helped to select TALE gene-containing clones from a 20 kb DNA fragment library of *Xoo* XF89b and sequence all the TALE genes from *Xoo* XF89b. C.-Y.C. was responsible for the assembly of reads generated from the Illumina sequencing technique.

## Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-018-36575-x>.

**Competing Interests:** The authors declare no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019