

Variation and inheritance of the *Xanthomonas raxX-raxSTAB* gene cluster required for activation of XA21-mediated immunity

FURONG LIU^{1,†}, MEGAN MCDONALD^{2,†}, BENJAMIN SCHWESSINGER^{1,2,†}, ANNA JOE¹, RORY PRUITT^{1,†}, TERESA ERICKSON¹, XIUXIANG ZHAO³, VALLEY STEWART^{1,3} AND PAMELA C. RONALD^{1,*}

¹Department of Plant Pathology and the Genome Center, University of California, Davis, CA 95616, USA

²Research School of Biology, Australian National University, Canberra 0200, Australia

³Department of Microbiology & Molecular Genetics, University of California, Davis, CA 95616, USA

SUMMARY

The rice XA21-mediated immune response is activated on recognition of the RaxX peptide produced by the bacterium *Xanthomonas oryzae* pv. *oryzae* (Xoo). The 60-residue RaxX precursor is post-translationally modified to form a sulfated tyrosine peptide that shares sequence and functional similarity with the plant sulfated tyrosine (PSY) peptide hormones. The 5-kb *raxX-raxSTAB* gene cluster of Xoo encodes RaxX, the RaxST tyrosyl-protein sulfotransferase, and the RaxA and RaxB components of a predicted type I secretion system. To assess *raxX-raxSTAB* gene cluster evolution and to determine its phylogenetic distribution, we first identified *rax* gene homologues in other genomes. We detected the complete *raxX-raxSTAB* gene cluster only in *Xanthomonas* spp., in five distinct lineages in addition to *X. oryzae*. The phylogenetic distribution of the *raxX-raxSTAB* gene cluster is consistent with the occurrence of multiple lateral (horizontal) gene transfer events during *Xanthomonas* speciation. RaxX natural variants contain a restricted set of missense substitutions, as expected if selection acts to maintain peptide hormone-like function. Indeed, eight RaxX variants tested all failed to activate the XA21-mediated immune response, yet retained peptide hormone activity. Together, these observations support the hypothesis that the XA21 receptor evolved specifically to recognize Xoo RaxX.

Keywords: *raxX-raxSTAB* gene cluster, XA21, Xoo, plant immunity.

INTRODUCTION

Host receptors activate innate immunity pathways on pathogen recognition (Ronald and Beutler, 2010). The gene encoding

the rice XA21 receptor kinase (Song *et al.*, 1995) confers resistance against most strains of the gamma-proteobacterium *Xanthomonas oryzae* pv. *oryzae* (Xoo) (Wang *et al.*, 1996). Xoo causes bacterial leaf blight disease of rice, which seriously constrains yields in Africa, Asia and South America. This well-studied XA21–Xoo interaction provides a basis from which to understand the molecular and evolutionary mechanisms of host–microbe interactions.

Four Xoo genes required for the activation of XA21-mediated immunity are located in the *raxX-raxSTAB* gene cluster (Fig. 1). The 60-residue RaxX predicted precursor protein undergoes sulfation by the RaxST tyrosylprotein sulfotransferase at residue tyrosine-41 (Tyr-41) (Pruitt *et al.*, 2015). The RaxB component of the RaxAB type I secretion complex (da Silva *et al.*, 2004) further processes the sulfated RaxX precursor by removing its double-glycine leader peptide prior to secretion (Holland *et al.*, 2016; Luu *et al.*, 2018). Core ('housekeeping') genes encode the predicted outer membrane TolC channel for the RaxAB complex (da Silva *et al.*, 2004), as well as enzymes to assimilate sulfate into 3'-phosphoadenosine 5'-phosphosulfate (PAPS) (Shen *et al.*, 2002), the sulfodonor for the RaxST sulfotransferase (Han *et al.*, 2012).

In both plants and animals, the post-translational modification catalysed by tyrosylprotein sulfotransferase is restricted to a subset of cell surface and secreted proteins that influence a variety of eukaryotic physiological processes (Matsubayashi, 2014; Stone *et al.*, 2009). For example, tyrosine sulfation of the chemokine receptors CCR5 and CXCR4 is essential for their functions, including as coreceptors for the human immunodeficiency virus gp120 envelope glycoprotein (Farzan *et al.*, 1999; Kleist *et al.*, 2016). In plants, sulfated tyrosine peptides influence cellular proliferation and expansion in root growth and/or plant immune signalling (Matsubayashi, 2014; Tang *et al.*, 2017). In contrast with these and other examples of protein tyrosine sulfation in animals and plants, RaxX sulfation by the RaxST enzyme is the only example of tyrosine sulfation documented in bacteria (Han *et al.*, 2012; Pruitt *et al.*, 2015).

*Correspondence: Email: pcronald@ucdavis.edu

†Co-first authors, in alphabetical order.

‡Present address: Department of Plant Biochemistry, University of Tübingen, Tübingen, Germany.

§Present address: Department of Plant Pathology, Shenyang Agricultural University, Shenyang 72076, China.

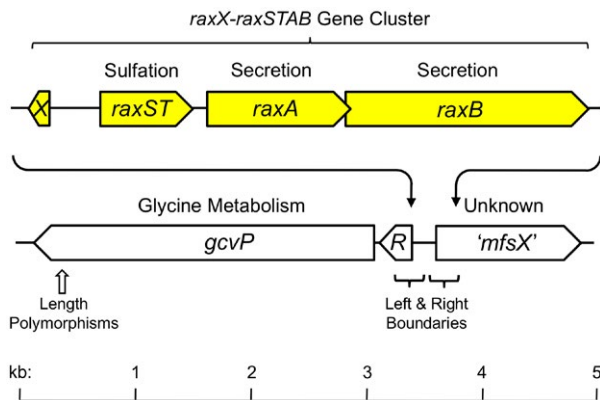


Fig. 1 The *raxX-raxSTAB* gene cluster. The *raxX-raxSTAB* gene cluster is located between the flanking *gcvP* and '*mfsX*' genes. Gene cluster acquisition through lateral gene transfer is hypothesized to occur by general recombination in the flanking *gcvP* and '*mfsX*' sequences as described in the text. Sequences at the left and right boundaries are shown in Fig. S2 (see Supporting Information). Sequences for length polymorphisms in the *gcvP* gene are shown in Fig. S3 (see Supporting Information).

Mature RaxX is predicted to comprise the carboxyl-terminal residues 40–60, numbered according to the precursor protein (Pruitt *et al.*, 2015, 2017). RaxX residues 40–52 share sequence similarity with mature plant sulfated tyrosine (PSY) peptide hormones (Amano *et al.*, 2007; Pruitt *et al.*, 2015, 2017) (Fig. 2). RaxX, like PSY1, can enhance root growth in diverse plant species (Pruitt *et al.*, 2017). The XA21-mediated response in rice requires residues 40–55, whereas growth stimulation requires only residues 40–52 (Pruitt *et al.*, 2015).

RaxX sequences are generally well conserved within different *Xanthomonas* species (Pruitt *et al.*, 2017). In *Xoo*, however, RaxX from strain IXO685, which evades XA21-mediated immunity, differs from active RaxX at the critical positions proline-44 (Pro-44) and Pro-48 (Pruitt *et al.*, 2015). Nevertheless, this RaxX protein stimulates root growth, as do two other RaxX Pro-48 variants from other *Xanthomonas* spp. (Pruitt *et al.*, 2017).

These results suggest that RaxX recognition by XA21 is restrained by different sequence and length requirements compared with its recognition by the root growth-promoting receptor(s) for PSY hormone(s). It also suggests that the recognition of RaxX by XA21 is specific to *Xoo*, whereas PSY mimicry is a general feature of RaxX from other *Xanthomonas* spp. Accordingly, we have hypothesized previously that PSY hormone mimicry is the original function of RaxX, whereas immune recognition by XA21 evolved later in response to *Xoo* (Pruitt *et al.*, 2017).

Here, we report tests of two general predictions derived from this hypothesis. The first prediction, that PSY hormone mimicry is broadly selective, is supported here by the presence of the *raxX-raxSTAB* gene cluster in a range of *Xanthomonas* spp., and by the ability of all RaxX variants tested to stimulate root growth in an assay for PSY function. The second prediction, that

recognition by XA21 is restricted to *X. oryzae* lineages, is validated here by the observation that XA21-mediated immunity is not activated by RaxX variants from other *Xanthomonas* spp. These results illustrate how a pathogen protein has evolved to retain its ability to modulate host physiology without being recognized by the host immune system.

RESULTS

The *raxX-raxSTAB* gene cluster is present in a subset of *Xanthomonas* spp.

We searched databases at the National Center for Biotechnology Information (NCBI) to identify bacterial genomes with the *raxX-raxSTAB* gene cluster. We found the intact *raxX-raxSTAB* gene cluster exclusively in *Xanthomonas* spp., and detected it in more than 200 unique genome sequences (File S1, see Supporting Information) among 413 accessed through the RefSeq database (O'Leary *et al.*, 2016).

Xanthomonas taxonomy has undergone several changes over the years (Vauterin *et al.*, 2000; Young, 2008) (for a representative example, see Midha and Patil, 2014). At one point, many strains were denoted as pathovars of either *X. campestris* or *X. axonopodis*, but, today, over 20 species are distinguished, several with multiple pathovars (Rademaker *et al.*, 2005; Vauterin *et al.*, 1995). Because many of the genome sequences we examined are from closely related strains, in some cases associated with different species designations, we constructed a whole-genome phylogenetic tree as described in Experimental procedures in order to organize these sequences by relatedness (Fig. S1, see Supporting Information). The topology of the resulting tree shares broad similarity with several other *Xanthomonas* phylogenetic trees in defining relationships between well-sampled species (Ferreira-Tonin *et al.*, 2012; Gardiner *et al.*, 2014; Hauben *et al.*, 1997; Midha and Patil, 2014; Parkinson *et al.*, 2007, 2009; Rademaker *et al.*, 2005; Triplett *et al.*, 2015; Young, 2008).

To examine *raxX-raxSTAB* gene cluster organization and inheritance more closely, we selected 15 genomes from strains that represent the phylogenetic range of *Xanthomonas* spp. (Table 1 and Fig. S1). Where possible, we chose complete genome sequences that are accompanied by published descriptions. The close relative *Stenotrophomonas maltophilia*, which does not contain the *raxX-raxSTAB* gene cluster, served as the outgroup (Moore *et al.*, 1997).

To facilitate discussion, we represent phylogenetic relationships between these strains as a cladogram that emphasizes the relative positions of the *raxX-raxSTAB* gene cluster-positive lineages (Fig. 3). Six distinct *Xanthomonas* lineages contain the *raxX-raxSTAB* gene cluster, one being *X. oryzae*. A second lineage includes related strains, currently denoted as *X. vasicola* or *X. campestris* pv. *musacearum* (Aritua *et al.*, 2008); for concise

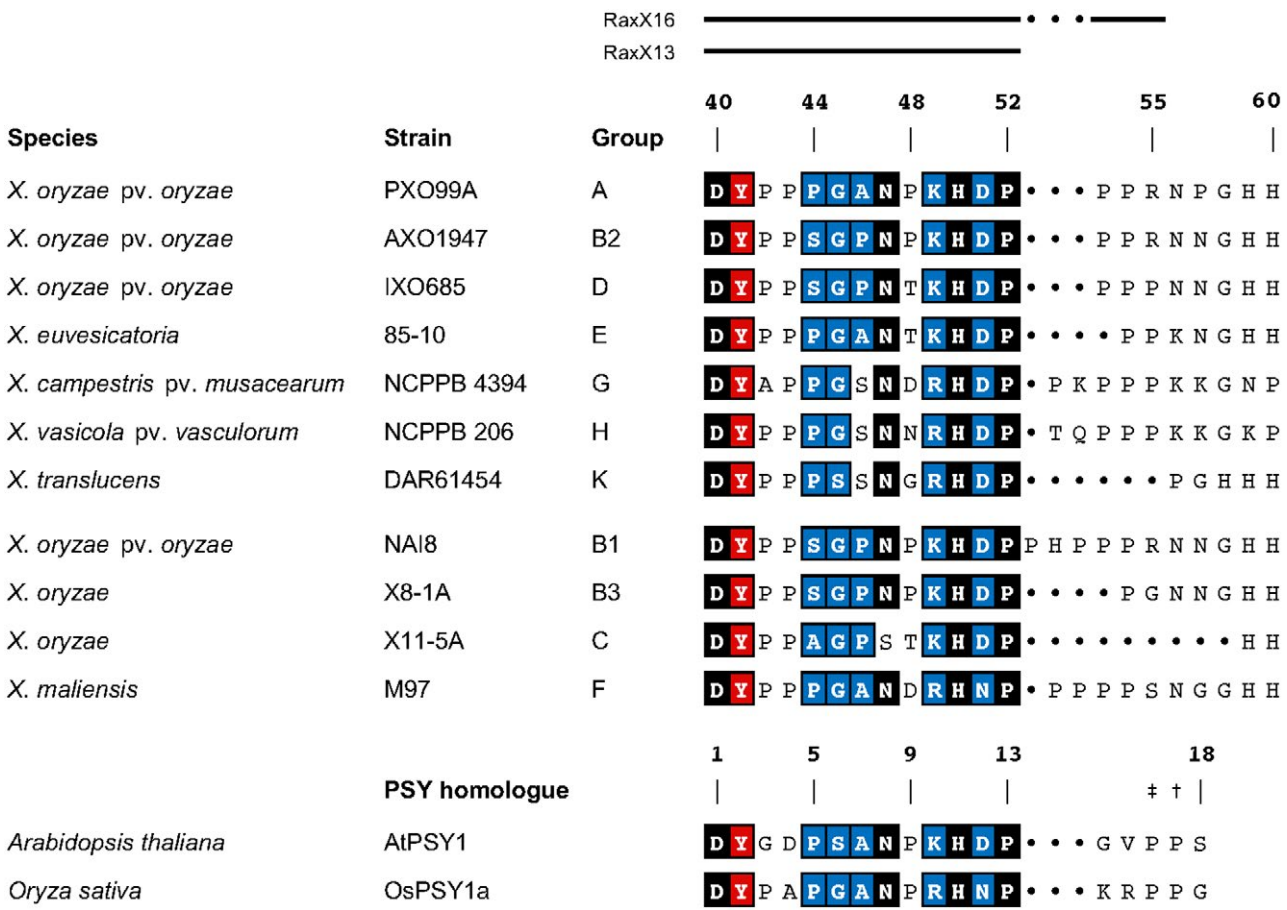


Fig. 2 RaxX and plant sulfated tyrosine (PSY) sequences. RaxX sequences show the presumed leader-cleaved forms of RaxX, numbered from the beginning of the precursor sequence. The extent of sequence comprising the RaxX16 and RaxX13 synthetic peptides is indicated above the alignment. Residues are shaded according to conservation in PSY sequences (Pruitt *et al.*, 2017): positions with nearly invariant residues are shaded black, and those with only two or three substitutions are shaded blue. The sulfated tyrosine (Tyr) residue is shaded red. Gaps are indicated by dots. Sequence groups are described elsewhere in detail (Pruitt *et al.*, 2017). The subgroups B1–B3 differ only in the carboxyl-terminal sequence beginning with residue 53. *Xanthomonas oryzae* strains X8-1A and X11-5A are non-pathogenic and therefore do not have pathovar designations. The mature form of *Arabidopsis thaliana* PSY1 (Amano *et al.*, 2007) and the corresponding region from *Oryza sativa* PSY1a (Amano *et al.*, 2007; Pruitt *et al.*, 2017) are shown for comparison. Residues Pro-16 and Pro-17 in AtPSY1 are both hydroxylated [†,‡], and Pro-16 is glycosylated with L-Ara₃ [‡] (Amano *et al.*, 2007).

presentation, we refer to these collectively as *X. vasicola*. The third lineage includes *X. euvesicatoria* and related species (Rademaker group 9.2; Barak *et al.*, 2016; Rademaker *et al.*, 2005). The fourth lineage includes strains denoted as *X. axonopodis*, such as pv. *manihotis* (Rademaker group 9.4; Mhedbi-Hajri *et al.*, 2013; Rademaker *et al.*, 2005). The fifth lineage includes *X. translucens* (Langlois *et al.*, 2017), within the distinct cluster of ‘early-branching’ species whose divergence from the remainder apparently occurred relatively early during *Xanthomonas* speciation (Parkinson *et al.*, 2007). The sixth lineage comprises *X. maliensis*, associated with, but non-pathogenic on, rice (Triplett *et al.*, 2015). Phylogenetic analyses place this species between the ‘early-branching’ species and the remainder (Triplett *et al.*, 2015). Notably, the *raxX-raxSTAB* gene cluster is absent from the group of strains classified as *X. citri* pathovars (Rademaker group

9.5; Bansal *et al.*, 2017; Rademaker *et al.*, 2005). These strains (some of which are denoted as *X. axonopodis* or *X. campestris*) cluster phylogenetically among four of the *raxX-raxSTAB* gene cluster-positive groups: *X. oryzae*, *X. vasicola*, *X. euvesicatoria* and *X. axonopodis* pv. *manihotis* (Midha and Patil, 2014; Rademaker *et al.*, 2005; Vauterin *et al.*, 1995). The simplest explanation for this pattern is that the *raxX-raxSTAB* gene cluster was lost from an ancestor of the *X. citri* lineage (Fig. 3); other explanations are not excluded.

Sequence conservation of the *raxX-raxSTAB* gene cluster suggests lateral gene transfer between *Xanthomonas* spp.

Both the organization and size of the *raxX-raxSTAB* gene cluster are conserved across all six lineages. To assess inheritance

Table 1 Reference strains for sequence comparisons.

Species	Strain	raxX-raxSTAB	Accession	Reference
<i>Stenotrophomonas maltophilia</i>	K279a	–	NC_010943.1	Crossman <i>et al.</i> (2008)
<i>Xanthomonas albilineans</i>	GPE PC73	–	NC_013722.1	Pieretti <i>et al.</i> (2015)
<i>X. arboricola</i> pv. <i>juglandis</i>	Xaj 417	–	NZ_CP012251.1	Pereira <i>et al.</i> (2015)
<i>X. axonopodis</i> pv. <i>manihotis</i>	UA536	+	NZ_AKEQ00000000	Bart <i>et al.</i> (2012)
<i>X. campestris</i> pv. <i>campestris</i>	ATCC 33913	–	NC_003902.1	da Silva <i>et al.</i> (2002)
<i>X. campestris</i> pv. <i>musacearum</i>	NCPBP 4392	+	NZ_AKB100000000.1	Wasukira <i>et al.</i> (2012)
<i>X. cannabis</i>	NCPBP 2877	–	NZ_JSZE00000000.1	Jacobs <i>et al.</i> (2015)
<i>X. citri</i> ssp. <i>citri</i>	306	–	NC_003919.1	da Silva <i>et al.</i> (2002)
<i>X. euvesicatoria</i>	85-10	+	NZ_CP017190.1	Thieme <i>et al.</i> (2005)
<i>X. fragariae</i>	LMG 25863	–	NZ_AJRZ00000000.1	Vandroemme <i>et al.</i> (2013)
<i>X. hyacinthi</i>	DSM 19077	–	JPLD00000000.1	Naushad <i>et al.</i> (2015)
<i>X. maliensis</i>	M97	+	NZ_AQPR00000000.1	Triplett <i>et al.</i> (2015)
<i>X. oryzae</i> pv. <i>oryzae</i>	PXO99 ^A	+	NC_010717.2	Salzberg <i>et al.</i> (2008)
<i>X. sacchari</i>	R1	–	NZ_CP010409.1	Studholme <i>et al.</i> (2011)
<i>X. translucens</i>	DAR61454	+	GCA_000334075.1	Gardiner <i>et al.</i> (2014)
<i>X. vesicatoria</i>	15b	–	NZ_JSXZ00000000.1	Vancheva <i>et al.</i> (2015)

patterns, we constructed a phylogenetic tree for the *raxX-raxSTAB* gene cluster (as the catenation of the four *rax* genes; Fig. 4) (Kuo and Ochman, 2009). The *rax* genes in *X. translucens*, in the early-branching group, cluster separately from their homologues in the other lineages. This finding is consistent with the hypothesis that *X. translucens* acquired the *raxX-raxSTAB* gene cluster relatively early during *Xanthomonas* speciation. For *X. maliensis*, the *raxX-raxSTAB* genes are most similar to those from *X. euvesicatoria* and the *X. axonopodis* pathovars *manihotis* and *phaseoli* (Fig. 4), even though the *X. maliensis* genome sequence itself is more distantly related (Fig. 3). This finding suggests that *X. maliensis* acquired the *raxX-raxSTAB* gene cluster relatively late during *Xanthomonas* speciation.

Boundaries flanking the *raxX-raxSTAB* gene cluster and adjacent genes suggest lateral gene transfer through general recombination

The *raxX-raxSTAB* gene cluster lies between two core (house-keeping) genes (Fig. 1). One, *gcvP*, encodes the pyridoxal-phosphate subunit of glycine dehydrogenase. An approximately 170-nucleotide riboswitch (*gcvR* in Fig. 1) controls GcvP protein synthesis in response to glycine (Mandal *et al.*, 2004). The other, '*mfsX*', encodes a major facilitator subfamily (MFS) transporter related to Bcr and CflA efflux proteins (da Silva *et al.*, 2004). Here, '*mfsX*' is only a provisional designation in the absence of functional characterization.

Comparing the *gcvP*–[*raxX-raxSTAB*]–'*mfsX*' region from the reference genomes reveals sharp boundaries flanking the position

of the *raxX-raxSTAB* gene cluster. On the left flank, substantial nucleotide identity spans the *gcvP* gene, the *gcvR* riboswitch and a predicted *gcvR* promoter –10 element (Mitchell *et al.*, 2003) (Fig. S2, see Supporting Information). On the right flank, identity begins shortly after the '*mfsX*' initiation codon. Accordingly, upstream sequence elements for initiating '*mfsX*' gene transcription (Mitchell *et al.*, 2003) and translation (Ma *et al.*, 2002) are conserved within, but not between, *raxX-raxSTAB* gene cluster-positive and cluster-negative sequences (Fig. S2).

Between these boundaries in genomes that lack the *raxX-raxSTAB* gene cluster, the compact (≤200 nucleotide) *gcvP*–'*mfsX*' intergenic sequence is modestly conserved in most genomes (about 60%–80% overall identity; Fig. S2). Much of this identity comes from the '*mfsX*' potential transcription and translation initiation sequences described above. The overall intergenic sequence is less conserved in the early-branching species (*X. albilineans*, *X. hyacinthi* and *X. sacchari*), displaying about 50%–65% overall identity.

We hypothesize that *raxX-raxSTAB* gene cluster phylogenetic distribution results from general recombination between conserved genes flanking each side (e.g. in or beyond the *gcvP* and '*mfsX*' genes). Two observations are consistent with this hypothesis. First, we observed that the sequences flanking the *raxX-raxSTAB* gene cluster are different from the *gcvP*–'*mfsX*' intergenic sequence in genomes that lack the *raxX-raxSTAB* gene cluster (Fig. S2). This is inconsistent with a mechanism through which the *raxX-raxSTAB* gene cluster integrated into the *gcvP*–'*mfsX*' intergenic sequence during lateral gene transfer events.

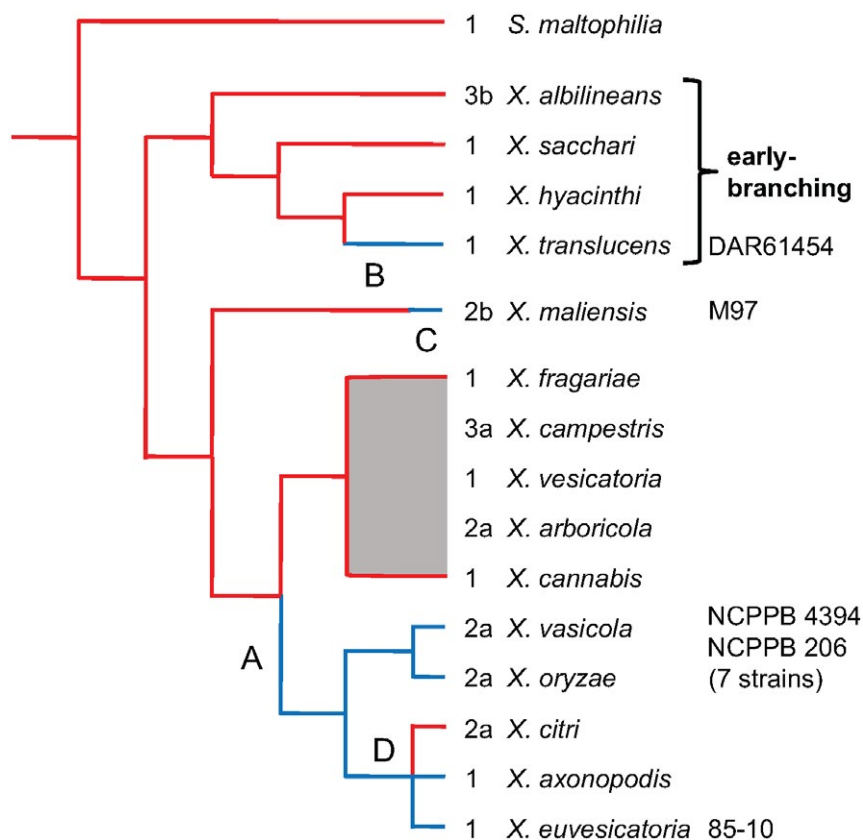


Fig. 3 Model for *raxX-raxSTAB* inheritance during *Xanthomonas* speciation. The *Xanthomonas* spp. cladogram is based on published phylogenetic trees (see text for references). Red lines depict lineages for strains that lack the *raxX-raxSTAB* gene cluster, whereas blue lines depict those that carry the cluster. Numbers indicate *gcvP* length polymorphism in each species (Fig. S3, see Supporting Information). Hypothetical events are: A, formation of the *raxX-raxSTAB* gene cluster; B, lateral gene transfer to *X. translucens* relatively early during speciation (indicated by the long blue line); C, lateral gene transfer to *X. maliensis* relatively late during speciation (indicated by the short blue line); D, loss from *X. citri*. Strain numbers denote sources of RaxX proteins chosen for functional tests, as described in the text.

The second observation consistent with lateral gene transfer via general recombination is that *gcvP* length polymorphisms (Figs 1 and S3, see Supporting Information) do not align with *Xanthomonas* phylogenetic relationships (Fig. 3). Inheritance patterns such as this often result from general recombination in the vicinity (Nelson *et al.*, 1997).

Notably, this *gcvP*–*mfsX* intergenic region is also conserved in the *X. citri* lineage (Fig. S2). If the *raxX-raxSTAB* gene cluster was lost during formation of this lineage (see above), general recombination would replace the resident *raxX-raxSTAB* gene cluster with a donor conserved *gcvP*–*mfsX* region.

***raxST*, but not *raxX*, homologues are present in genomes from diverse bacterial species**

Our GenBank database searches identified *raxX* homologues and the *raxX-raxSTAB* gene cluster only in *Xanthomonas* spp. However, these searches did identify *raxST* homologues encoding proteins with about 40% identity to, and approximately the

same length as, the *Xoo* RaxST protein. These sequences include the PAPS binding motifs that define sulfotransferase activity (Negishi *et al.*, 2001; da Silva *et al.*, 2004). Regardless of its current function, a *raxST* homologue potentially could evolve to encode tyrosylprotein sulfotransferase activity.

None of these *raxST* homologues is associated with a *raxX* homologue, and most are also not associated with *raxA* or *raxB* homologues. Presumably, the enzymes by these *raxST* homologues act on substrates other than RaxX. These *raxST* homologues support the hypothesis that the *raxSTAB* cluster arose from a new combination of pre-existing *raxST*, *raxA* and *raxB* homologues. Proteolytic maturation and ATP-dependent peptide secretion systems are broadly distributed, and so *raxA* and *raxB* homologues are plentiful in bacterial genomes (Holland *et al.*, 2016).

These *raxST* homologues occur in diverse genetic contexts in a range of bacterial phyla, including Proteobacteria and Cyanobacteria (Fig. S4, see Supporting Information). Nevertheless, for most species represented by multiple genome sequences, the

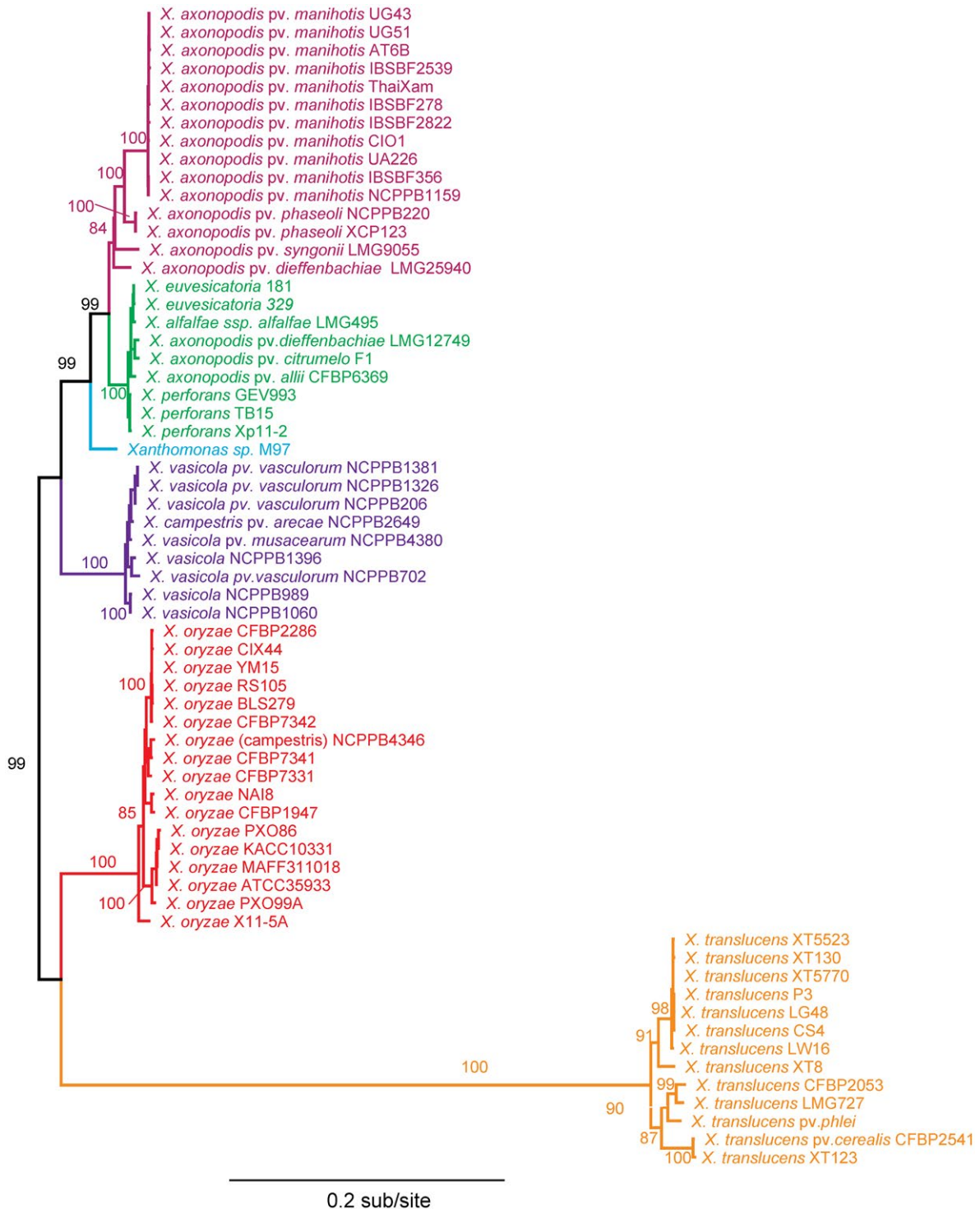


Fig. 4 Phylogenetic tree for *raxX-raxSTAB* nucleotide sequences. The best scoring maximum likelihood tree for the catenated *raxA*, *raxB*, *raxX* and *raxST* coding sequences. Numbers shown on the branches represent the proportion of branches supported by 10 000 bootstrap replicates (0–100). Bootstraps are not shown for branches with less than 50% support, or for branches too short to easily distinguish. Species names are coloured according to phylogenetic group.

raxST homologue was detected in a minority of individuals, and so it is not part of the core genome in these strains. Moreover, relationships between species in a *raxST* gene phylogenetic tree bear no resemblance to those in the overall tree of bacterial species. For example, in the *raxST* gene tree, sequences from Cyanobacteria are flanked on both sides by sequences from Proteobacteria (Fig. S4). Together, these findings provide evidence for lateral gene transfer of *raxST* homologues (Kuo and Ochman, 2009).

RaxX protein sequence variants representing all six *raxX-raxSTAB* gene cluster-positive lineages

RaxX protein sequences from diverse *Xanthomonas* spp. assort into several sequence groups differentiated by polymorphisms within the predicted mature peptide sequence (Fig. 2) (Pruitt *et al.*, 2017). Many of these groups are subdivided further according to polymorphisms in the predicted leader protein sequence (residues

1–39) or carboxyl-terminal region distal to residue Pro-52. Most leader polymorphisms lie between residues 2 and 24, and are unlikely to affect the function of the mature RaxX protein. Here, we only consider polymorphisms in the predicted mature form.

To assess the function of RaxX variants, we focused on frequently observed variants in species represented by numerous genome sequences (Fig. S1). These include sequence groups A, B and D from *X. oryzae* pv. *oryzae* and *X. oryzae* pv. *oryzicola*, as well as sequence groups E, G and H, representing most genomes for the *X. euvesicatoria* and *X. vasicola* groups (Fig. 2). Finally, sequence group K is most numerous among *X. translucens* genomes. The comparison reference is the RaxX protein sequence from the Philippines *Xoo* strain PXO99^A (sequence group A). Examples from lower frequency (mostly unique) sequence groups were analysed by complementation, as described below.

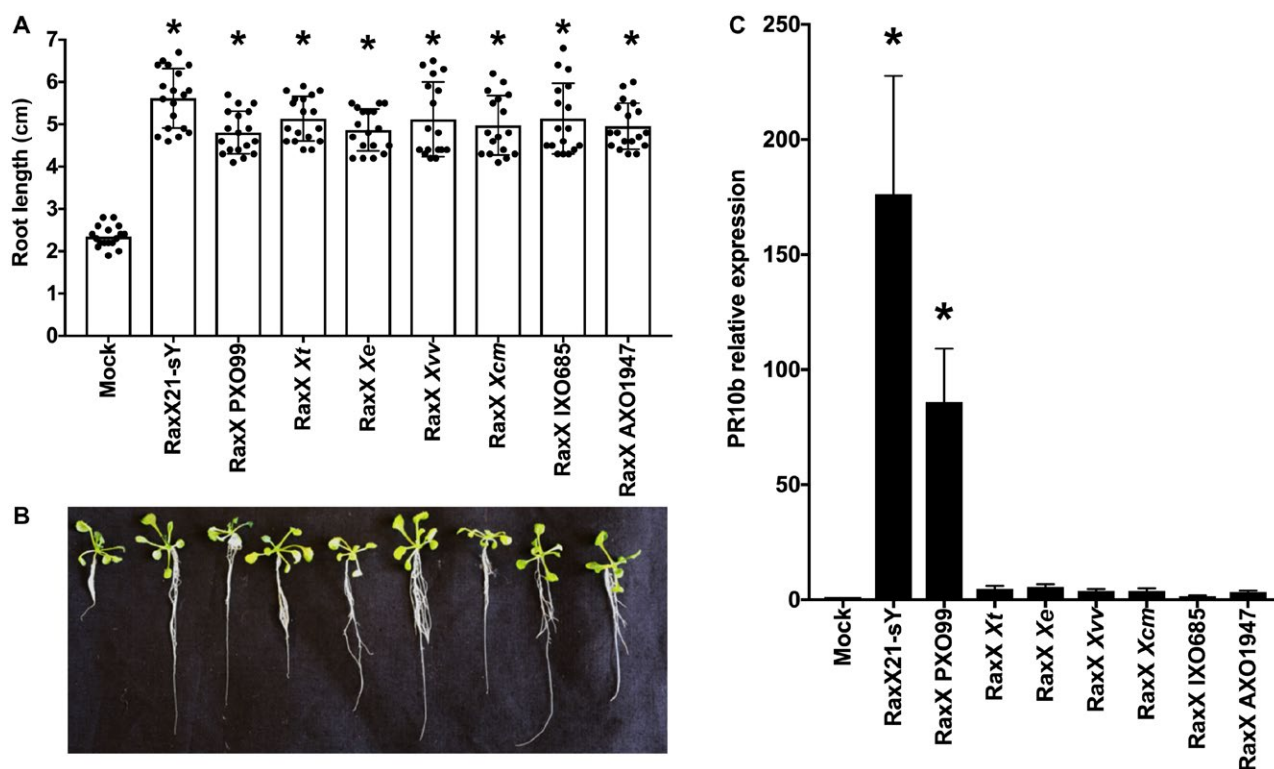


Fig. 5 RaxX variant peptides promote root growth. (A) Stimulation of *Arabidopsis* root growth. Fourteen-day-old *tpst-1* seedlings were grown on half-strength Murashige and Skoog (½ MS) vertical plates with or without 100 nM of the indicated full-length peptides. Bars indicate the average seedling root length measured after 14 days ($n > 10$). Error bars show the standard deviation. Asterisk indicates a statistically significant difference from Mock using Dunnett's test ($P < 0.05$). Peptide RaxX sY21 is a 21-residue sulfated peptide with potent RaxX activity (Pruitt *et al.*, 2015). Strain abbreviations: *Xvv*, *Xanthomonas vasicola* pv. *vasculorum*; *Xt*, *X. translucens*; *Xe*, *X. euvesicatoria*; *Xcm*, *X. campestris* pv. *musacearum*; PXO99^A, IXO685, AXO1947, strains of *X. oryzae* pv. *oryzae*. (B) *Arabidopsis* seedlings from a representative experiment. (C) Activation of rice *PR10b* gene expression. Purified peptide (500 nM) was used to treat detached leaves as described in Experimental procedures. Expression levels of the *PR10b* gene (normalized to actin gene expression) were determined after 12 h. Data are the mean values from four biological replicates. Error bars show the standard deviation. Asterisk indicates a statistically significant difference from Mock using Dunnett's test ($P < 0.05$).

RaxX variants promote root growth, but fail to activate the XA21-mediated immune response

We generated and purified tyrosine-sulfated, full-length (unprocessed) RaxX peptides for these seven variants using an expanded genetic code approach (see Experimental procedures) (Fig. 2), together representing all five pathogenic lineages that contain the *raxX-raxSTAB* gene cluster. The positive control is RaxX21-sY, a synthetic 21-residue tyrosine-sulfated peptide with strong activity; non-sulfated peptides have undetectable activity (Pruitt *et al.*, 2015). The seven RaxX variant peptides were used in two separate assays for function. First, we performed root growth experiments with an *Arabidopsis thaliana* *tpst-1* mutant lacking tyrosylprotein sulfotransferase, which is required for all known sulfated tyrosine peptide hormones, including PSY (Matsubayashi, 2014). This eliminates endogenous PSY activity, so that the effects of the added peptides are more easily observed (Matsubayashi, 2014; Pruitt *et al.*, 2017). Root lengths for seedlings grown without added peptide averaged 23.5 mm, whereas root lengths for seedlings grown with 100 nM peptide were at least twice as long (Fig. 5A,B). This observation is consistent with the hypothesis that these peptides mimic PSY1 peptide hormone activity. It should be noted that three of these variants (groups D, E and G) have been examined previously (Pruitt *et al.*, 2017) and are included here to facilitate direct comparisons, as well as to monitor the consistency of the results.

In the second assay, we tested each RaxX peptide for direct activation of XA21-mediated immunity by assaying the induction of the *PR10b* marker gene as a readout for immune activation (Pruitt *et al.*, 2015; Thomas *et al.*, 2016). In contrast with the results from the root growth assay, only the group A RaxX protein (from *Xoo* strain PXO99^A) was able to induce XA21-mediated *PR10b* marker gene expression (Fig. 5C).

In a separate test for the activation of XA21-mediated immunity, we used a Δ raxX deletion mutant of *Xoo* strain PXO99^A as a host for genetic complementation. We tested each of the *raxX* alleles shown in Fig. 2, which includes examples from lower frequency (mostly unique) sequence groups. We introduced each *raxX* allele into the Δ raxX test strain, and monitored disease progression in leaves of whole plants. Only the group A *raxX* allele (from *Xoo* strain PXO99^A) was able to complement the *Xoo* PXO99^A Δ raxX strain to activate XA21-mediated immunity (Fig. 6). The expression of each *raxX* allele was confirmed by quantitative polymerase chain reaction (qPCR) (Fig. S5, see Supporting Information).

Together, these results provide direct evidence that the activation of XA21-mediated immunity is restricted to RaxX proteins from sequence group A, found in most strains of *Xoo*. None of the other *X. oryzae* RaxX variants tested (including RaxX from *X. oryzae* pv. *oryzicola*, for which the mature sequence is identical to that of *Xoo* strain IXO685), was able

to activate XA21-mediated immunity. The observation that all RaxX proteins tested stimulated *Arabidopsis* root growth

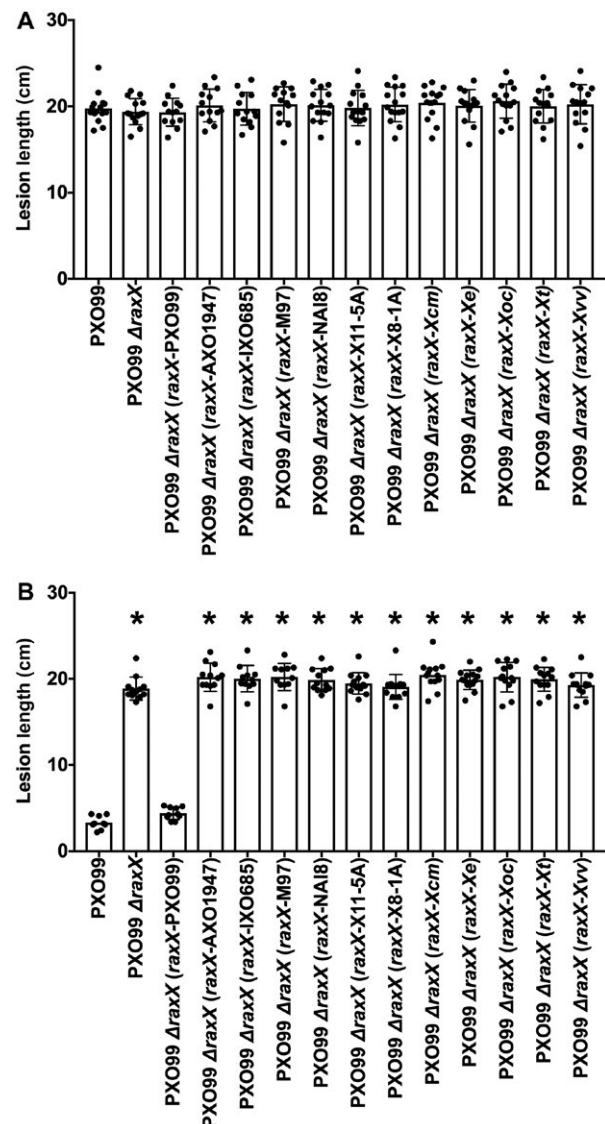


Fig. 6 RaxX variants fail to activate XA21-mediated immunity. Different *raxX* genes were cloned into vector pVSP6 (see Experimental procedures) to test for complementation of the *Xanthomonas oryzae* pv. *oryzae* (*Xoo*) strain PXO99^A Δ raxX strain. Leaf tips of rice varieties TP309 (A) or XA21-expressing TP309 (B) were inoculated by clipping with scissors dipped in bacterial suspensions (approximate cell density of 8×10^8 cells/mL). Lesion lengths were measured 14 days after inoculation. Data are the mean values from measurements of 10–20 leaves. Error bars show the standard error of the mean, and asterisks indicate a statistically significant difference from *Xoo* strain PXO99^A according to Dunnett's multiple comparison procedure ($P < 0.05$). Values in (A) are insignificantly different. Strain abbreviations: Xvv, *X. vasculorum*; Xt, *X. translucens*; Xoc, *X. oryzae* pv. *oryzicola*; Xe, *X. euvesicatoria*; Xcm, *X. campestris* pv. *musacearum*; X8-1A, X11-5A, strains of *X. oryzae*; M97, *X. maliensis* M97; PXO99^A, IXO685, AXO1947, strains of *X. oryzae* pv. *oryzae*.

suggests that the RaxX PSY peptide mimicry function is not restricted to rice.

African *Xoo* strain AXO1947 RaxX and RaxST natural variants both lead to evasion of the XA21 immune receptor

The *raxX* alleles from *Xoo* strains IXO685 and AXO1947 failed to complement the Δ *raxX* mutant of *Xoo* strain PXO99^A for XA21 immune activation (Fig. 6). In addition to its variant *raxX* allele (Fig. 2), we noted that *Xoo* strain AXO1947 (Huguet-Tapia *et al.*, 2016) carries seven missense polymorphisms in the *raxST* gene (Fig. S6, see Supporting Information) not present in other *Xoo* strains, such as IXO685. To determine if the variant *raxST* allele from strain AXO1947 encodes a functional protein, we performed additional complementation tests.

We found that the *raxX* allele from strain PXO99^A conferred the XA21 immune activation phenotype on strain IXO685, but not on strain AXO1947 (Fig. 7B). This result suggests that the *raxX* variant allele is not the only factor that prevents strain AXO1947 from activating the XA21 immune response. Consistent with this hypothesis, the *raxST* allele from strain PXO99^A failed to confer the XA21 immune activation phenotype on strain AXO1947 (Fig. 7D). In contrast, the addition of both the *raxX* and *raxST* alleles from strain PXO99^A was sufficient to confer the XA21 immune activation phenotype on strain AXO1947 (Fig. 7F).

Taken together, these results suggest that *Xoo* strain AXO1947 contains mutant versions of both genes, *raxST* and *raxX*. Analysis by quantitative reverse transcription-polymerase chain reaction (qRT-PCR) confirms that these genes are expressed in the complemented strains (Fig. S7, see Supporting Information).

RaxST variants from *Xoo* strain AXO1947

To determine which of the RaxST missense polymorphisms is responsible for the apparent reduction in enzyme activity, we used site-specific mutagenesis to introduce each individually into the *raxST* gene from strain PXO99^A. Genes encoding two of these [histidine-50 (His-50) to aspartic acid (Asp) (H50D) and arginine-129 (Arg-129) to leucine (Leu) (R129L)] were unable to complement the Δ *raxST* mutant of *Xoo* strain PXO99^A for XA21 immune activation (Fig. 8), indicating that both His-50 and Arg-129 are necessary for RaxST activity.

Little is known about RaxST structure and function. Diverse sulfotransferases share limited sequence similarity, mostly comprising two relatively short sequence motifs involved in PAPS binding (Negishi *et al.*, 2001). These motifs are conserved in the *Xoo* RaxST sequence (da Silva *et al.*, 2004). Research with diverse sulfotransferases has identified three essential residues: a positively charged residue (corresponding to Arg-11 in RaxST) in one PAPS binding motif, an invariant serine (Ser) (corresponding

to Ser-118 in RaxST) in the other and a catalytic base [His or glutamic acid (Glu)] located between the two PAPS binding motifs (Negishi *et al.*, 2001).

We generated a RaxST molecular model with the program iTasser (Yang and Zhang, 2015) using the crystal structure of human tyrosylprotein sulfotransferase-2 (TPST2) as a template (PDB: 3AP1). The sequence alignment is shown in Fig. S8 (see Supporting Information). TPST2 is a functional dimer (Teramoto *et al.*, 2013), which is replicated in the RaxST structural model (Fig. S9, see Supporting Information). The two essential residues identified from *Xoo* strain AXO1947, His-50 and Arg-129, display surface-exposed side chains in close proximity to the corresponding position for the bound substrate peptide co-crystallized with TPST2. These residues are distal to the catalytic site. Therefore, we hypothesize that these RaxST residues are involved in RaxX peptide binding.

DISCUSSION

Previously, we hypothesized that RaxX mimics the actions of PSY hormones, and that the XA21 receptor evolved specifically to recognize RaxX from *Xoo* (Pruitt *et al.*, 2015, 2017). This prediction is supported here by our finding that all the RaxX variants tested stimulate root growth (Fig. 5A,B) (Pruitt *et al.*, 2017), but fail to activate the XA21-mediated immune response (Figs 5C and 6). Thus, RaxX sequence determinants are more stringent for XA21-mediated immunity activation than for root growth stimulation. In this discussion, we consider two questions: (1) what are the potential selective pressures acting on RaxX that affect sequence variation; and (2) how was the *raxX-raxSTAB* gene cluster inherited in *Xanthomonas* spp.?

Opposing selection pressures drive RaxX natural variation

Maintenance of the *raxX-raxSTAB* gene cluster (Fig. 3) suggests that RaxX provides fitness benefits to diverse *Xanthomonas* spp., presumably during their interactions with hosts that collectively encompass a range of monocot and dicot species. This hypothesis is supported by *in vivo* data showing that *Xoo* strains lacking the *raxX* or *raxST* genes are compromised for virulence (Pruitt *et al.*, 2015, 2017). On the other hand, rice-restricted XA21-mediated immunity would select specifically against RaxX maintenance by *Xoo*. Analysis of *raxX-raxSTAB* gene cluster sequence polymorphisms suggests that both types of selection occur.

The *Xa21* gene has been introgressed into commercial rice varieties (Khush *et al.*, 1990; Midha *et al.*, 2017). Widespread planting of *Xa21* rice presumably increases selection for *Xoo* variants that evade XA21-mediated immunity. All RaxX missense variants examined mimicked PSY hormone activity (Fig. 5A,B) (Pruitt *et al.*, 2017), suggesting that this property confers a selective

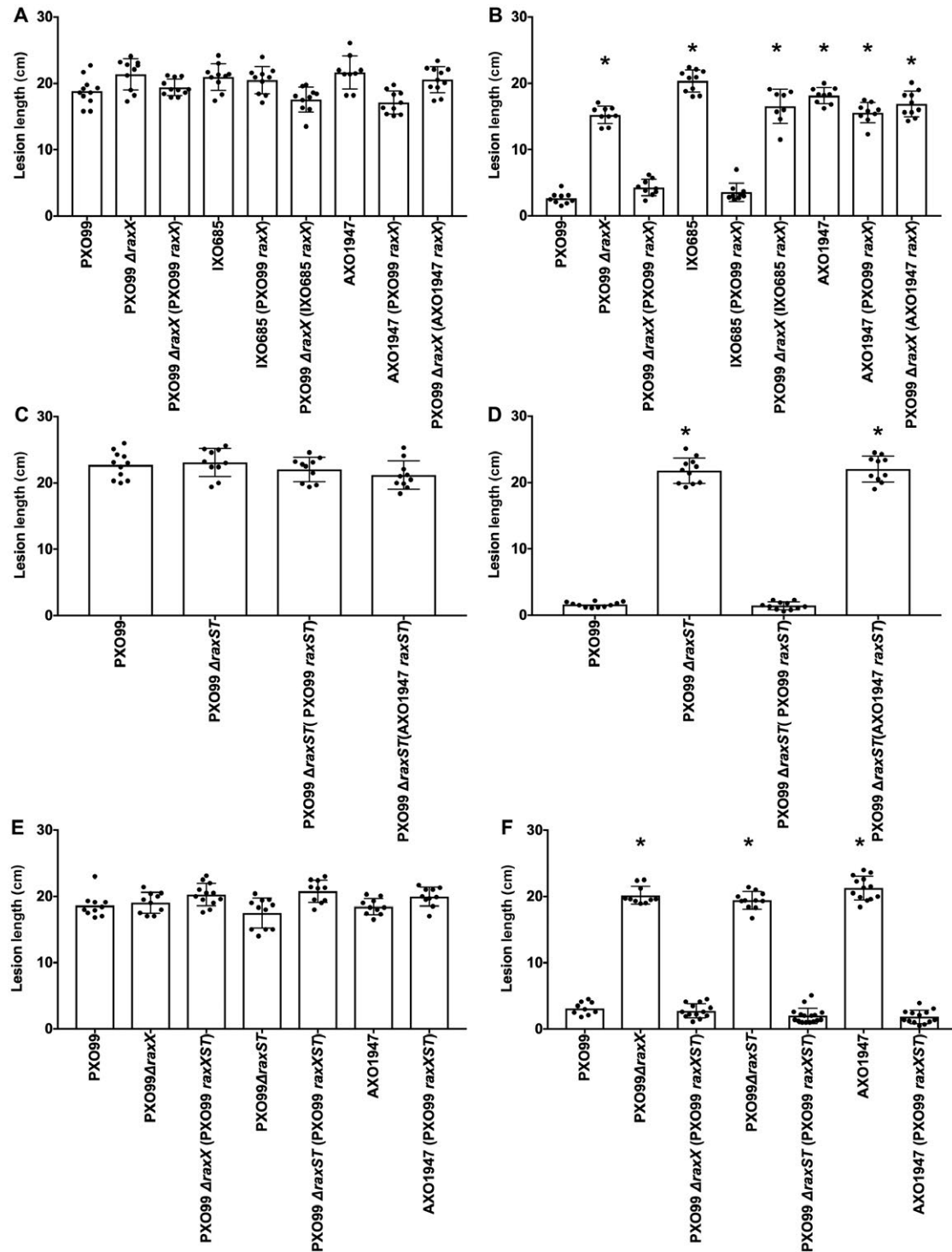


Fig. 7 The *raxX* and *raxST* genes are dysfunctional in *Xanthomonas oryzae* pv. *oryzae* (Xoo) strain AXO1947. Different combinations of the *raxX* and *raxST* genes were cloned into vector pVSP61 (see Experimental procedures) to test for complementation. Leaf tips of rice varieties TP309 (A, C and E) or XA21-expressing TP309 (B, D and F) were inoculated by clipping with scissors dipped in bacterial suspensions (approximate cell density of 8×10^8 cells/mL). Lesion measurements were taken 14 days after inoculation. Data are the mean values from measurements of 10–20 leaves. Error bars show the standard error of the mean, and asterisks indicate a statistically significant difference from Xoo strain PAXO99^A according to Dunnett's multiple comparison procedure ($P < 0.05$). Values in (A), (C) and (E) are not significantly different. (A) and (B) show complementation results for the *raxX* gene, (C) and (D) show results for the *raxST* gene, and (E) and (F) show results for the combination of both the *raxX* and *raxST* genes. Specific combinations of genes and complementation hosts are described in the figure labels.

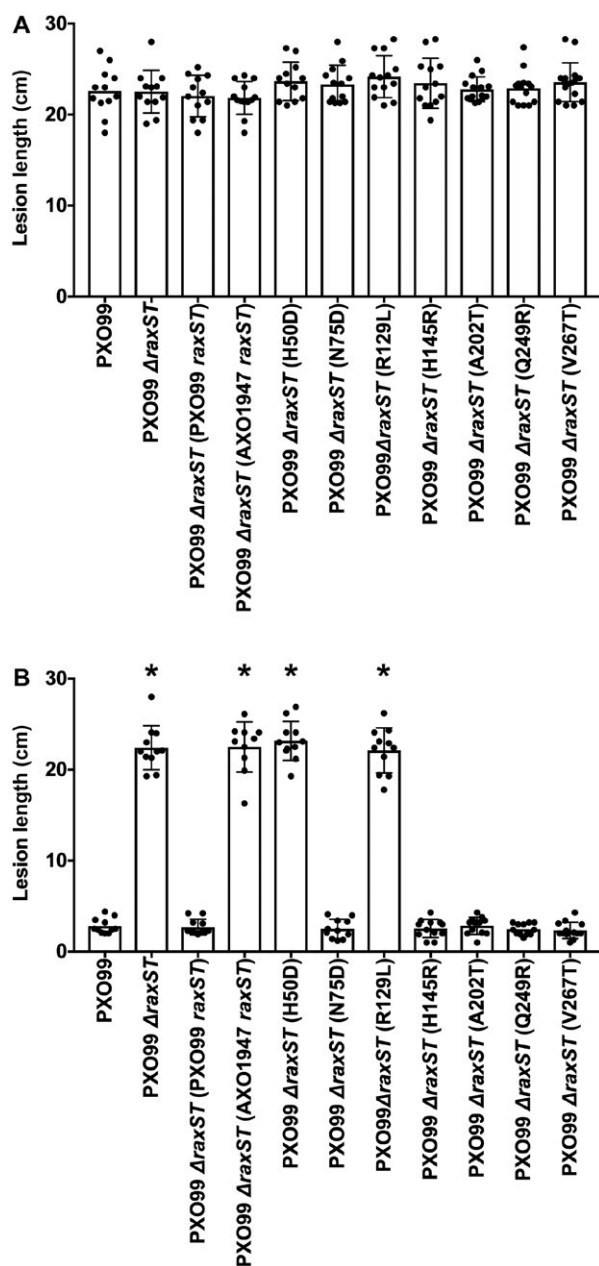


Fig. 8 Two missense substitutions inactivate RaxST in *Xoo* strain AXO1947. Each of the seven *raxST* missense polymorphisms from *Xoo* strain AXO1947 was introduced singly into the wild-type *raxST* gene from *Xoo* strain PXO99^A (see Experimental Procedures). These mutant alleles then were tested for complementation of the *Xoo* strain PXO99^A Δ raxST strain. Leaf tips of rice varieties TP309 (panel A) or XA21-expressing TP309 (panel B) were inoculated by clipping with scissors dipped in bacterial suspensions (approximate cell density of 8×10^8 cells mL⁻¹). Lesion measurements were taken 14 days after inoculation. Data are the mean values from measurements of 10–20 leaves. Error bars show the standard error of the mean, and “*” indicates a statistically significant difference from *Xoo* strain PXO99^A according to Dunnett’s multiple comparison procedure ($P < 0.05$).

advantage. Consistent with this, we did not observe any *raxX* frameshift or nonsense alterations. Instead, RaxX variant sequences contain a restricted set of missense substitutions, consistent with the hypothesis that selection acts to retain PSY-like function (Fig. 2; see Pruitt *et al.*, 2017).

Among all RaxX variants tested, only that from *Xoo* strain PXO99^A (which represents the large majority of *Xoo* *raxX* alleles) activated the XA21-mediated immune response (Figs 5C and 6). This result demonstrates that recognition of RaxX by XA21 is strictly limited to *Xoo*, and confirms and extends a previous conclusion from our laboratory that residues Pro-44 and Pro-48 are both required for *Xoo* RaxX recognition by XA21 (Pruitt *et al.*, 2015).

Thus, it appears that some *Xoo* strains that evade activation of XA21-mediated immunity arise from a restricted set of *raxX* missense substitution alleles encoding variants that retain PSY-like function. This observation suggests that it may be possible to engineer novel XA21 variants that recognize these variant RaxX proteins. If so, it may then be possible to engineer broad-spectrum resistance against *Xoo* (and other *raxX-raxSTAB* gene cluster-positive *Xanthomonas* spp.) by expressing multiple XA21 proteins that collectively recognize multiple RaxX variants.

We also identified *raxST* and/or *raxA* gene loss-of-function alterations in *Xoo* field isolates (Fig. 7; da Silva *et al.*, 2004), which presumably cannot express the PSY mimicry phenotype of RaxX. Such loss-of-function alterations could temper the effectiveness of production strategies that rely on engineered *Xa21* alleles.

raxX-raxSTAB gene cluster origin

The *raxAB* genes are homologous to those encoding proteolytic maturation and ATP-dependent peptide secretion complexes (Lin *et al.*, 2015; da Silva *et al.*, 2004), related to type I secretion systems, but specialized for the secretion of small peptides, such as bacteriocins and peptide pheromones (Holland *et al.*, 2016). Frequently, the gene encoding the secreted substrate is adjacent to genes encoding components of the secretion complex (Dirix *et al.*, 2004). We hypothesize that the intact *raxX-raxSTAB* gene cluster originated in an ancestor to the lineage containing *X. oryzae*, *X. euvesicatoria* and related species, with subsequent gains or losses through lateral gene transfer (Fig. 2). Relatively few events appear to have been necessary to form the *raxX-raxSTAB* gene cluster. The *raxX* gene might have evolved from the gene for the secreted peptide substrate of the RaxAB ancestor. The complete cluster would result from the incorporation of the ancestral *raxST* gene, homologues of which are distributed broadly (Fig. S4).

Role of the *raxX-raxSTAB* gene cluster in *Xanthomonas* biology

The *raxX-raxSTAB* gene cluster does not exhibit features characteristic of self-mobile genomic islands, such as a gene for a site-specific recombinase (Hacker *et al.*, 1997). Instead, evidence suggests that *raxX-STAB* gene cluster lateral gene transfer occurred through general recombination between genes flanking each side of the *raxX-STAB* gene cluster (Figs 1 and S2). In bacteria, gene acquisition through lateral gene transfer contributes to the emergence of new pathovars (for one example, see Ogura *et al.*, 2009). Conceivably, lateral acquisition of the *raxX-raxSTAB* gene cluster might allow a particular strain to infect a previously inaccessible host.

Xanthomonas pathovar phenotypes (Jacques *et al.*, 2016) are not predicted by the presence or absence of the *raxX-raxSTAB* gene cluster. For example, some *raxX-raxSTAB* gene cluster-positive species can infect only monocots (e.g. *X. oryzae*, *X. translucens*) or only dicots (e.g. *X. euvesicatoria*), just as some *raxX-raxSTAB* gene cluster-negative species can also infect only monocots (e.g. *X. arboricola*, *X. hyacinthi*) or only dicots (e.g. *X. campestris* pv. *campestris*, *X. citri*). Similarly, some *raxX-raxSTAB* gene cluster-positive species are specific for vascular tissue (e.g. *Xoo*, *X. vasicola*) or for non-vascular tissue (e.g. *X. oryzae* pv. *oryzicola*, *X. euvesicatoria*), just as some *raxX-raxSTAB* gene cluster-negative species are also specific for vascular tissue (e.g. *X. hortorum*, *X. albilineans*) or for non-vascular tissue (e.g. *X. citri*, *X. arboricola*). Thus, selective function(s) for the *raxX-raxSTAB* gene cluster in *Xanthomonas* spp. remain to be determined.

EXPERIMENTAL PROCEDURES

Survey of *raxX-STAB* homologues in publicly available databases

We used the 5-kb-long *Xoo* PX099^A *raxX-raxSTAB* genomic region, including 600 bp upstream of *raxST* and 70 bp downstream of *raxB*, as query to search the following NCBI databases with BLASTn and MEGABLAST using an e-value cut-off of 1e-3: nr/nt, htgs, refseq_genomic_representative_genomes, refseq_genomic and gss. To identify RaxX homologues, we used the protein sequence of RaxX from *Xoo* PX099^A as query to search the same databases using tBLASTn with a PAM30 scoring matrix to account for the short sequence length of RaxX. In case of *raxST* from *Xoo* PX099^A, we used the genomic coding sequence to search the same databases using the same cut-offs. In addition, we used the RaxST protein sequence to search the following database using BLASTp with an e-value cut-off of 1e-3 and a BLOSUM62 scoring matrix: nr, refseq_protein, env_nr. The databases were last accessed on 6 January 2016 for the initial manuscript submission and 25 June

2018 during preparation of the resubmission. Searches were restricted to bacteria (taxid: 2) in the case of refseq_genomic_representative_genomes. The observations of the specificity of *raxX* and the intact *raxX-raxSTAB* gene cluster to the genus *Xanthomonas* were consistent across all queries.

Whole-genome-based phylogenetic tree for *Xanthomonas* spp.

All available *Xanthomonas* genomes were downloaded from the NCBI ftp server on 29 January 2016 (413 genome accessions). The genome fasta files were used to build a local BLAST database using BLASTV2.27+ (Camacho *et al.*, 2009). For all genes in and surrounding the *raxSTAB* cluster, BLASTn (e-value cut-off of 1e-3) was used to identify homologues in the local BLAST database. As a result of the small size of RaxX, tBLASTn was required to identify homologues (e-value cut-off of 1e-3). Fasta files for each BLAST hit were generated using a custom python script (available on request). Alignments of all genes were performed with Muscle v3.5 (Edgar, 2004) implemented in the desktop tool Geneious v9.1.8 (Kearse *et al.*, 2012). Alignment ends were trimmed so that each sequence was equal in length and in the first coding frame. Maximum likelihood (ML) trees were built with RaxML v8.2.4 (Stamatakis, 2014) with the following settings: (-m GTRGAMMA F -f a -x 3298589 -N 10000 -p 23). Trees shown in all figures are the highest scoring ML trees, and the numbers shown on the branches are the resampled bootstrap values from 1000 replicates. Trees were drawn in FigTree v1.4.0 (<http://tree.bio.ed.ac.uk/software/figtree/>).

Whole-genome phylogenies were generated using the entire genome assembly with the program Andi v0.10 (Haubold *et al.*, 2015; Klotzl and Haubold, 2016). These distance matrices were plotted as neighbour-joining trees using Phylip v3.695 (Felsenstein, 1981). The numbers on the branches represent the proportion (0–100) that the branch appeared in the 'bootstrapped' distance matrices using Andi.

Sequence analyses

Nucleotide and deduced amino acid sequences were edited and analysed with the programs EditSeqTM (version 14.1.0), MegAlignTM (version 14.1.0) and SeqBuilderTM (version 14.1.0), DNASTAR, Madison, WI, USA. The Integrated Microbial Genomes interface (Chen *et al.*, 2017) was used to compare genome segments from different species.

Bacterial growth

Xanthomonas strains were cultured at 28 °C. Solid medium was peptone sucrose agar (PSA; pH 7.0), which contains (per litre) peptone (10 g), sucrose (10 g), sodium glutamate (1 g) and agar (15 g). Liquid cultures were aerated at 230 rpm in YEB medium (pH 7.3), which contains (per litre) yeast extract (5 g), tryptone

(10 g), NaCl (5 g), sucrose (5 g) and MgSO₄ (0.5 g). Antibiotics were kanamycin, carbenicillin, spectinomycin (all at 50 µg/mL) and cephalixin (20 µg/mL).

Rice growth and inoculation

Oryza sativa ssp. *japonica* rice varieties were TP309 and XA21-TP309, which is a 106-17-derived transgenic line of TP309 carrying the *Xa21* gene expressed from its native promoter (Song *et al.*, 1995). TP309 rice does not contain the *Xa21* gene. Seeds were germinated in distilled water at 28 °C for 1 week and then transplanted into sandy soil (80% sand, 20% peat; Redi-Gro, Sacramento, CA) in 5.5-in square pots with two seedlings per pot. Plants were grown in tubs in a glasshouse, and were top watered daily with fertilizer water (N, 58 ppm; P, 15 ppm; K, 55 ppm; Ca, 20 ppm; Mg, 13 ppm; S, 49 ppm; Fe, 1 ppm; Cu, 0.06 ppm; Mn, 0.4 ppm; Mo, 0.02 ppm; Zn, 0.1 ppm; B, 0.4 ppm) for 4 weeks, followed by water for 2 weeks. Six weeks after planting, rice pots were transferred to a growth chamber with the following day/night settings: 28 °C/24 °C, 80%/85% humidity and 14-h/10-h lighting. Plants were inoculated 2–3 days after transfer using the scissors clipping method (Song *et al.*, 1995). Bacteria for inoculation were taken from PSA plates and resuspended in water at a density of approximately 8×10^8 colony-forming units (CFU)/mL. Water-soaked lesions were measured 14 days after inoculation.

Complementation tests

The *Xoo* strain PXO99^A marker-free deletions Δ *raxX* and Δ *raxST* have been described previously (Pruitt *et al.*, 2015). Site-specific mutational alterations were introduced by PCR using the In-Fusion HD cloning system (Takara, Mountain View, CA). The *raxX* and *raxST* genes from different *Xanthomonas* spp. were cloned into plasmid vector pVS61 and electrotransformed into the appropriate recipient strains, as described previously (Pruitt *et al.*, 2015). qRT-PCR was performed as described previously (Pruitt *et al.*, 2015). Gene expression was normalized to the chromosomal gene PXO_01660 (annotated as a homologue of the *ampC* gene encoding β -lactamase). DNA primers for qRT-PCR were: *ampC*-F, GACTCGTAATGCCTACGACC; *ampC*-R, AATTGCTCGTAGAAGCTGCC; *qraxST*-F, CTTCACCGTGCAGATCGAC; *qraxST*-R, TATCGACGATCCAACCAAC; *qraxX*-F, AAAATCGCCCGCAAGGGT; *qraxX*-R, TCAATGGTGCCCGGGGTTG.

RaxX peptide stimulation of *PR10b* gene expression

Full-length sulfated RaxX proteins were purified from an *Escherichia coli* strain with an expanded genetic code that directs the incorporation of sulfotyrosine at the appropriate position (Schwessinger *et al.*, 2016). The resulting MBP-3C–RaxX–His fusion proteins were incubated with 3C protease, followed

by anion exchange chromatography, in order to remove the amino-terminal maltose binding protein tag, as described previously (Schwessinger *et al.*, 2016). The control peptide, sulfated RaxX21-sY, has been described previously (Pruitt *et al.*, 2015).

Rice plants were grown in a hydroponic system in growth chambers at 24 or 28 °C with a 14-h/10-h light–dark cycle at 80% humidity. Seedlings were grown in A-OK Starter Plugs (Grodan, Milton, ON, Canada) and watered with Hoagland's solution twice a week. Peptide influence on *PR10b* marker gene expression was measured as described previously (Pruitt *et al.*, 2015). Briefly, leaves of 4-week-old hydroponically grown rice plants were cut into 2-cm-long strips and incubated for at least 12 h in double-distilled H₂O to reduce residual wound signals. Leaf strips were treated with the indicated peptides and then snap-frozen in liquid nitrogen before processing. qRT-PCR was performed as described previously (Pruitt *et al.*, 2015). Gene expression was normalized to the actin gene expression level and to the respective mock-treated control at 0 or 9 h. DNA primers for qRT-PCR were: *PR10b*-F, TGTGGAAGGTCTGCTTGAC; *PR10b*-R, CCTTAGCACGTGAGTTGCG.

ACKNOWLEDGEMENTS

We thank Simon Williams from Research School of Biology at Australia National University, Canberra and Michael Steinwand from the Ronald laboratory for their assistance with RaxST homology modelling. This study was supported by Public Health Service research grants GM059962 and GM122968 from the National Institute of General Medical Sciences awarded to P.C.R. The authors declare no conflicts of interest.

REFERENCES

- Amano, Y., Tsubouchi, H., Shinohara, H., Ogawa, M. and Matsubayashi, Y. (2007) Tyrosine-sulfated glycopeptide involved in cellular proliferation and expansion in *Arabidopsis*. *Proc. Natl. Acad. Sci. USA*, **104**, 18 333–18 338.
- Aritua, V., Parkinson, N., Thwaites, R., Heeney, J.V., Jones, D.R., Tushemereirwe, W., Crozier, J., Reeder, R., Stead, D.E. and Smith, J. (2008) Characterization of the *Xanthomonas* sp. causing wilt of onset and banana and its proposed reclassification as a strain of *X. vasicola*. *Plant Pathol.* **57**, 170–177.
- Bansal, K., Midha, S., Kumar, S. and Patil, P.B. (2017) Ecological and evolutionary insights into pathovar diversity of *Xanthomonas citri*. *Appl. Environ. Microbiol.* AEM-02993.
- Barak, J.D., Vancheva, T., Lefeuvre, P., Jones, J.B., Timilsina, S., Minsavage, G.V., Vallad, G.E. and Koebnik, R. (2016) Whole-genome sequences of *Xanthomonas euvesicatoria* strains clarify taxonomy and reveal a stepwise erosion of Type 3 effectors. *Front. Plant Sci.* **7**, 1805.
- Bart, R., Cohn, M., Kassen, A., McCallum, E.J., Shybut, M., Petriello, A., Krasileva, K., Dahlbeck, D., Medina, C., Alicai, T., Kumar, L., Moreira, L.M., Neto, J.R., Verdier, V., Santana, M.A., Kositcharoenkul, N., Vanderschuren, H., Gruijssem, W., Bernal, A. and Staskawicz, B.J. (2012) High-throughput genomic sequencing of cassava bacterial blight strains identifies conserved effectors to target for durable resistance. *Proc. Natl. Acad. Sci. USA*, **109**, E1972–E1979.

- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K. and Madden, T.L. (2009) BLAST+: architecture and applications. *BMC Bioinformatics*, **10**, 421.
- Chen, I.A., Markowitz, V.M., Chu, K., Palaniappan, K., Szeto, E., Pillay, M., Ratner, A., Huang, J., Andersen, E., Huntemann, M., Varghese, N., Hadjithomas, M., Tennessen, K., Nielsen, T., Ivanova, N.N. and Kyrpides, N.C. (2017) IMG/M: integrated genome and metagenome comparative data analysis system. *Nucleic Acids Res.* **45**, D507–D516.
- Crossman, L.C., Gould, V.C., Dow, J.M., Vernikos, G.S., Okazaki, A., Sebahia, M., Saunders, D., Arrowsmith, C., Carver, T., Peters, N., Adlem, E., Kerhornou, A., Lord, A., Murphy, L., Seeger, K., Squares, R., Rutter, S., Quail, M.A., Rajandream, M.-A., Harris, D., Churcher, C., Bentley, S.D., Parkhill, J., Thomson, N.R. and Avison, M.B. (2008) The complete genome, comparative and functional analysis of *Stenotrophomonas maltophilia* reveals an organism heavily shielded by drug resistance determinants. *Genome Biol.* **9**, R74.
- Dirix, G., Monsieurs, P., Dombrecht, B., Daniels, R., Marchal, K., Vanderleyden, J. and Michiels, J. (2004) Peptide signal molecules and bacteriocins in Gram-negative bacteria: a genome-wide in silico screening for peptides containing a double-glycine leader sequence and their cognate transporters. *Peptides*, **25**, 1425–1440.
- Edgar, R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797.
- Farzan, M., Mirzabekov, T., Kolchinsky, P., Wyatt, R., Cayabyab, M., Gerard, N.P., Gerard, C., Sodroski, J. and Choe, H. (1999) Tyrosine sulfation of the amino terminus of CCR13 facilitates HIV-1 entry. *Cell*, **96**, 667–676.
- Felsenstein, J. (1981) Evolutionary trees from DNA sequences: a maximum likelihood approach. *J. Mol. Evol.* **17**, 368–376.
- Ferreira-Tonin, M., Rodrigues-Neto, J., Harakava, R. and Destefano, S.A. (2012) Phylogenetic analysis of *Xanthomonas* based on partial *rpoB* gene sequences and species differentiation by PCR-RFLP. *Int. J. Syst. Evol. Microbiol.* **62**, 1419–1424.
- Gardiner, D.M., Upadhyaya, N.M., Stiller, J., Ellis, J.G., Dodds, P.N., Kazan, K. and Manners, J.M. (2014) Genomic analysis of *Xanthomonas translucens* pathogenic on wheat and barley reveals cross-kingdom gene transfer events and diverse protein delivery systems. *PLoS ONE*, **9**, e84995.
- Hacker, J., Blum-Oehler, G., Mühldorfer, I. and Tschäpe, H. (1997) Pathogenicity islands of virulent bacteria: structure, function and impact on microbial evolution. *Mol. Microbiol.* **23**, 1089–1097.
- Han, S.W., Lee, S.W., Bahar, O., Schwessinger, B., Robinson, M.R., Shaw, J.B., Madsen, J.A., Brodbelt, J.S. and Ronald, P.C. (2012) Tyrosine sulfation in a Gram-negative bacterium. *Nat. Commun.* **3**, 1153.
- Hauben, L., Vauterin, L., Swings, J. and Moore, E.R. (1997) Comparison of 16S ribosomal DNA sequences of all *Xanthomonas* species. *Int. J. Syst. Bacteriol.* **47**, 328–335.
- Haubold, B., Klotzl, F. and Pfaffelhuber, P. (2015) Andi: fast and accurate estimation of evolutionary distances between closely related genomes. *Bioinformatics*, **31**, 1169–1175.
- Holland, I.B., Peherstorfer, S., Kanonenberg, K., Lenders, M., Reimann, S. and Schmitt, L. (2016) Type I protein secretion—deceptively simple yet with a wide range of mechanistic variability across the family. *EcoSal Plus*, **7**, <https://doi.org/10.1128/ecosalplus.ESP-0019-2015>.
- Huguet-Tapia, J.C., Peng, Z., Yang, B., Yin, Z., Liu, S. and White, F.F. (2016) Complete genome sequence of the African strain AX01947 of *Xanthomonas oryzae* pv. *oryzae*. *Genome Announc.* **4**, e01730-15.
- Jacobs, J.M., Pesce, C., Lefeuvre, P. and Koebnik, R. (2015) Comparative genomics of a cannabis pathogen reveals insight into the evolution of pathogenicity in *Xanthomonas*. *Front. Plant Sci.* **6**, 431.
- Jacques, M.A., Arlat, M., Boulanger, A., Boureau, T., Carrere, S., Cesbron, S., Chen, N.W., Cociancich, S., Darrasse, A., Denancé, N. and Fischer-Le Saux, M. (2016) Using ecology, physiology, and genomics to understand host specificity in *Xanthomonas*. *Annu. Rev. Phytopathol.* **54**, 163–187.
- Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P. and Drummond, A. (2012) Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, **28**, 1647–1649.
- Khush, G.S., Bacalangco, E. and Ogawa, T. (1990) A new gene for resistance to bacterial blight from *O. longistaminata*. *Rice Genet. Newsl.* **7**, 121–122.
- Kleist, A.B., Getschman, A.E., Ziarek, J.J., Nevins, A.M., Gauthier, P.A., Chevigne, A., Szpakowska, M. and Volkman, B.F. (2016) New paradigms in chemokine receptor signal transduction: moving beyond the two-site model. *Biochem. Pharmacol.* **114**, 53–68.
- Klotzl, F. and Haubold, B. (2016) Support values for genome phylogenies. *Life (Basel)*, **6**, 11.
- Kuo, C.H. and Ochman, H. (2009) The fate of new bacterial genes. *FEMS Microbiol. Rev.* **33**, 38–43.
- Langlois, P.A., Snelling, J., Hamilton, J.P., Bragard, C., Koebnik, R., Verdier, V., Triplett, L.R., Blom, J., Tisserat, N.A. and Leach, J.E. (2017) Characterization of the *Xanthomonas translucens* complex using draft genomes, comparative genomics, phylogenetic analysis, and diagnostic LAMP assays. *Phytopathology*, **107**, 519–527.
- Lin, D.Y., Huang, S. and Chen, J. (2015) Crystal structures of a polypeptide processing and secretion transporter. *Nature*, **523**, 425–430.
- Luu, D. D., Joe, A., Chen, Y., Parys, K., Bahar, O., Pruitt, R., Chen, L. J., Petzold, C. J., Long, K. and Adamchak, C., Stewart, V. (2018) Sulfated RaxX, which represents an unclassified group of ribosomally synthesized post-translationally modified peptides, binds a host immune receptor. *bioRxiv*, 442517.
- Ma, J., Campbell, A. and Karlin, S. (2002) Correlations between Shine–Dalgarno sequences and gene features such as predicted expression levels and operon structures. *J. Bacteriol.* **184**, 5733–5745.
- Mandal, M., Lee, M., Barrick, J.E., Weinberg, Z., Emilsson, G.M., Ruzzo, W.L. and Breaker, R.R. (2004) A glycine-dependent riboswitch that uses cooperative binding to control gene expression. *Science*, **306**, 275–279.
- Matsubayashi, Y. (2014) Posttranslationally modified small-peptide signals in plants. *Annu. Rev. Plant Biol.* **65**, 385–413.
- Mhedbi-Hajri, N., Hajri, A., Boureau, T., Darrasse, A., Durand, K., Brin, C., Saux, M.F.-L., Manceau, C., Poussier, S., Pruvost, O., Lemaire, C. and Jacques, M.-A. (2013) Evolutionary history of the plant pathogenic bacterium *Xanthomonas axonopodis*. *PLoS ONE*, **8**, e58474.
- Midha, S., Bansal, K., Kumar, S., Girija, A.M., Mishra, D., Brahma, K., Laha, G.S., Sundaram, R.M., Sonti, R.V. and Patil, P.B. (2017) Population genomic insights into variation and evolution of *Xanthomonas oryzae* pv. *oryzae*. *Sci. Rep.* **7**, 40 694.
- Midha, S. and Patil, P.B. (2014) Genomic insights into the evolutionary origin of *Xanthomonas axonopodis* pv. *citri* and its ecological relatives. *Appl. Environ. Microbiol.* **80**, 6266–6279.
- Mitchell, J.E., Zheng, D., Busby, S.J. and Minchin, S.D. (2003) Identification and analysis of 'extended-10' promoters in *Escherichia coli*. *Nucleic Acids Res.* **31**, 4689–4695.
- Moore, E.R., Kruger, A.S., Hauben, L., Seal, S.E., Daniels, M.J., De Baere, R., Wachter, R., Timmis, K.N. and Swings, J. (1997) 16S rRNA gene sequence analyses and inter- and intragenetic relationships of *Xanthomonas* species and *Stenotrophomonas maltophilia*. *FEMS Microbiol. Lett.* **151**, 145–153.
- Naushad, S., Adeolu, M., Wong, S., Sohail, M., Schellhorn, H.E. and Gupta, R.S. (2015) A phylogenomic and molecular marker based taxonomic framework for the order *Xanthomonadales*: proposal to transfer

- the families *Algiphilaceae* and *Solimonadaceae* to the order *Nevskiales* ord. nov. and to create a new family within the order *Xanthomonadales*, the family *Rhodanobacteraceae* fam. nov., containing the genus *Rhodanobacter* and its closest relatives. *Antonie Van Leeuwenhoek*, **107**, 467–485.
- Negishi, M., Pedersen, L.G., Petrotchenko, E., Shevtsov, S., Gorokhov, A., Kakuta, Y. and Pedersen, L.C. (2001) Structure and function of sulfo-transferases. *Arch. Biochem. Biophys.* **390**, 149–157.
- Nelson, K., Wang, F.S., Boyd, E.F. and Selander, R.K. (1997) Size and sequence polymorphism in the isocitrate dehydrogenase kinase/phosphatase gene (*aceK*) and flanking regions in *Salmonella enterica* and *Escherichia coli*. *Genetics*, **147**, 1509–1520.
- Ogura, Y., Ooka, T., Iguchi, A., Toh, H., Asadulghani, M., Oshima, K., Kodama, T., Abe, H., Nakayama, K., Kurokawa, K., Tobe, T., Hattori, M. and Hayashi, T. (2009) Comparative genomics reveal the mechanism of the parallel evolution of O157 and non-O157 enterohemorrhagic *Escherichia coli*. *Proc. Natl. Acad. Sci. USA*, **106**, 17 939–17 944.
- O'Leary, N.A., Wright, M.W., Brister, J.R., Ciufo, S., Haddad, D., McVeigh, R., Rajput, B., Robbertse, B., Smith-White, B., Ako-Adjei, D., Astashyn, A., Badrettdin, A., Bao, Y., Blinkova, O., Brover, V., Chetvernin, V., Choi, J., Cox, E., Ermolaeva, O., Farrell, C.M., Goldfarb, T., Gupta, T., Haft, D., Hatcher, E., Hlavina, W., Joardar, V.S., Kodali, V.K., Li, W., Maglott, D., Masterson, P., McGarvey, K.M., Murphy, M.R., O'Neill, K., Pujar, S., Rangwala, S.H., Rausch, D., Riddick, L.D., Schoch, C., Shkeda, A., Storz, S.S., Sun, H., Thibaud-Nissen, F., Tolstoy, I., Tully, R.E., Vatsan, A.R., Wallin, C., Webb, D., Wu, W., Landrum, M.J., Kimchi, A., Tatusova, T., DiCuccio, M., Kitts, P., Murphy, T.D. and Pruitt, K.D. (2016) Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* **44**, D733–D745.
- Parkinson, N., Aritua, V., Heeney, J., Cowie, C., Bew, J. and Stead, D. (2007) Phylogenetic analysis of *Xanthomonas* species by comparison of partial gyrase B gene sequences. *Int. J. Syst. Evol. Microbiol.* **57**, 2881–2887.
- Parkinson, N., Cowie, C., Heeney, J. and Stead, D. (2009) Phylogenetic structure of *Xanthomonas* determined by comparison of *gyrB* sequences. *Int. J. Syst. Evol. Microbiol.* **59**, 264–274.
- Pereira, U.P., Gouran, H., Nascimento, R., Adaskaveg, J.E., Goulart, L.R. and Dandekar, A.M. (2015) Complete genome sequence of *Xanthomonas arboricola* pv. *juglandis* 417, a copper-resistant strain isolated from *Juglans regia* L. *Genome Announc.* **3**, e01126–e01135.
- Pieretti, I., Cociancich, S., Bolot, S., Carrere, S., Morisset, A., Rott, P. and Royer, M. (2015) Full genome sequence analysis of two isolates reveals a novel *Xanthomonas* species close to the sugarcane pathogen *Xanthomonas albilineans*. *Genes (Basel)*, **6**, 714–733.
- Pruitt, R.N., Joe, A., Zhang, W., Feng, W., Stewart, V., Schwessinger, B., Dinnyeny, J.R. and Ronald, P.C. (2017) A microbially derived tyrosine-sulfated peptide mimics a plant peptide hormone. *New Phytol.* **215**, 725–736.
- Pruitt, R.N., Schwessinger, B., Joe, A., Thomas, N., Liu, F., Albert, M., Robinson, M.R., Chan, L.J.G., Luu, D.D., Chen, H., Bahar, O., Daudi, A., De Vleeschauwer, D., Caddell, D., Zhang, W., Zhao, X., Li, X., Heazlewood, J.L., Ruan, D., Majumder, D., Chern, M., Kalbacher, H., Midha, S., Patil, P.B., Sonti, R.V., Petzold, C.J., Liu, C.C., Brodbelt, J.S., Felix, G. and Ronald, P.C. (2015) The rice immune receptor XA21 recognizes a tyrosine-sulfated protein from a Gram-negative bacterium. *Sci. Adv.* **1**, e1500245.
- Rademaker, J.L.W., Louws, F.J., Schultz, M.H., Rossbach, U., Vauterin, L., Swings, J. and De Bruijn, F.J. (2005) A comprehensive species to strain taxonomic framework for *Xanthomonas*. *Phytopathology*, **95**, 1098–1111.
- Robert, X. and Gouet, P. (2014) Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* **42**, W320–W324.
- Ronald, P.C. and Beutler, B. (2010) Plant and animal sensors of conserved microbial signatures. *Science*, **330**, 1061–1064.
- Salzberg, S.L., Sommer, D.D., Schatz, M.C., Phillippy, A.M., Rabinowicz, P.D., Tsuge, S., Furutani, A., Ochiai, H., Delcher, A.L., Kelley, D., Madupu, R., Puiu, D., Radune, D., Shumway, M., Trapnell, C., Aparna, G., Jha, G., Pandey, A., Patil, P.B., Ishihara, H., Meyer, D.F., Szurek, B., Verdier, V., Koebnik, R., Dow, J.M., Ryan, R.P., Hirata, H., Tsuyumu, S., Won Lee, S., Ronald, P.C., Sonti, R.V., Van Sluys, M.-A., Leach, J.E., White, F.F. and Bogdanove, A.J. (2008) Genome sequence and rapid evolution of the rice pathogen *Xanthomonas oryzae* pv. *oryzae* PXO99A. *BMC Genomics*, **9**, 204.
- Schwessinger, B., Li, X., Ellinghaus, T.L., Chan, L.J., Wei, T., Joe, A., Thomas, N., Pruitt, R., Adams, P.D., Chern, M.S., Petzold, C.J., Liu, C.C. and Ronald, P.C. (2016) A second-generation expression system for tyrosine-sulfated proteins and its application in crop protection. *Integr. Biol. (Camb.)* **8**, 542–545.
- Shen, Y., Sharma, P., da Silva, F.G. and Ronald, P. (2002) The *Xanthomonas oryzae* pv. *oryzae* *raxP* and *raxQ* genes encode an ATP sulphurylase and adenosine-5'-phosphosulphate kinase that are required for AvrXa21 avirulence activity. *Mol. Microbiol.* **44**, 37–48.
- da Silva, A.C., Ferro, J.A., Reinach, F.C., Farah, C.S., Furlan, L.R., Quaggio, R.B., Monteiro-Vitorello, C.B., Van Sluys, M.A., Almeida, N.F., Alves, L.M.C., do Amaral, A.M., Bertolini, M.C., Camargo, L.E.A., Camarotte, G., Cannavan, F., Cardozo, J., Chambergo, F., Ciapina, L.P., Cicarelli, R.M.B., Coutinho, L.L., Cursino-Santos, J.R., El-Dorry, H., Faria, J.B., Ferreira, A.J.S., Ferreira, R.C.C., Ferro, M.I.T., Formighieri, E.F., Franco, M.C., Greggio, C.C., Gruber, A., Katsuyama, A.M., Kishi, L.T., Leite, R.P., Lemos, E.G.M., Lemos, M.V.F., Locali, E.C., Machado, M.A., Madeira, A.M.B.N., Martinez-Rossi, N.M., Martins, E.C., Meidanis, J., Menck, C.F.M., Miyaki, C.Y., Moon, D.H., Moreira, L.M., Novo, M.T.M., Okura, V.K., Oliveira, M.C., Oliveira, V.R., Pereira, H.A., Rossi, A., Sena, J.A.D., Silva, C., de Souza, R.F., Spinola, L.A.F., Takita, M.A., Tamura, R.E., Teixeira, E.C., Tezza, R.I.D., Trindade dos Santos, M., Truffi, D., Tsai, S.M., White, F.F., Setubal, J.C. and Kitajima, J.P. (2002) Comparison of the genomes of two *Xanthomonas* pathogens with differing host specificities. *Nature*, **417**, 459–463.
- da Silva, F.G., Shen, Y., Dardick, C., Burdman, S., Yadav, R.C., de Leon, A.L. and Ronald, P.C. (2004) Bacterial genes involved in type I secretion and sulfation are required to elicit the rice Xa21-mediated innate immune response. *Mol. Plant–Microbe Interact.* **17**, 593–601.
- Song, W.Y., Wang, G.L., Chen, L.L., Kim, H.S., Pi, L.Y., Holsten, T., Gardner, J., Wang, B., Zhai, W.-X., Zhu, L.-H., Fauquet, C. and Ronald, P. (1995) A receptor kinase-like protein encoded by the rice disease resistance gene, *Xa21*. *Science*, **270**, 1804–1806.
- Stamatakis, A. (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, **30**, 1312–1313.
- Stone, M.J., Chuang, S., Hou, X., Shoham, M. and Zhu, J.Z. (2009) Tyrosine sulfation: an increasingly recognised post-translational modification of secreted proteins. *Nat. Biotechnol.* **25**, 299–317.
- Studholme, D.J., Wasukira, A., Paszkiewicz, K., Aritua, V., Thwaites, R., Smith, J. and Grant, M. (2011) Draft genome sequences of *Xanthomonas sacchari* and two banana-associated *Xanthomonas* reveal insights into the *Xanthomonas* group 1 clade. *Genes (Basel)*, **2**, 1050–1065.
- Tang, D., Wang, G. and Zhou, J.M. (2017) Receptor kinases in plant–pathogen interactions: more than pattern recognition. *Plant Cell*, **29**, 618–637.
- Teramoto, T., Fujikawa, Y., Kawaguchi, Y., Kurogi, K., Soejima, M., Adachi, R., Nakanishi, Y., Mishihiro-Sato, E., Liu, M.-C., Sakakibara, Y., Suiko, M., Kimura, M. and Kakuta, Y. (2013) Crystal structure of human tyrosylprotein sulfotransferase-2 reveals the mechanism of protein tyrosine sulfation reaction. *Nat. Commun.* **4**, 1572.

- Thieme, F., Koebnik, R., Bekel, T., Berger, C., Boch, J., Buttner, D., Caldana, C., Gaigalat, L., Goesmann, A., Kay, S., Kirchner, O., Lanz, C., Linke, B., McHardy, A.C., Meyer, F., Mittenhuber, G., Nies, D.H., Niesbach-Klosgen, U., Patschkowski, T., Ruckert, C., Rupp, O., Schneiker, S., Schuster, S.C., Vorholter, F.-J., Weber, E., Puhler, A., Bonas, U., Bartels, D. and Kaiser, O. (2005) Insights into genome plasticity and pathogenicity of the plant pathogenic bacterium *Xanthomonas campestris* pv. *vesicatoria* revealed by the complete genome sequence. *J. Bacteriol.* **187**, 7254–7266.
- Thomas, N.C., Schwessinger, B., Liu, F., Chen, H., Wei, T., Nguyen, Y.P., Shaker, I.W. and Ronald, P.C. (2016) XA21-specific induction of stress-related genes following *Xanthomonas* infection of detached rice leaves. *PeerJ*, **4**, e2446.
- Triplett, L.R., Verdier, V., Campillo, T., Van Malderghem, C., Cleenwerck, I., Maes, M., Deblais, L., Corral, R., Koita, O., Cottyn, B. and Leach, J.E. (2015) Characterization of a novel clade of *Xanthomonas* isolated from rice leaves in Mali and proposal of *Xanthomonas maliensis* sp. nov. *Antonie Van Leeuwenhoek*, **107**, 869–881.
- Vancheva, T., Bogatzevska, N., Moncheva, P., Lefeuvre, P. and Koebnik, R. (2015) Draft genome sequences of two *Xanthomonas vesicatoria* strains from the Balkan peninsula. *Genome Announc.* **3**, e01558-14.
- Vandroemme, J., Cottyn, B., Baeyen, S., De Vos, P. and Maes, M. (2013) Draft genome sequence of *Xanthomonas fragariae* reveals reductive evolution and distinct virulence-related gene content. *BMC Genomics*, **14**, 829.
- Vauterin, L., Hoste, B., Kersters, K. and Swings, J. (1995) Reclassification of *Xanthomonas*. *Int. J. Syst. Bacteriol.* **45**, 472–489.
- Vauterin, L., Rademaker, J. and Swings, J. (2000) Synopsis on the taxonomy of the genus *Xanthomonas*. *Phytopathology*, **90**, 677–682.
- Wang, G.L., Song, W.Y., Ruan, D.L., Sideris, S. and Ronald, P.C. (1996) The cloned gene, Xa21, confers resistance to multiple *Xanthomonas oryzae* pv. *oryzae* isolates in transgenic plants. *Mol. Plant-Microbe Interact.* **9**, 850–855.
- Wasukira, A., Tayebwa, J., Thwaites, R., Paszkiewicz, K., Aritua, V., Kubiriba, J., Smith, J., Grant, M. and Studholme, D.J. (2012) Genome-wide sequencing reveals two major sub-lineages in the genetically monomorphic pathogen *Xanthomonas campestris* pathovar *musacearum*. *Genes (Basel)*, **3**, 361–377.
- Yang, J. and Zhang, Y. (2015) I-TASSER server: new development for protein structure and function predictions. *Nucleic Acids Res.* **43**, W174–W181.
- Young, J.M. (2008) An overview of bacterial nomenclature with special reference to plant pathogens. *Syst. Appl. Microbiol.* **31**, 405–424.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article at the publisher's web site:

Fig. S1 Whole-genome-based *Xanthomonas* phylogenetic tree. This tree was constructed by the analysis of whole-genome sequences, as described in Experimental procedures. Blue indicates genomes that contain the *raxX-raxSTAB* gene cluster; red indicates genomes that do not. Group numbers are arbitrary.

Fig. S2 Sequences flanking the *raxX-raxSTAB* gene cluster. Sequences are from the reference strains described in Table 1. Sequences conserved within a group, but different from other groups, are coloured green ('early-branching' species), brown (*raxX-raxSTAB* cluster-negative strains) or yellow (*raxX-raxSTAB* cluster-positive strains). For presentation, the sequence is divided

into left and right boundaries. The green and brown sequences are contiguous, whereas the yellow sequences are interrupted by the c. 5-kb *raxX-raxSTAB* gene cluster, depicted as a yellow rectangle. For presentation, approximately 60–80 nucleotides with relatively low similarity were removed from the sequence shown in the right boundary panel. These conceptual deletions are denoted by the number of nucleotides removed in each case. Black sequences are conserved in all lineages, and include both coding regions as well as matches to transcription and translation initiation consensus sequences, which are described in the text. An 'mfsX' + 1 frameshift in *Xoo* sequences is indicated by the vertical red line. Abbreviations are in red for *raxX-raxSTAB* cluster-negative strains and in blue for *raxX-raxSTAB* cluster-positive strains: *Sm*, *Stenotrophomonas maltophilia*; *Xa*, *Xanthomonas albilineans*; *Xac*, *X. citri* ssp. *citri*; *Xaj*, *X. arboricola* pv. *juglandis*; *Xam*, *X. axonopodis* pv. *manihotis*; *Xc*, *X. cannabis*; *Xcc*, *X. campestris* pv. *campestris*; *Xcm*, *X. campestris* pv. *musacearum*; *Xe*, *X. euvesicatoria*; *Xf*, *X. fragariae*; *Xh*, *X. hyacinthi*; *Xm*, *X. maliensis*; *Xoo*, *X. oryzae* pv. *oryzae*; *Xs*, *X. sacchari*; *Xt*, *X. translucens*; *Xv*, *X. vesicatoria*.

Fig. S3 GcvP length polymorphisms in different *Xanthomonas* lineages. The relevant portion of the GcvP amino acid sequence is shown for each of the reference strains. Species in red lack the *raxX-raxSTAB* gene cluster, whereas those in blue carry the cluster. Numbers denote different allelic types for reference to Fig. 3. The positions of residues Gly-733 and Val-738 (numbering for allelic type 1) are indicated. Abbreviations: *Sm*, *Stenotrophomonas maltophilia*; *Xa*, *Xanthomonas albilineans*; *Xac*, *X. citri* ssp. *citri*; *Xaj*, *X. arboricola* pv. *juglandis*; *Xam*, *X. axonopodis* pv. *manihotis*; *Xc*, *X. cannabis*; *Xcc*, *X. campestris* pv. *campestris*; *Xcm*, *X. campestris* pv. *musacearum*; *Xe*, *X. euvesicatoria*; *Xf*, *X. fragariae*; *Xh*, *X. hyacinthi*; *Xm*, *X. maliensis*; *Xoo*, *X. oryzae* pv. *oryzae*; *Xs*, *X. sacchari*; *Xt*, *X. translucens*; *Xv*, *X. vesicatoria*.

Fig. S4 Phylogenetic tree for *raxST* homologues. Distribution of *raxST* homologues across bacterial genera, including the major groups of proteobacteria as well as cyanobacteria. The tree shown was constructed by neighbour-joining with 1000 bootstrap replicates; branches with <50% bootstrap support are not drawn. The *raxST* sequence from *Xanthomonas oryzae* pv. *oryzae* (*Xoo*) strain PX099^A was used as query for tBLASTn.

Fig. S5 *raxX* expression in *Xanthomonas oryzae* pv. *oryzae* (*Xoo*) PX099^A complemented strains. Data show *raxX* gene expression in the complemented strains with different *raxX* alleles with its promoter region on plasmids. The expression is shown as the logarithm of raw data using quantitative reverse transcription-polymerase chain reaction (qRT-PCR). Data are the mean values from two biological replicates. Error bars show the standard deviation.

Fig. S6 *RaxST* sequence polymorphisms in *Xanthomonas oryzae* pv. *oryzae* (*Xoo*) strain AX01947. The *RaxST* sequence from *Xoo* strain PX099^A is shown. The seven missense substitutions in the sequence from *Xoo* strain AX01947 (Huguet-Tapia *et al.*,

2016) are indicated. The boundaries of the 3'-phosphoadenosine 5'-phosphosulfate (PAPS) binding motifs (5'-PSB and 3'-PB; Negishi *et al.*, 2001), enclosed in boxes, include the catalytic residues Arg-11 and Ser-118.

Fig. S7 *raxX* and *raxST* expression in *Xanthomonas oryzae* pv. *oryzae* (Xoo) PXO99^A complemented strains. Data show *raxX* and *raxST* gene expression in the complemented strains (with *raxX* and *raxST* on plasmids) relative to expression in Xoo strain PXO899^A (with *raxX* and *raxST* on the chromosome). Expression was determined by quantitative reverse transcription-polymerase chain reaction (qRT-PCR) (see Experimental procedures), and is shown as the logarithm of the fold change. Gene expression was normalized to the chromosomal gene PXO_01660 (annotated as an *ampC* gene homologue encoding-lactamase). Data are the mean values from two biological replicates. Error bars show the standard deviation.

Fig. S8 RaxST structural alignment. Sequence alignment of the human tyrosylprotein sulfotransferase-2 (TPST2) and *Xanthomonas oryzae* pv. *oryzae* (Xoo) RaxST sequences

formatted with ESPript 3.0 (Robert & Gouet, 2014). Secondary structure elements derived from the respective structural models are shown. Stars show TPST2 residues involved in 3'-phosphoadenosine 5'-phosphosulfate (PAPS) binding, and arrows show RaxST missense substitutions.

Fig. S9 Model for RaxST structure. Predicted RaxST structure shown in cartoon and surface representation, based on the dimeric structure of tyrosylprotein sulfotransferase-2 (TPST2). The two RaxST monomers are coloured in dark and light green. The 3'-phosphoadenosine 5'-phosphate (PAP) and C4 substrate peptide that were co-crystallized with TPST2 are superimposed on the RaxST model. PAP is represented as labelled and the substrate peptide is shown in yellow cartoon with the acceptor tyrosine represented as labelled. Residues His-50 and Arg-129 are coloured in magenta and highlighted.

File S1 *Xanthomonas* strains analysed for whole-genome phylogeny. Excel file (.XLS format).