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

**Furong Liu,[†,1] Megan McDonald,[†,2] Benjamin Schwessinger,[†,1,2]**

**Anna Joe,[1] Rory Pruitt,[1,a] Teresa Erickson,[1] Xiuxiang Zhao,[1,b]**

**Valley Stewart,[*,3] and Pamela C. Ronald[*,1]**

[†] co-first author, in alphabetical order

* corresponding authors

[1] Department of Plant Pathology and the Genome Center, University of California, Davis

[2] Research School of Biology, Australian National University, Canberra

[3] Department of Microbiology & Molecular Genetics, University of California, Davis

Present Address:

[a] Department of Plant Biochemistry, University of Tübingen, Germany

[b] Department of Plant Pathology, Shenyang Agricultural University, China

**Running head:** *raxX-raxSTAB* gene cluster in *Xanthomonas* spp.

**Keywords:** *Xanthomonas*; tyrosine sulfating; plant innate immunity

**Summary**

The rice XA21-mediated immune response is activated upon recognition of the RaxX peptide produced by the bacterium *Xanthomonas oryzae* pv. *oryzae* (*Xoo*). The 60 residue RaxX precursor is posttranslationally modified to form a sulfated tyrosine peptide that shares sequence and functional similarity with the plant sulfated tyrosine (PSY) peptide hormones. The five kb *raxX-raxSTAB* gene cluster of *Xoo* encodes RaxX, the RaxST tyrosylprotein sulfotransferase, and the RaxA and RaxB components of a predicted type one secretion system. The identified the complete *raxX-raxSTAB* gene cluster is present 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 transfer events during *Xanthomonas* speciation. RaxX variants representing each of the five lineages, and three *Xoo* RaxX variants, fail to activate the XA21-mediated immune response yet retain peptide hormone activity. These RaxX variants contain a restricted set of missense mutations, consistent with the hypothesis that selection acts to maintain peptide hormone-like function. These observations are also consistent with the hypothesis that the XA21 receptor evolved specifically to recognize *Xoo* RaxX.

## INTRODUCTION

15 Host receptors activate innate immunity pathways upon pathogen recognition (Ronald & Beutler,
16 2010). The gene encoding the rice XA21 receptor kinase (Song *et al.*, 1995) confers resistance
17 against most strains of the gamma-proteobacterium *Xanthomonas oryzae* pv. *oryzae* (*Xoo*)
18 (Wang *et al.*, 1996). This well-studied XA21-*Xoo* interaction provides a basis from which to
19 understand molecular and evolutionary mechanisms of host-microbe interactions.

20 Four *Xoo* genes that are **r**equired for **a**ctivation of **X**A21-mediated immunity, are located in the
21 *raxX-raxSTAB* gene cluster (**Fig. 1**). The 60-residue RaxX predicted precursor protein
22 undergoes sulfation by the RaxST tyrosylprotein sulfotransferase at residue Tyr-41 (Pruitt *et al.*,
23 2015). We hypothesize that the RaxB proteolytic maturation and ATP-dependent peptide
24 secretion complex (da Silva *et al.*, 2004) further processes the sulfated RaxX precursor by
25 removing its double-glycine leader peptide prior to secretion (Holland *et al.*, 2016). Located
26 outside the *raxX-raxSTAB* gene cluster, the *raxC* gene, an ortholog of the *tolC* gene, encodes the
27 predicted outer membrane channel for this secretion complex (da Silva et al., 2004). Finally, the
28 *raxPQ* genes encode enzymes to assimilate sulfate into 3'-**p**hospho**a**denosine 5'-**p**hospho**s**ulfate
29 (PAPS) (Shen *et al.*, 2002), the sulfodonor for the RaxST sulfotransferase (Han *et al.*, 2012).

30 In both plants and animals, the post-translational modification catalyzed by tyrosylprotein
31 sulfotransferase is restricted to a subset of cell surface and secreted proteins that influence a
32 variety of eukaryotic physiological processes (Matsubayashi, 2014, Stone *et al.*, 2009). For
33 example, tyrosine sulfation of the chemokine receptors CCR5 and CXCR4 is essential for their

34 functions including as coreceptors for the human immunodeficiency virus gp120 envelope
35 glycoprotein (Farzan *et al.*, 1999, Kleist *et al.*, 2016). In plants, sulfated tyrosine peptides
36 influence cellular proliferation and expansion in root growth, and/or plant immune signaling
37 (Matsubayashi, 2014, Tang *et al.*, 2017). In contrast to these and other examples of protein
38 tyrosine sulfation in animals and plants, RaxX sulfation by the RaxST enzyme is the only
39 example of tyrosine sulfation documented in bacteria (Pruitt et al., 2015, Han et al., 2012).

40 Mature RaxX is predicted to comprise the carboxyl-terminal residues 40-60, numbered according
41 to the precursor protein (Pruitt et al., 2015, Pruitt *et al.*, 2017). RaxX residues 40-52 share
42 sequence similarity with mature **p**lant peptide containing **s**ulfated t**y**rosine (PSY) hormones
43 (Pruitt et al., 2015, Amano *et al.*, 2007, Pruitt et al., 2017). RaxX, like PSY1, can enhance root
44 growth in diverse plant species (Pruitt et al., 2017). The XA21-mediated response in rice
45 requires residues 40-55 (RaxX16 peptide), whereas plant growth stimulation requires only
46 residues 40-52 (RaxX13 peptide) (Pruitt et al., 2015). **Fig. 2** shows sequences for the RaxX
47 variants examined in this study, together with two representative PSY sequences for comparison.

48 RaxX sequences generally are well conserved within different *Xanthomonas* species (Pruitt et al.,
49 2017). In *Xoo* however, RaxX from a strain IXO685, which evades XA21-mediated immunity
50 differs from active RaxX at the critical positions Pro-44 and Pro-48 (**Fig. 2**) (Pruitt et al., 2015).
51 Nevertheless, this RaxX protein stimulates root growth, as do two other RaxX Pro-48 variants
52 from other *Xanthomonas* spp. (Fig. 6 in (Pruitt et al., 2017)).

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

59 Two general predictions derive from this hypothesis. The first prediction, that PSY hormone
60 mimicry is broadly selective, is supported here by the presence of the *raxX-raxSTAB* gene cluster
61 *Xanthomonas* spp., and by the ability of all RaxX variants tested to stimulate root growth in an
62 assay for PSY function. The second prediction, that recognition by XA21 is restricted to *X.*
63 *oryzae* lineages, is validated here by the observation that XA21-mediated immunity is not
64 activated by RaxX variants from other *Xanthomonas* spp. These results illustrate how a
65 pathogen protein has evolved to retain its ability to modulate host physiology without being
66 recognized by the host immune system.

## RESULTS

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

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

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

83 To examine *raxX-raxSTAB* gene cluster organization and inheritance more closely, we selected
84 15 genomes from strains that represent the phylogenetic range of *Xanthomonas* spp. (**Table 1**
85 and **Fig. S1**). Where possible, we chose complete genome sequences that are accompanied by
86 published descriptions. Throughout the analyses described below, species for which relatively
87 large numbers of sequences are available also were monitored broadly for exceptional features.
88 The close relative *Stenotrophomonas maltophilia*, which does not contain the *raxX-raxSTAB*
89 gene cluster, serves as the outgroup (Moore *et al.*, 1997).

90 To facilitate discussion, we represent phylogenetic relationships between these strains as a
91 cladogram that emphasizes relative positions of the *raxX-raxSTAB* gene cluster-positive lineages
92 (**Fig. 3**). Six distinct *Xanthomonas* lineages contain the *raxX-raxSTAB* gene cluster, one being *X.*
93 *oryzae*. A second lineage includes related strains currently denoted as *X. vasicola* or *X.*
94 *campestris* pv. *musacearum* (Aritua *et al.*, 2008); for concise presentation, we refer to these
95 collectively as *X. vasicola*. The third lineage includes *X. euvesicatoria* and related species
96 (Rademaker group 9.2; (Rademaker et al., 2005, Barak *et al.*, 2016). The fourth lineage includes
97 strains denoted as *X. axonopodis*, such as pv. *manihotis* (Rademaker group 9.4; (Rademaker et
98 al., 2005, Mhedbi-Hajri *et al.*, 2013). The fifth lineage includes *X. translucens* (Langlois *et al.*,
99 2017), within the distinct cluster of "early-branching" species whose divergence from the
100 remainder apparently occurred relatively early during *Xanthomonas* speciation (Parkinson et al.,
101 2007). The sixth lineage comprises *X. maliensis*, associated with but nonpathogenic on rice
102 (Triplett et al., 2015). Phylogenetic analyses place this species between the "early-branching"
103 species and the remainder (Triplett et al., 2015).

104 Notably, the *raxX-raxSTAB* gene cluster is absent from the group of strains classified as *X. citri*
105 pathovars (Rademaker group 9.5; (Rademaker et al., 2005, Bansal *et al.*, 2017). These strains
106 (some of which are denoted as *X. axonopodis* or *X. campestris*) cluster phylogenetically among
107 four of the *raxX-raxSTAB* gene cluster-positive groups: *X. oryzae*, *X. vasicola*, *X. euvesicatoria*
108 and *X. axonopodis* pv. *manihotis* (Midha & Patil, 2014, Vauterin et al., 1995, Rademaker et al.,
109 2005). The simplest explanation for this pattern is that the *raxX-raxSTAB* gene cluster was lost
110 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 transfer between *Xanthomonas* spp.**

111 Both the organization and size of the *raxX-raxSTAB* gene cluster are conserved across all six
112 lineages. To assess inheritance patterns, we constructed a phylogenetic tree for the *raxX-*
113 *raxSTAB* gene cluster (as the catenation of the four *rax* genes; **Fig. 4**) (Kuo & Ochman, 2009).
114 The *rax* genes in *X. translucens*, in the early-branching group, cluster separately from their
115 homologs in the other lineages. This finding is consistent with the hypothesis that *X. translucens*
116 acquired the *raxX-raxSTAB* gene cluster relatively early during *Xanthomonas* speciation. For *X.*
117 *maliensis*, the *raxX-raxSTAB* genes are most similar to those from *X. euvesicatoria* and the *X.*
118 *axonopodis* pathovars *manihotis* and *phaseoli* (**Fig. 4**), even though the *X. maliensis* genome
119 sequence itself is more distantly related (**Fig. 3**). This finding suggests that *X. maliensis* acquired
120 the *raxX-raxSTAB* gene cluster relatively late during *Xanthomonas* speciation.

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

121 The *raxX-raxSTAB* gene cluster lies between two core (housekeeping) genes (**Fig. 1**). One,
122 *gcvP*, encodes the pyridoxal-phosphate subunit of glycine dehydrogenase. An approximately
123 170 nt riboswitch (*gcvR* in **Fig. 1**) controls GcvP protein synthesis in response to glycine
124 (Mandal *et al.*, 2004). The other, "*mfsX*", encodes a **m**ajor **f**acilitator **s**ubfamily (MFS)
125 transporter related to Bcr and CflA efflux proteins (da Silva et al., 2004). Here, "*mfsX*" is only a
126 provisional designation absent functional characterization.

127 Comparing the *gcvP - [raxX-raxSTAB] -* "*mfsX*" region from the reference genomes reveals
128 sharp boundaries flanking the position of the *raxX-raxSTAB* gene cluster. On the left flank,
129 substantial nucleotide identity spans the *gcvP* gene, the *gcvR* riboswitch, and a predicted *gcvR*
130 promoter –10 element (Mitchell *et al.*, 2003) (**Fig. S2**). On the right flank, identity begins
131 shortly after the "*mfsX*" initiation codon. Accordingly, upstream sequence elements for initiating
132 "*mfsX*" gene transcription (Mitchell et al., 2003) and translation (Ma *et al.*, 2002) are conserved
133 within, but not between, *raxX-raxSTAB* gene cluster-positive and -negative sequences (**Fig. S2**).

134 Between these boundaries in *raxX-raxSTAB* gene cluster-negative species, the compact (≤ 200
135 nt) *gcvP*-"*mfsX*" intergenic sequence is modestly conserved in most genomes (about 60-80%
136 overall identity; **Fig. S2**). Much of this identity comes from the "*mfsX*" potential transcription
137 and translation initiation sequences described above. The overall intergenic sequence is less

138 conserved in the early-branching species (*X. albilineans*, *X. hyacinthi* and *X. sacchari*),
139 displaying about 50-65% overall identity.

140 We hypothesize that *raxX-raxSTAB* gene cluster phylogenetic distribution results from general
141 recombination between conserved genes flanking each side (e.g., in or beyond the *gcvP* and
142 *"mfsX"* genes). Two observations are consistent with the hypothesis, First, we observed that the
143 sequences flanking the *raxX-raxSTAB* gene cluster are different from the *gcvP*-"*mfsX*" intergenic
144 sequence in *raxX-raxSTAB* gene cluster-negative strains (**Fig. S2**). This argues against models in
145 which the *raxX-raxSTAB* gene cluster has integrated into the *gcvP*-"*mfsX*" intergenic sequence
146 during lateral transfer events.

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

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

**_raxST_ but not _raxX_ homologs are present in genomes from diverse bacterial species**

155 Our GenBank database searches identified *raxX* homologs and the *raxX-raxSTAB* gene cluster
156 only in *Xanthomonas* spp. However, these searches did identify *raxST* homologs encoding
157 proteins with about 40% identity to, and approximately the same length as, the *Xoo* RaxST
158 protein. These sequences include the PAPS-binding motifs that define sulfotransferase activity
159 (da Silva et al., 2004, Negishi *et al.*, 2001). Regardless of its current function, a *raxST* homolog
160 potentially could evolve to encode tyrosylprotein sulfotransferase activity.

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

167 These *raxST* homologs are in diverse genetic contexts in a range of bacterial phyla including
168 Proteobacteria and Cyanobacteria (**Fig. S4**). Nevertheless, for most species represented by
169 multiple genome sequences, the *raxST* homolog was detected in a minority of individuals, so it is
170 not part of the core genome in these strains. Moreover, relationships between species in a *raxST*
171 gene phylogenetic tree bear no resemblance to those in the overall tree of bacterial species. For
172 example, in the *raxST* gene tree, sequences from Cyanobacteria are flanked on both sides by
173 sequences from Proteobacteria (**Fig. S4**). Together, these findings provide evidence for lateral
174 transfer of *raxST* homologs (Kuo & Ochman, 2009).

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

175 RaxX protein sequences from diverse *Xanthomonas* spp. assort into several sequence groups
176 differentiated by polymorphisms within the predicted mature peptide sequence (**Fig. 2**) (Pruitt et
177 al., 2017). Many of these groups are subdivided further according to polymorphisms in the
178 predicted leader protein sequence (residues 1-39) or carboxyl-terminal region distal to residue
179 Pro-52. Most leader polymorphisms lie between residues 2-24, and are unlikely to affect
180 function of mature RaxX protein. Here we only consider polymorphisms in the predicted mature
181 form.

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

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

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

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

210 In a separate test for activation of XA21-mediated immunity, we used a □*raxX* deletion mutant
211 of *Xoo* strain PXO99$^{A}$ as a host for genetic complementation. We tested each of the *raxX* alleles
212 shown in **Fig. 2**, which includes examples from lower frequency (mostly unique) sequence
213 groups. We introduced each *raxX* allele into the □*raxX* test strain, and monitored disease
214 progression in leaves of whole plants. Only the group A *raxX* allele (from *Xoo* strain PXO99$^{A}$)
215 was able to complement the *Xoo* PXO99$^{A}$ □*raxX* strain to activate XA21-mediated immunity
216 (**Fig. 6**). Expression of each *raxX* allele was confirmed by qPCR (**Fig. S5**).

217 Together, these results provide direct evidence that activation of XA21-mediated immunity is
218 restricted to RaxX proteins from sequence group A, found in most strains of *Xoo*. None of the
219 other *X. oryzae* RaxX variants tested (including RaxX from *X. oryzae* pv. *oryzicola*) was able to
220 activate XA21-mediated immunity. The observation that all RaxX proteins tested stimulated
221 *Arabidopsis* root growth suggests that the RaxX PSY peptide mimicry function is not restricted
222 to rice.

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

223 The *raxX* alleles from *Xoo* strains IXO685 and AXO1947 failed to complement the □*raxX*
224 mutant of *Xoo* strain PXO99$^{A}$ for XA21 immune activation (**Fig. 6**). In addition to its variant
225 *raxX* allele (**Fig. 2**), we noted that *Xoo* strain AXO1947 (Huguet-Tapia *et al.*, 2016) carries seven
226 missense polymorphisms in the *raxST* gene (**Fig. S6**) not present in other *Xoo* strains such as

227 IXO685. To determine if the variant *raxST* allele from strain AXO1947 encodes a functional
228 protein, we performed additional complementation tests.

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

236 Taken together, these results suggest that *Xoo* strain AXO1947 has mutant versions of both
237 genes, *raxST* and *raxX*. Analysis by qRT-PCR confirms that these genes were expressed in the
238 complemented strains (**Fig. S7**).

**RaxST variants from *Xoo* strain AXO1947**

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

245 Little is known about RaxST structure and function. Diverse sulfotransferases share limited
246 sequence similarity, mostly comprising two relatively short sequence motifs involved in PAPS
247 binding (Negishi et al., 2001). These motifs are conserved in the *Xoo* RaxST sequence (da Silva
248 et al., 2004). Research with diverse sulfotransferases has identified three essential residues: a
249 positively-charged residue (corresponding to Arg-11 in RaxST) in one PAPS binding motif, an
250 invariant Ser (corresponding to Ser-118 in RaxST) in the other, and a catalytic base (His or Glu)
251 located between the two PAPS binding motifs (Negishi et al., 2001).

252 We generated a RaxST molecular model with the program iTasser (Yang & Zhang, 2015) using
253 the crystal structure of human tyrosylprotein sulfotransferase-2 (TPST2) as a template (PDB:
254 3AP1). The sequence alignment is shown in **Fig. S8**. TPST2 is a functional dimer (Teramoto *et*
255 *al.*, 2013), which is replicated in the RaxST structural model (**Fig. S9)**. The two essential
256 residues identified from *Xoo* strain AXO1947, His-50 and Arg-129, display surface exposed side
257 chains in close proximity to the corresponding position for the bound substrate peptide co-
258 crystalized with TPST2. These residues are distal to the catalytic site. Therefore, we
259 hypothesize that these RaxST residues are involved in RaxX peptide binding.

## DISCUSSION

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

### Opposing selection pressures drive RaxX natural variation

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

275 The *Xa21* gene has been introgressed into commercial rice varieties (Khush *et al.*, 1990, Midha
276 *et al.*, 2017). Widespread planting of *Xa21* rice presumably increases selection for *Xoo* variants

277 that evade XA21-mediated immunity. All RaxX missense variants examined mimicked PSY
278 hormone activity (**Fig. 5A** and **Fig. 5B**) (Pruitt et al., 2017), suggesting that this property confers
279 a selective advantage. Consistent with this, we did not observe any *raxX* frameshift or nonsense
280 alterations. Instead, RaxX variant sequences contain a restricted set of missense substitutions,
281 consistent with the hypothesis that selection acts to retain PSY-like function (**Fig. 2**; see
282 reference (Pruitt et al., 2017)).

283 Among all RaxX variants tested, only that from *Xoo* strain PXO99$^{A}$ (which represents the large
284 majority of *Xoo raxX* alleles) activated the XA21-mediated immune response (**Fig. 5C** and **Fig.**
285 **6**). This result demonstrates that recognition of RaxX by XA21 is strictly limited to *Xoo*, and
286 confirms and extends a prior conclusion from our laboratory, that residues Pro-44 and Pro-48
287 both are required for *Xoo* RaxX recognition by XA21 (Pruitt et al., 2015).

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

294 We also have identified *raxST* and/or *raxA* gene loss of function alterations in *Xoo* field isolates
295 (**Fig. 7**; reference (da Silva et al., 2004), which presumably cannot express the PSY mimicry

296 phenotype of RaxX). Such loss of function alterations could temper the effectiveness of
297 production strategies that rely on engineered *Xa21* alleles.

### ***raxX-raxSTAB* gene cluster origin**

298 The *raxAB* genes are homologous to those encoding proteolytic maturation and ATP-dependent
299 peptide secretion complexes (da Silva et al., 2004, Lin *et al.*, 2015), related to type 1 secretion
300 systems but specialized for secreting small peptides such as bacteriocins and peptide pheromones
301 (Holland et al., 2016). Frequently, the gene encoding the secreted substrate is adjacent to genes
302 encoding components of the secretion complex (Dirix *et al.*, 2004). We hypothesize that the
303 intact *raxX-raxSTAB* gene cluster originated in an ancestor to the lineage containing *X. oryzae*,
304 *X. euvesicatoria*, and related species, with subsequent gains or loss through lateral transfer (**Fig.**
305 **2**). Relatively few events appear to have been necessary to form the *raxX-raxSTAB* gene cluster.
306 The *raxX* gene might have evolved from the gene for the secreted peptide substrate of the
307 RaxAB ancestor. The complete cluster would result from incorporation of the ancestral *raxST*
308 gene, homologs of which are distributed broadly (**Fig. S4**).

### **Role for the *raxX-raxSTAB* gene cluster in *Xanthomonas* biology**

309 The *raxX-raxSTAB* gene cluster does not exhibit features, such as a gene for a site-specific
310 recombinase, characteristic of self-mobile genomic islands (Hacker *et al.*, 1997). Instead,
311 evidence suggests that *raxX-STAB* gene cluster lateral transfer occurred through general
312 recombination between genes flanking each side of the *raxX-STAB* gene cluster (**Fig. 1** and **Fig.**

313 S2). In bacteria, gene acquisition through lateral transfer contributes to emergence of new
314 pathovars (see reference (Ogura *et al.*, 2009) for one example). Conceivably, lateral acquisition
315 of the *raxX-raxSTAB* gene cluster might allow a particular strain to infect a previously
316 inaccessible host.

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

## Experimental Procedures

### Survey of the RaxX, RaxST and the *raxX-STAB* genomic region in publicly available databases

328 We used the 5kb long *Xoo* PXO99$^{A}$ *raxX-raxSTAB* genomic region, including 600 bp upstream
329 of *raxST* and 70 bp downstream of *raxB*, as query to search the following NCBI databases with
330 blastn and megablast using e-value cut-off of 1e-3; nr/nt, htgs,
331 refseq_genomic_representative_genomes, refseq_genomic, and gss. To identify RaxX homologs
332 we used the protein sequence of RaxX from *Xoo* PXO99$^{A}$ as query to search the same databases
333 using tblastn with a PAM30 scoring matrix to account for the short sequence length of RaxX. In
334 case of *raxST* from *Xoo* PXO99$^{A}$ we used the genomic coding sequence to search the same
335 databases using the same cut-offs. In addition, we used the RaxST protein sequence to search the
336 following database using blastp with an e-value cut-off of 1e-3 and a BLOSUM62 scoring
337 matrix; nr, refseq_protein, env_nr. The databases were last accessed 2016/01/06 for the initial
338 manuscript submission and 2018/06/25 during preparation of the resubmission. Searches were
339 restricted to bacteria (taxid: 2) in case of refseq_genomic_representative_genomes. The
340 observations of specificity of *raxX* and the intact *raxX-raxSTAB* gene cluster to the genus
341 *Xanthomonas* was consistent across all queries.

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

342 All available *Xanthomonas* genomes were downloaded from the NCBI ftp server on January 29,
343 2016 (413 genome accessions). The genome fasta files were used to build a local blast database
344 using BLASTv2.27+ (Camacho *et al.*, 2009). For all genes in and surrounding the *raxSTAB*
345 cluster blastn (evalue cutoff of 1e-3) was used to identify homologs in the local blast database.
346 Due to the small size of RaxX, tblastn was required to identify homologs (evalue cutoff of 1e-3).
347 Fasta files for each blast hit were generated using a custom python script (available upon
348 request). Alignments of all genes were performed with Muscle v3.5 (Edgar, 2004) implemented
349 in the desktop tool Geneious v9.1.8 (Kearse *et al.*, 2012). Alignment ends were trimmed so that
350 each sequence was equal in length and in the first coding frame. Maximum likelihood trees were
351 built with RaxML v8.2.4 (Stamatakis, 2014) with the following settings: (-m GTRGAMMA F -f
352 a -x 3298589 -N 10000 -p 23). Trees shown in all figures are the highest scoring ML tree and
353 numbers shown on branches are the resampled bootstrap values from 1000 replicates. Trees
354 were drawn in FigTree v1.4.0 (http://tree.bio.ed.ac.uk/software/figtree/).

355 Whole genome phylogenies were generated using the entire genome assembly with the program
356 Andi v0.10 (Haubold *et al.*, 2015, Klotzl & Haubold, 2016). These distance matrices were
357 plotted as neighbor-joining tree using Phylip v3.695 (Felsenstein, 1981). Numbers on the
358 branches represent the proportion (0-100) that the branch appeared in the "bootstrapped"
359 distance matrices using Andi.

**Sequence analyses**

360 Nucleotide and deduced amino acid sequences were edited and analyzed with the programs
361 EditSeq™ (version 14.1.0), MegAlign™ (version 14.1.0) and SeqBuilder™ (version 14.1.0),
362 DNASTAR, Madison, WI. The Integrated Microbial Genomes interface (Chen *et al.*, 2017) was
363 used to compare genome segments from different species.

**Bacterial growth**

364 *Xanthomonas* strains were cultured at 28°C. Solid medium was peptone sucrose agar (PSA; pH
365 7.0), which contains (per liter) peptone (10 g), sucrose (10 g), sodium glutamate (1 g) and agar
366 (15 g). Liquid cultures were aerated at 230 rpm in YEB medium (pH 7.3), which contains (per
367 liter) yeast extract (5 g), tryptone (10 g), NaCl (5 g), sucrose (5 g), and $MgSO_4$ (0.5 g).
368 Antibiotics were kanamycin, carbenicillin, spectinomycin (all at 50 □g/ml), and cephalexin (20
369 □g/ml).

**Rice growth and inoculation**

370 *Oryza sativa* ssp. *japonica* rice varieties were TP309 and XA21-TP309, which is a 106-17-
371 derived transgenic line of TP309 carrying the *Xa21* gene expressed from its native promoter
372 (Song et al., 1995). TP309 rice does not contain the *Xa21* gene. Seeds were germinated in
373 distilled water at 28°C for one week and then transplanted into sandy soil (80% sand, 20% peat;
374 Redi-Gro) in 5.5-inch square pots with two seedlings per pot. Plants were grown in tubs in a
375 greenhouse, and were top watered daily with fertilizer water [N, 58 ppm (parts per million); P,
376 15 ppm; K, 55 ppm; Ca, 20 ppm; Mg, 13 ppm; S, 49 ppm; Fe, 1 ppm; Cu, 0.06 ppm; Mn, 0.4

377 ppm; Mo, 0.02 ppm; Zn, 0.1 ppm; B, 0.4 ppm] for four weeks, followed by water for two weeks.
378 Six weeks after planting, rice pots were transferred to a growth chamber with the following
379 day/night settings: 28°C/24°C, 80%/85% humidity, and 14/10-hour lighting. Plants were
380 inoculated 2 to 3 days after transfer using the scissors clipping method (Song et al., 1995).
381 Bacteria for inoculation were taken from PSA plates and resuspended in water at a density of
382 approximately $8 \times 10^8$ CFU/ml. Water-soaked lesions were measured 14 days after inoculation.

**Complementation tests**

383 The *Xoo* strain PXO99$^A$ marker-free deletions □*raxX* and □*raxST* were described previously
384 (Pruitt et al., 2015). The *raxX* and *raxST* genes from different *Xanthomonas* spp. were cloned
385 into plasmid vector pVS61 and electrotransformed into the appropriate recipient strains as
386 described previously (Pruitt et al., 2015). Site-specific mutational alterations were introduced by
387 PCR using the In-Fusion HD cloning system (Takara).

**RaxX peptide stimulation of *PR10b* gene expression**

388 Full-length sulfated RaxX proteins were purified from an *E. coli* strain with an expanded genetic
389 code that directs incorporation of sulfotyrosine at the appropriate position (Schwessinger *et al.*,
390 2016). The resulting MBP-3C–RaxX-His fusion proteins were incubated with 3C protease
391 followed by anion exchange chromatography in order to remove the amino-terminal maltose
392 binding protein tag, as described previously (Schwessinger et al., 2016). The control peptide,
393 sulfated RaxX21-sY, has been described (Pruitt et al., 2015).

394 Rice plants were grown in a hydroponic system in growth chambers at 24° or 28°C with a 14-
395 hour/10-hour light-dark cycle at 80% humidity. Seedlings were grown in A-OK Starter Plugs
396 (Grodan) and watered with Hoagland's solution twice a week. Peptide influence on *PR10b*
397 marker gene expression was measured as described previously (Pruitt et al., 2015). Briefly,
398 leaves of 4-week-old hydroponically grown rice plants were cut into 2-cm-long strips and
399 incubated for at least 12 hours in $ddH_2O$ to reduce residual wound signals. Leaf strips were
400 treated with the indicated peptides and then snap-frozen in liquid nitrogen before processing.
401 Quantitative reverse transcription polymerase chain reaction (qRT-PCR) was done as described
402 previously (Pruitt et al., 2015). Gene expression was normalized to the actin gene expression
403 level and to the respective mock-treated control at 0 or 9 hour.

404 DNA primers for qRT-PCR were: ampC-F, GACTCGTAATGCCTACGACC; ampC-R,
405 AATTGCTCGTAGAAGCTGCC; qraxST-F, CTTCCAACGTGCAGATCGAC; qraxST-R,
406 TATCGACGATCCAACCAAC; qRaxX-F, AAAATCGCCCGCCAAGGGT; qRaxX-R,
407 TCAATGGTGCCCGGGGTTG; PR10b-F, TGTGGAAGGTCTGCTTGGAC; PR10b-R,
408 CCTTTAGCACGTGAGTTGCG

**Acknowledgements**

Supported by NIH GM59962 and GM122968 to Pamela C. Ronald. We are grateful to anonymous reviewer 2 for homology modeling. 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 modeling.

## REFERENCES


413 Amano, Y., Tsubouchi, H., Shinohara, H., Ogawa, M. and Matsubayashi, Y. (2007) Tyrosine-
414 sulfated glycopeptide involved in cellular proliferation and expansion in *Arabidopsis*. *Proc*
415 *Natl Acad Sci U S A,* **104,** 18333-18338.
416 Aritua, V., Parkinson, N., Thwaites, R., Heeney, J. V., Jones, D. R., Tushemereirwe, W.*, et al.*
417 (2008) Characterization of the *Xanthomonas* sp. causing wilt of enset and banana and its
418 proposed reclassification as a strain of *X. vasicola*. *Plant Pathol.,* **57,** 170-177.
419 Bansal, K., Midha, S., Kumar, S. and Patil, P. B. (2017) Ecological and evolutionary insights
420 into pathovar diversity of Xanthomonas citri. *Appl Environ Microbiol,* **83,** e02993-02916.
421 Barak, J. D., Vancheva, T., Lefeuvre, P., Jones, J. B., Timilsina, S., Minsavage, G. V.*, et al.*
422 (2016) Whole-genome sequences of *Xanthomonas euvesicatoria* strains clarify taxonomy
423 and reveal a stepwise erosion of Type 3 effectors. *Front Plant Sci,* **7,** 1805.
424 Bart, R., Cohn, M., Kassen, A., McCallum, E. J., Shybut, M., Petriello, A.*, et al.* (2012) High-
425 throughput genomic sequencing of cassava bacterial blight strains identifies conserved
426 effectors to target for durable resistance. *Proc Natl Acad Sci U S A,* **109,** E1972-1979.
427 Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K.*, et al.* (2009)
428 BLAST+: architecture and applications. *BMC Bioinformatics,* **10,** 421.
429 Chen, I. A., Markowitz, V. M., Chu, K., Palaniappan, K., Szeto, E., Pillay, M.*, et al.* (2017)
430 IMG/M: integrated genome and metagenome comparative data analysis system. *Nucleic*
431 *Acids Res,* **45,** D507-D516.
432 Crossman, L. C., Gould, V. C., Dow, J. M., Vernikos, G. S., Okazaki, A., Sebaihia, M.*, et al.*
433 (2008) The complete genome, comparative and functional analysis of *Stenotrophomonas*
434 *maltophilia* reveals an organism heavily shielded by drug resistance determinants. *Genome*
435 *Biol,* **9,** R74.
436 da Silva, A. C., Ferro, J. A., Reinach, F. C., Farah, C. S., Furlan, L. R., Quaggio, R. B.*, et al.*
437 (2002) Comparison of the genomes of two *Xanthomonas* pathogens with differing host
438 specificities. *Nature,* **417,** 459-463.
439 da Silva, F. G., Shen, Y., Dardick, C., Burdman, S., Yadav, R. C., de Leon, A. L.*, et al.* (2004)
440 Bacterial genes involved in type I secretion and sulfation are required to elicit the rice Xa21-
441 mediated innate immune response. *Mol Plant Microbe Interact,* **17,** 593-601.
442 Dirix, G., Monsieurs, P., Dombrecht, B., Daniels, R., Marchal, K., Vanderleyden, J.*, et al.* (2004)
443 Peptide signal molecules and bacteriocins in Gram-negative bacteria: a genome-wide in
444 silico screening for peptides containing a double-glycine leader sequence and their cognate
445 transporters. *Peptides,* **25,** 1425-1440.
446 Edgar, R. C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high
447 throughput. *Nucleic Acids Res,* **32,** 1792-1797.
448 Farzan, M., Mirzabekov, T., Kolchinsky, P., Wyatt, R., Cayabyab, M., Gerard, N. P.*, et al.*
449 (1999) Tyrosine sulfation of the amino terminus of CCR5 facilitates HIV-1 entry. *Cell,* **96,**
450 667-676.
451 Felsenstein, J. (1981) Evolutionary trees from DNA sequences: a maximum likelihood approach.
452 *J Mol Evol,* **17,** 368-376.
453 Ferreira-Tonin, M., Rodrigues-Neto, J., Harakava, R. and Destefano, S. A. (2012) Phylogenetic
454 analysis of *Xanthomonas* based on partial *rpoB* gene sequences and species differentiation by
455 PCR-RFLP. *Int J Syst Evol Microbiol,* **62,** 1419-1424.

456 Gardiner, D. M., Upadhyaya, N. M., Stiller, J., Ellis, J. G., Dodds, P. N., Kazan, K., *et al.* (2014)
457 Genomic analysis of *Xanthomonas translucens* pathogenic on wheat and barley reveals
458 cross-kingdom gene transfer events and diverse protein delivery systems. *PLoS One,* **9,**
459 e84995.
460 Hacker, J., Blum-Oehler, G., Mühldorfer, I. and Tschäpe, H. (1997) Pathogenicity islands of
461 virulent bacteria: structure, function and impact on microbial evolution. *Mol Microbiol,* **23,**
462 1089-1097.
463 Han, S. W., Lee, S. W., Bahar, O., Schwessinger, B., Robinson, M. R., Shaw, J. B., *et al.* (2012)
464 Tyrosine sulfation in a Gram-negative bacterium. *Nature communications,* **3,** 1153.
465 Hauben, L., Vauterin, L., Swings, J. and Moore, E. R. (1997) Comparison of 16S ribosomal
466 DNA sequences of all *Xanthomonas* species. *Int J Syst Bacteriol,* **47,** 328-335.
467 Haubold, B., Klotzl, F. and Pfaffelhuber, P. (2015) andi: fast and accurate estimation of
468 evolutionary distances between closely related genomes. *Bioinformatics,* **31,** 1169-1175.
469 Holland, I. B., Peherstorfer, S., Kanonenberg, K., Lenders, M., Reimann, S. and Schmitt, L.
470 (2016) Type I protein secretion-deceptively simple yet with a wide range of mechanistic
471 variability across the family. *EcoSal Plus,* **7,** doi:10.1128/ecosalplus.ESP-0019-2015.
472 Huguet-Tapia, J. C., Peng, Z., Yang, B., Yin, Z., Liu, S. and White, F. F. (2016) Complete
473 Genome Sequence of the African Strain AXO1947 of Xanthomonas oryzae pv. oryzae.
474 *Genome announcements,* **4**.
475 Jacobs, J. M., Pesce, C., Lefeuvre, P. and Koebnik, R. (2015) Comparative genomics of a
476 cannabis pathogen reveals insight into the evolution of pathogenicity in *Xanthomonas*. *Front*
477 *Plant Sci,* **6,** 431.
478 Jacques, M. A., Arlat, M., Boulanger, A., Boureau, T., Carrere, S., Cesbron, S., *et al.* (2016)
479 Using ecology, physiology, and genomics to understand host specificity in *Xanthomonas*.
480 *Annu Rev Phytopathol,* **54,** 163-187.
481 Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., *et al.* (2012)
482 Geneious Basic: an integrated and extendable desktop software platform for the organization
483 and analysis of sequence data. *Bioinformatics,* **28,** 1647-1649.
484 Khush, G. S., Bacalangco, E. and Ogawa, T. (1990) A new gene for resistance to bacterial blight
485 from *O. longistaminata*. *Rice Genet Newsl,* **7,** 121-122.
486 Kleist, A. B., Getschman, A. E., Ziarek, J. J., Nevins, A. M., Gauthier, P. A., Chevigne, A., *et al.*
487 (2016) New paradigms in chemokine receptor signal transduction: Moving beyond the two-
488 site model. *Biochem Pharmacol,* **114,** 53-68.
489 Klotzl, F. and Haubold, B. (2016) Support values for genome phylogenies. *Life (Basel),* **6,** 11.
490 Kuo, C. H. and Ochman, H. (2009) The fate of new bacterial genes. *FEMS Microbiol Rev,* **33,**
491 38-43.
492 Langlois, P. A., Snelling, J., Hamilton, J. P., Bragard, C., Koebnik, R., Verdier, V., *et al.* (2017)
493 Characterization of the *Xanthomonas translucen*s complex using draft genomes, comparative
494 genomics, phylogenetic analysis, and diagnostic LAMP assays. *Phytopathology,* **107,** 519-
495 527.
496 Lin, D. Y., Huang, S. and Chen, J. (2015) Crystal structures of a polypeptide processing and
497 secretion transporter. *Nature,* **523,** 425-430.
498 Ma, J., Campbell, A. and Karlin, S. (2002) Correlations between Shine-Dalgarno sequences and
499 gene features such as predicted expression levels and operon structures. *J Bacteriol,* **184,**
500 5733-5745.

501 Mandal, M., Lee, M., Barrick, J. E., Weinberg, Z., Emilsson, G. M., Ruzzo, W. L., *et al.* (2004)
502 A glycine-dependent riboswitch that uses cooperative binding to control gene expression.
503 *Science,* **306,** 275-279.
504 Matsubayashi, Y. (2014) Posttranslationally modified small-peptide signals in plants. *Annu Rev*
505 *Plant Biol,* **65,** 385-413.
506 Mhedbi-Hajri, N., Hajri, A., Boureau, T., Darrasse, A., Durand, K., Brin, C., *et al.* (2013)
507 Evolutionary history of the plant pathogenic bacterium *Xanthomonas axonopodis*. *PloS one,*
508 **8,** e58474.
509 Midha, S., Bansal, K., Kumar, S., Girija, A. M., Mishra, D., Brahma, K., *et al.* (2017) Population
510 genomic insights into variation and evolution of *Xanthomonas oryzae* pv. *oryzae*. *Sci Rep,* **7,**
511 40694.
512 Midha, S. and Patil, P. B. (2014) Genomic insights into the evolutionary origin of *Xanthomonas*
513 *axonopodis* pv. *citri* and its ecological relatives. *Appl Environ Microbiol,* **80,** 6266-6279.
514 Mitchell, J. E., Zheng, D., Busby, S. J. and Minchin, S. D. (2003) Identification and analysis of
515 'extended -10' promoters in *Escherichia coli*. *Nucleic Acids Res,* **31,** 4689-4695.
516 Moore, E. R., Kruger, A. S., Hauben, L., Seal, S. E., Daniels, M. J., De Baere, R., *et al.* (1997)
517 16S rRNA gene sequence analyses and inter- and intrageneric relationships of *Xanthomonas*
518 species and *Stenotrophomonas maltophilia*. *FEMS Microbiol Lett,* **151,** 145-153.
519 Naushad, S., Adeolu, M., Wong, S., Sohail, M., Schellhorn, H. E. and Gupta, R. S. (2015) A
520 phylogenomic and molecular marker based taxonomic framework for the order
521 *Xanthomonadales*: proposal to transfer the families *Algiphilacea*e and *Solimonadaceae* to the
522 order *Nevskiales* ord. nov. and to create a new family within the order *Xanthomonadales*, the
523 family *Rhodanobacteraceae* fam. nov., containing the genus *Rhodanobacter* and its closest
524 relatives. *Antonie Van Leeuwenhoek,* **107,** 467-485.
525 Negishi, M., Pedersen, L. G., Petrotchenko, E., Shevtsov, S., Gorokhov, A., Kakuta, Y., *et al.*
526 (2001) Structure and function of sulfotransferases. *Arch Biochem Biophys,* **390,** 149-157.
527 Nelson, K., Wang, F. S., Boyd, E. F. and Selander, R. K. (1997) Size and sequence
528 polymorphism in the isocitrate dehydrogenase kinase/phosphatase gene (*aceK*) and flanking
529 regions in *Salmonella enterica* and *Escherichia coli*. *Genetics,* **147,** 1509-1520.
530 O'Leary, N. A., Wright, M. W., Brister, J. R., Ciufo, S., Haddad, D., McVeigh, R., *et al.* (2016)
531 Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and
532 functional annotation. *Nucleic Acids Res,* **44,** D733-745.
533 Ogura, Y., Ooka, T., Iguchi, A., Toh, H., Asadulghani, M., Oshima, K., *et al.* (2009)
534 Comparative genomics reveal the mechanism of the parallel evolution of O157 and non-
535 O157 enterohemorrhagic *Escherichia coli*. *Proc Natl Acad Sci U S A,* **106,** 17939-17944.
536 Parkinson, N., Aritua, V., Heeney, J., Cowie, C., Bew, J. and Stead, D. (2007) Phylogenetic
537 analysis of *Xanthomonas* species by comparison of partial gyrase B gene sequences. *Int J*
538 *Syst Evol Microbiol,* **57,** 2881-2887.
539 Parkinson, N., Cowie, C., Heeney, J. and Stead, D. (2009) Phylogenetic structure of
540 *Xanthomonas* determined by comparison of *gyrB* sequences. *Int J Syst Evol Microbiol,* **59,**
541 264-274.
542 Pereira, U. P., Gouran, H., Nascimento, R., Adaskaveg, J. E., Goulart, L. R. and Dandekar, A. M.
543 (2015) Complete genome sequence of *Xanthomonas arboricola* pv. *juglandis* 417, a copper-
544 resistant strain isolated from *Juglans regia* L. *Genome announcements,* **3,** e01126-01115.

545 Pieretti, I., Cociancich, S., Bolot, S., Carrere, S., Morisset, A., Rott, P., *et al.* (2015) Full genome
546 sequence analysis of two isolates reveals a novel Xanthomonas species close to the sugarcane
547 pathogen *Xanthomonas albilineans*. *Genes (Basel),* **6,** 714-733.
548 Pruitt, R. N., Joe, A., Zhang, W., Feng, W., Stewart, V., Schwessinger, B., *et al.* (2017) A
549 microbially derived tyrosine-sulfated peptide mimics a plant peptide hormone. *New Phytol,*
550 **215,** 725-736.
551 Pruitt, R. N., Schwessinger, B., Joe, A., Thomas, N., Liu, F., Albert, M., *et al.* (2015) The rice
552 immune receptor XA21 recognizes a tyrosine-sulfated protein from a Gram-negative
553 bacterium. *Sci Adv,* **1,** e1500245.
554 Rademaker, J. L. W., Louws, F. J., Schultz, M. H., Rossbach, U., Vauterin, L., Swings, J., *et al.*
555 (2005) A comprehensive species to strain taxonomic framework for Xanthomonas.
556 *Phytopathology,* **95,** 1098-1111.
557 Robert, X. and Gouet, P. (2014) Deciphering key features in protein structures with the new
558 ENDscript server. *Nucleic Acids Res,* **42,** W320-324.
559 Ronald, P. C. and Beutler, B. (2010) Plant and animal sensors of conserved microbial signatures.
560 *Science,* **330,** 1061-1064.
561 Salzberg, S. L., Sommer, D. D., Schatz, M. C., Phillippy, A. M., Rabinowicz, P. D., Tsuge, S., *et*
562 *al.* (2008) Genome sequence and rapid evolution of the rice pathogen Xanthomonas oryzae
563 pv. oryzae PXO99A. *BMC Genomics,* **9,** 204.
564 Schwessinger, B., Li, X., Ellinghaus, T. L., Chan, L. J., Wei, T., Joe, A., *et al.* (2016) A second-
565 generation expression system for tyrosine-sulfated proteins and its application in crop
566 protection. *Integrative biology : quantitative biosciences from nano to macro,* **8,** 542-545.
567 Shen, Y., Sharma, P., da Silva, F. G. and Ronald, P. (2002) The *Xanthomonas oryzae* pv. *oryzae*
568 *raxP* and *raxQ* genes encode an ATP sulphurylase and adenosine-5'-phosphosulphate kinase
569 that are required for AvrXa21 avirulence activity. *Mol Microbiol,* **44,** 37-48.
570 Song, W. Y., Wang, G. L., Chen, L. L., Kim, H. S., Pi, L. Y., Holsten, T., *et al.* (1995) A
571 receptor kinase-like protein encoded by the rice disease resistance gene, *Xa21*. *Science,* **270,**
572 1804-1806.
573 Stamatakis, A. (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of
574 large phylogenies. *Bioinformatics,* **30,** 1312-1313.
575 Stone, M. J., Chuang, S., Hou, X., Shoham, M. and Zhu, J. Z. (2009) Tyrosine sulfation: an
576 increasingly recognised post-translational modification of secreted proteins. *N Biotechnol,*
577 **25,** 299-317.
578 Studholme, D. J., Wasukira, A., Paszkiewicz, K., Aritua, V., Thwaites, R., Smith, J., *et al.* (2011)
579 Draft genome sequences of *Xanthomonas sacchari* and two banana-associated
580 Xanthomonads reveal insights into the *Xanthomonas* group 1 clade. *Genes (Basel),* **2,** 1050-
581 1065.
582 Tang, D., Wang, G. and Zhou, J. M. (2017) Receptor kinases in plant-pathogen interactions:
583 more than pattern recognition. *Plant Cell,* **29,** 618-637.
584 Teramoto, T., Fujikawa, Y., Kawaguchi, Y., Kurogi, K., Soejima, M., Adachi, R., *et al.* (2013)
585 Crystal structure of human tyrosylprotein sulfotransferase-2 reveals the mechanism of protein
586 tyrosine sulfation reaction. *Nat Commun,* **4,** 1572.
587 Thieme, F., Koebnik, R., Bekel, T., Berger, C., Boch, J., Buttner, D., *et al.* (2005) Insights into
588 genome plasticity and pathogenicity of the plant pathogenic bacterium *Xanthomonas*
589 *campestris* pv. *vesicatoria* revealed by the complete genome sequence. *J Bacteriol,* **187,**
590 7254-7266.

591 Thomas, N. C., Schwessinger, B., Liu, F., Chen, H., Wei, T., Nguyen, Y. P.*, et al.* (2016) XA21-
592 specific induction of stress-related genes following Xanthomonas infection of detached rice
593 leaves. *PeerJ,* **4,** e2446.
594 Triplett, L. R., Verdier, V., Campillo, T., Van Malderghem, C., Cleenwerck, I., Maes, M.*, et al.*
595 (2015) Characterization of a novel clade of *Xanthomonas* isolated from rice leaves in Mali
596 and proposal of *Xanthomonas maliensis* sp. nov. *Antonie Van Leeuwenhoek,* **107,** 869-881.
597 Vancheva, T., Bogatzevska, N., Moncheva, P., Lefeuvre, P. and Koebnik, R. (2015) Draft
598 genome sequences of two *Xanthomonas vesicatoria* strains from the balkan peninsula.
599 *Genome announcements,* **3,** e01558-01514.
600 Vandroemme, J., Cottyn, B., Baeyen, S., De Vos, P. and Maes, M. (2013) Draft genome
601 sequence of *Xanthomonas fragariae* reveals reductive evolution and distinct virulence-
602 related gene content. *BMC Genomics,* **14,** 829.
603 Vauterin, L., Hoste, B., Kersters, K. and Swings, J. (1995) Reclassification of *Xanthomonas*. *Int.*
604 *J. Syst. Bacteriol.,* **45**.
605 Vauterin, L., Rademaker, J. and Swings, J. (2000) Synopsis on the taxonomy of the genus
606 *Xanthomonas*. *Phytopathology,* **90,** 677-682.
607 Wang, G. L., Song, W. Y., Ruan, D. L., Sideris, S. and Ronald, P. C. (1996) The cloned gene,
608 Xa21, confers resistance to multiple *Xanthomonas oryzae* pv. *oryzae* isolates in transgenic
609 plants. *Mol Plant Microbe Interact,* **9,** 850-855.
610 Wasukira, A., Tayebwa, J., Thwaites, R., Paszkiewicz, K., Aritua, V., Kubiriba, J.*, et al.* (2012)
611 Genome-wide sequencing reveals two major sub-lineages in the genetically monomorphic
612 pathogen *Xanthomonas campestris* pathovar *musacearum*. *Genes (Basel),* **3,** 361-377.
613 Yang, J. and Zhang, Y. (2015) I-TASSER server: new development for protein structure and
614 function predictions. *Nucleic Acids Res,* **43,** W174-181.
615 Young, J. M. (2008) An overview of bacterial nomenclature with special reference to plant
616 pathogens. *Syst Appl Microbiol,* **31,** 405-424.

## FIGURE LEGENDS

**Fig. 1. The *raxX-raxSTAB* gene cluster.**

617 The *raxX-raxSTAB* gene cluster is located between the flanking *gcvRP* and *"mfsX"* genes. Gene
618 cluster acquisition through lateral transfer is hypothesized to occur by general recombination in
619 the flanking *gcvR* and "*mfsX*" sequences as described in the text. Sequences at the left and right
620 boundaries are shown in **Fig. S2**. Sequences for length polymorphisms in the *gcv*P gene are
621 shown in **Fig. S3**.

**Fig. 2. RaxX variants.**

622 Sequences show the presumed leader-cleaved forms of RaxX, numbered from the beginning of
623 the precursor sequence. The extent of sequence comprising the RaxX16 and RaxX13 synthetic
624 peptides is indicated above the alignment. Residues are shaded according to conservation in
625 PSY sequences (Pruitt et al., 2017): positions with nearly invariant residues are shaded black,
626 and those with only two or three substitutions are shaded blue. The sulfated Tyr residue is
627 shaded red. Gaps are indicated by dots. Sequence groups are described elsewhere in detail
628 (Pruitt et al., 2017). The subgroups B1-B3 differ only in the carboxyl-terminal sequence
629 beginning with residue 53. *X. oryzae* strains X8-1A and X11-5A are nonpathogenic and
630 therefore do not have pathovar designations. The mature form of *Arabidopsis thaliana* PSY1
631 (Amano et al., 2007) and the corresponding region from *Oryza sativa* PSY1a (Amano et al.,
632 2007, Pruitt et al., 2017) are shown for comparison. Residues Pro-16 and Pro-17 in AtPSY1
633 both are hydroxylated [†,‡], and Pro-16 is glycosylated with L-$Ara_3$ [‡] (Amano et al., 2007) .

**Fig. 3. Model for raxX-raxSTAB inheritance during Xanthomonas speciation.**

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

**Fig. 4. Phylogenetic tree for raxX-raxSTAB nucleotide sequences.**

642 The best scoring maximum likelihood tree for the catenated *raxA*, *raxB*, *raxX* and *raxST* coding
643 sequences. Numbers shown on branches represent the proportion of branches supported by
644 10,000 bootstrap replicates (0-100). Bootstraps are not shown for branches with less than 50%
645 support, nor for branches too short to easily distinguish. Species names are colored according to
646 phylogenetic group.

**Fig. 5. RaxX variant peptides promote root growth.**

647 (A) Stimulation of *Arabidopsis* root growth. Fourteen-day-old *tpst-1* seedlings were grown on
648 ½ MS vertical plates with or without 100 nM of the indicated full-length peptides. Bars indicate
649 the average seedling root length measured after 14 d (n>10). Error bars show the standard
650 deviation. The "*" indicates a statistically significant difference from Mock using Dunnett's test
651 ($p<0.05$). Peptide RaxX sY21 is a 21 residue sulfated peptide with potent RaxX activity (Pruitt

652 et al., 2015). Strain abbreviations are *Xvv, X. vasicola* pv. *vasculorum*; *Xt*, *X. translucens*; *Xe*, *X.*
653 *euvesicatoria*; *Xcm*, *X. campestris* pv. *musacearum*; PXO99$^{A}$, IXO685, AXO1947, strains of *X.*
654 *oryzae* pv. *oryzae*. (B) *Arabidopsis seedlings* from a representative experiment. (C) Activation
655 of rice *PR10b* gene expression. Purified peptide (500 nM) was used to treat detached leaves as
656 described in Materials and Methods. Expression levels of the *PR10b* gene (normalized to actin
657 gene expression) were determined after 12 h. Data are the mean values from four biological
658 replicates. Error bars show the standard deviation. The "*" indicates a statistically significant
659 difference from Mock using Dunnett's test ($p<0.05$).

**Fig. 6. RaxX variants fail to activate XA21-mediated immunity.**

660 Different *raxX* genes were cloned into vector pVSP6 (see Materials and Methods) to test for
661 complementation of the *Xoo* strain PXO99$^{A}$ Δ*raxX* strain. Leaf tips of rice varieties TP309
662 (panel A) or XA21-expressing TP309 (panel B) were inoculated by clipping with scissors dipped
663 in bacterial suspensions (approximate cell density of $8 \times 10^{8}$ cells $mL^{-1}$). Lesion lengths were
664 measured 14 days after inoculation. Data are the mean values from measurements of 10-20
665 leaves. Error bars show the standard error of the mean, and "*" indicates a statistically
666 significant difference from *Xoo* strain PXO99$^{A}$ according to Dunnett's multiple comparison
667 procedure ($p<0.05$). Values in panel A are insignificantly different. Strain abbreviations are
668 Xvv, *X. vasicola* pv. *vasculorum*; Xt, *X. translucens*; Xoc, *X. oryzae* pv. *oryzicola*; Xe, *X.*
669 *euvesicatoria*; Xcm, *X. campestris* pv. *musacearum*; X8-1A, X11-5A, strains of *X. oryzae*; M97,
670 *X. maliensis* M97; PXO99$^{A}$, IXO685, AXO1947, strains of *X. oryzae* pv. *oryzae*.

**Fig. 7. The *raxX* and *raxST* genes are dysfunctional in *Xoo* strain AXO1947.**

671 Different combinations of the *raxX* and *raxST* genes were cloned into vector pVSP61 (see
672 Materials and Methods) to test for complementation. Leaf tips of rice varieties TP309 (panels A,
673 C and E) or XA21-expressing TP309 (panels B, D and F) were inoculated by clipping with
674 scissors dipped in bacterial suspensions (approximate cell density of $8 \times 10^8$ cells $mL^{-1}$). Lesion
675 measurements were taken 14 days after inoculation. Data are the mean values from
676 measurements of 10-20 leaves. Error bars show the standard error of the mean, and "*" indicates
677 a statistically significant difference from *Xoo* strain $PXO99^A$ according to Dunnett's multiple
678 comparison procedure ($p<0.05$). Values in panels A, C and E are insignificantly different.
679 Panels A and B show complementation results for the *raxX* gene, panels C and D show results
680 for the *raxST* gene, and panels E and F show results for the combination of both the *raxX* and
681 *raxST* genes. Specific combinations of genes and complementation hosts are described in the
682 figure labels.

**Fig. 8. Two missense substitutions inactivate RaxST in *Xoo* strain AXO1947.**

683 Each of the seven *raxST* missense polymorphisms from *Xoo* strain AXO1947 was introduced
684 singly into the wild-type *raxST* gene from *Xoo* strain $PXO99^A$ (see Materials and Methods).
685 These mutant alleles then were tested for complementation of the *Xoo* strain $PXO99^A$ Δ*raxST*
686 strain. Leaf tips of rice varieties TP309 (panel A) or XA21-expressing TP309 (panel B) were
687 inoculated by clipping with scissors dipped in bacterial suspensions (approximate cell density of
688 $8 \times 10^8$ cells $mL^{-1}$). Lesion measurements were taken 14 days after inoculation. Data are the
689 mean values from measurements of 10-20 leaves. Error bars show the standard error of the
690 mean, and "*" indicates a statistically significant difference from *Xoo* strain $PXO99^A$ according
691 to Dunnett's multiple comparison procedure ($p<0.05$).

**Table 1.** Reference strains for sequence comparisons.

| Species | Strain | *raxX-raxSTAB* | Accession | Reference |
|---|---|---|---|---|
| *S. maltophilia* | K279a | – | NC_010943.1 | (Crossman *et al.*, 2008) |
| *X. albilineans* | GPE PC73 | – | NC_013722.1 | (Pieretti *et al.*, 2015) |
| *X. arboricola* pv. *juglandis* | Xaj 417 | – | NZ_CP012251.1 | (Pereira *et al.*, 2015) |
| *X. axonopodis* pv. *manihotis* | UA536 | + | NZ_AKEQ00000000 | (Bart *et al.*, 2012) |
| *X. campestris* pv. *campestris* | ATCC 33913 | – | NC_003902.1 | (da Silva *et al.*, 2002) |
| *X. campestris* pv. *musacearum* | NCPPB 4392 | + | NZ_AKBI00000000.1 | (Wasukira *et al.*, 2012) |
| *X. cannabis* | NCPPB 2877 | – | NZ_JSZE00000000.1 | (Jacobs *et al.*, 2015) |
| *X. citri* subsp. *citri* | 306 | – | NC_003919.1 | (da Silva et al., 2002) |
| *X. euvesicatoria* | 85-10 | + | NZ_CP017190.1 | (Thieme *et al.*, 2005) |
| *X. fragariae* | LMG 25863 | – | NZ_AJRZ00000000.1 | (Vandroemme *et al.*, 2013) |
| *X. hyacinthi* | DSM 19077 | – | JPLD00000000.1 | (Naushad *et al.*, 2015) |

| | | | |
|---|---|---|---|
| *X. maliensis* | M97 | + | NZ_AQPR00000000.1(Triplett et al., 2015) |
| *X. oryzae* pv. *oryzae* | PXO99[A] | + | NC_010717.2 (Salzberg *et al.*, 2008) |
| *X. sacchari* | R1 | – | NZ_CP010409.1 (Studholme *et al.*, 2011) |
| *X. translucens* | DAR61454 | + | GCA_000334075.1(Gardiner et al., 2014) |
| *X. vesicatoria* | 15b (Vancheva *et al.*, 2015) | – | NZ_JSXZ00000000.1 |

**Supporting Information**

**Fig. S1. Whole genome-based *Xanthomonas* phylogenetic tree.**

692 This tree was constructed by analysis of whole genome sequences as described in Materials and
693 Methods. Blue indicates genomes that contain the *raxX-raxSTAB* gene cluster; red indicates
694 genomes that do not. Group numbers are arbitrary.

**Fig. S2. Sequences flanking the *raxX-raxSTAB* gene cluster.**

695 Sequences are from the reference strains described in **Table 1**. Sequences conserved within a
696 group but different from other groups are colored green ("early-branching" species), brown
697 (*raxX-raxSTAB* cluster-negative strains), or yellow (*raxX-raxSTAB* cluster-positive strains). For
698 presentation, the sequence is divided into Left and Right boundaries. The green and brown
699 sequences are contiguous, whereas the yellow sequences are interrupted by the ca. 5 kb *raxX-*
700 *raxSTAB* gene cluster, depicted as a yellow rectangle. For presentation, approximately 60-80 nt
701 with relatively low similarity were removed from sequence shown in the Right boundary panel.
702 These conceptual deletions are denoted by the number of nt removed in each case. Black
703 sequences are conserved in all lineages, and include both coding regions as well as matches to
704 transcription and translation initiation consensus sequences, which are described in the text. An
705 "*mfsX*" +1 frameshift in *Xoo* sequences is indicated by the vertical red line. Abbreviations are in
706 red for *raxX-raxSTAB* cluster-negative strains and blue for *raxX-raxSTAB* cluster-positive
707 strains: *S. maltophilia, Sm; X. albilineans, Xa; X. arboricola pv. juglandis, Xaj; X. axonopodis*
708 *pv. manihotis, Xam; X. campestris pv. campestris, Xcc; X. campestris pv. musacearum, Xcm; X.*
709 *cannabis, Xc; X. citri subsp. citri, Xac; X. euvesicatoria, Xe; X. fragariae, Xf; X. hyacinthi, Xh;*

710 *X. maliensis, Xm; X. oryzae pv. oryzae,Xoo; X. sacchari,; Xs X. translucens, Xt; X. vesicatoria,*
711 *Xv.*

**Fig. S3. GcvP length polymorphisms in different *Xanthomonas* lineages.**

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

**Fig. S4. Phylogenetic tree for *raxST* homologs.**

720 Distribution of *raxST* homologs across bacterial genera, including the major groups of
721 proteobacteria as well as cyanobacteria. The tree shown was constructed by neighbor-joining
722 with 1000 bootstrap replicates; branches with < 50% bootstrap support are not drawn. The *raxST*
723 sequence from *Xoo* strain PXO99$^{A}$ was used as query for tBLASTn.

**Fig. S5. *raxX* expression in *Xoo* PXO99$^{A}$ complemented strains.**

724 Data show that *raxX* gene expression in the complemented strains with different *raxX* alleles
725 with its promoter region on plasmids. The expression is shown as the logarithm of raw data using
726 qRT-PCR. Gene expression was normalized to the chromosomal gene PXO_01660 (annotated as

727 an *ampC* gene homolog encoding ß-lactamase). Data are the mean values from two biological
728 replicates. Error bars show the standard deviation.

**Fig. S6. RaxST sequence polymorphisms in *Xoo* strain AXO1947.**

729 The RaxST sequence from *Xoo* strain PXO99$^{A}$ is shown. The seven missense substitutions in
730 the sequence from *Xoo* strain AXO1947 (Huguet-Tapia et al., 2016) are indicated. The
731 boundaries of the PAPS binding motifs (5'-PSB and 3'-PB; reference (Negishi et al., 2001), are
732 enclosed in boxes. These motifs encompass the catalytic residues Arg-11 and Ser-118.

**Fig. S7. *raxX* and *raxST* expression in *Xoo* PXO99$^{A}$ complemented strains.**

733 Data show *raxX* and *raxST* gene expression in the complemented strains (with *raxX* and *raxST*
734 on plasmids) relative to expression in *Xoo* strain PXO899$^{A}$ (with *raxX* and *raxST* on the
735 chromosome). Expression was determined by qRT-PCR (see Materials and Methods), and is
736 shown as the logarithm of the fold change. Gene expression was normalized to the chromosomal
737 gene PXO_01660 (annotated as an *ampC* gene homolog encoding □-lactamase). Data are the
738 mean values from two biological replicates. Error bars show the standard deviation.

**Fig. S8. RaxST structural alignment.**

739 Sequence alignment of the human TPST2 and *Xoo* RaxST sequences formatted with ESPript 3.0
740 (Robert & Gouet, 2014). Secondary structure elements derived from the respective structural
741 models are shown. Stars show TPST2 residues involved in PAPS binding, and arrows show
742 RaxST missense substitutions.

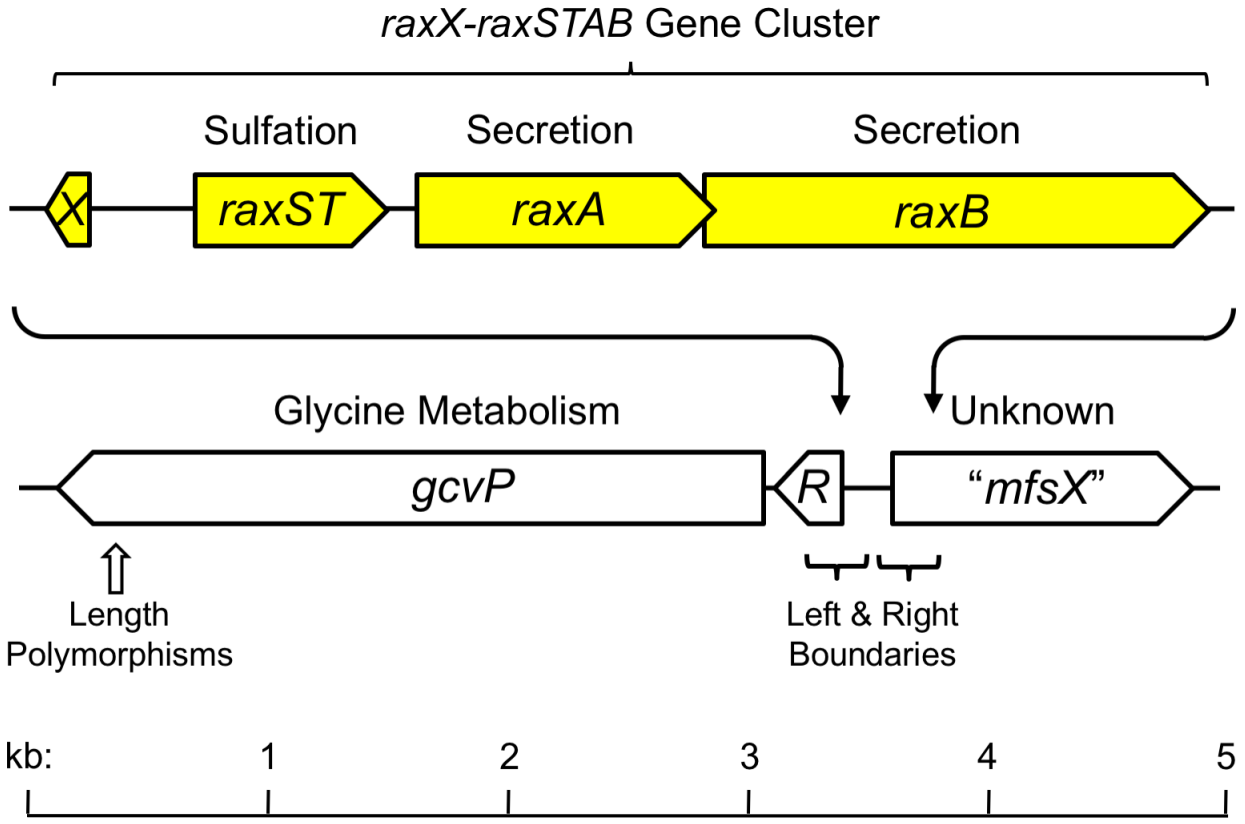
**Fig. S9. Model for RaxST structure.**

743 Predicted RaxST structure shown in cartoon and surface representation, based on the dimeric
744 structure of TPST2. The two RaxST monomers are colored in dark and light green. The 3'-
745 phosphoadenosine-5'-phosphate (PAP) and C4 substrate peptide that where co-crystallized with
746 TPST2 are superimposed on the RaxST model. PAP is represented as labelled and the substrate
747 peptide is shown in yellow-cartoon with the acceptor tyrosine represented as labelled. Residues
748 His-50 and Arg-129 are colored in magenta and highlighted.

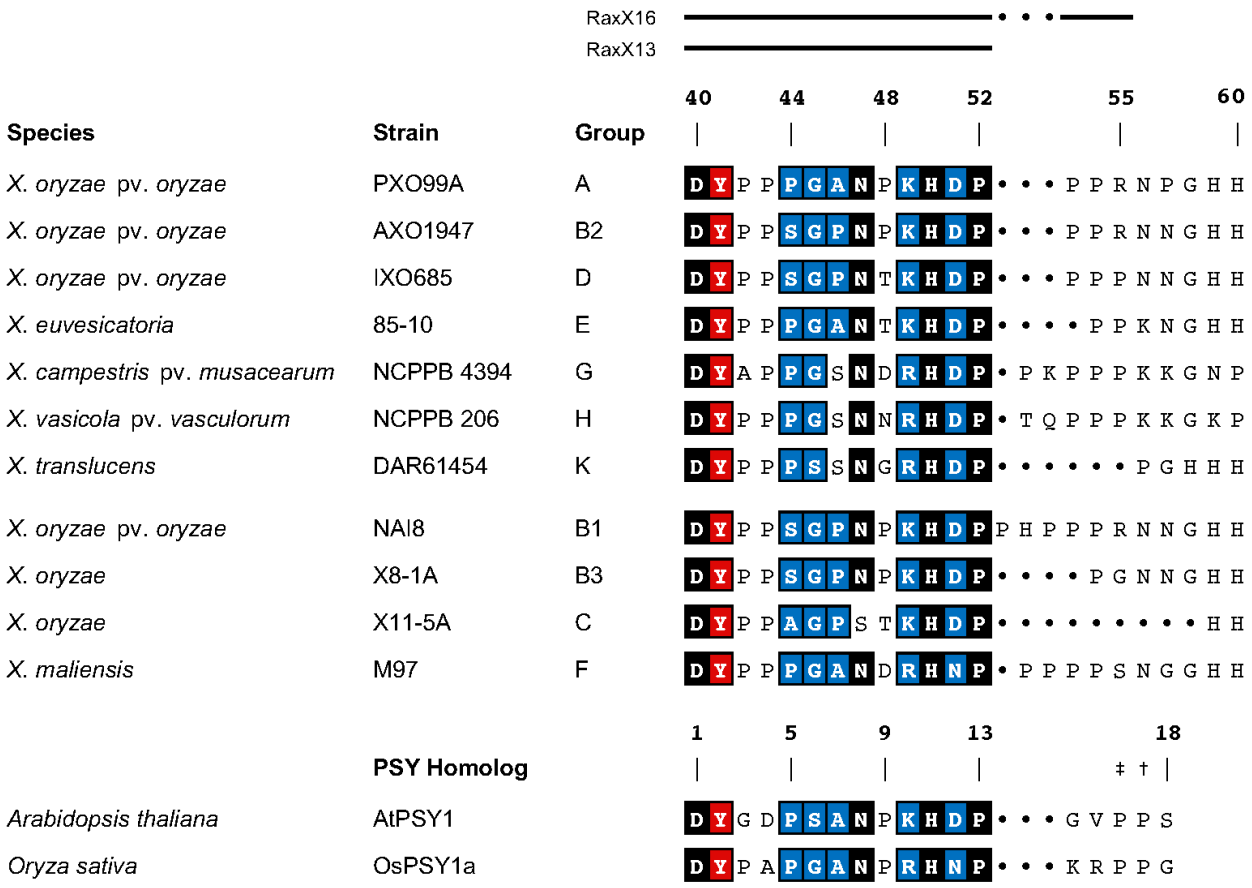
**File. S1. *Xanthomonas* strains analyzed for whole-genome phylogeny.**

749 Excel file (.XLS format).

750 **Fig. 1.**

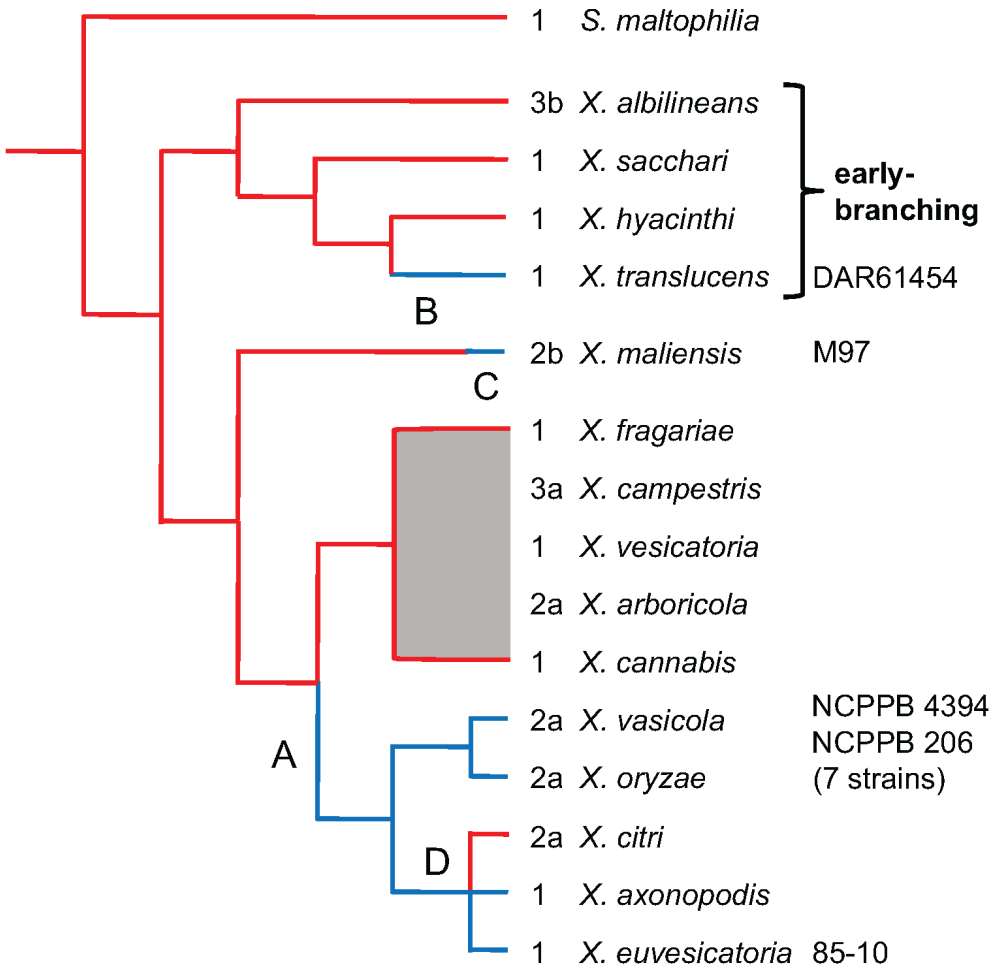

raxX-raxSTAB Gene Cluster
Sulfation
Secretion
Secretion
X
raxST
raxA
raxB
Glycine Metabolism
Unknown
gcvP
R
"mfsX"
Length
Polymorphisms
Left & Right
Boundaries
kb:
1
2
3
4
5

751 **Fig. 2.**

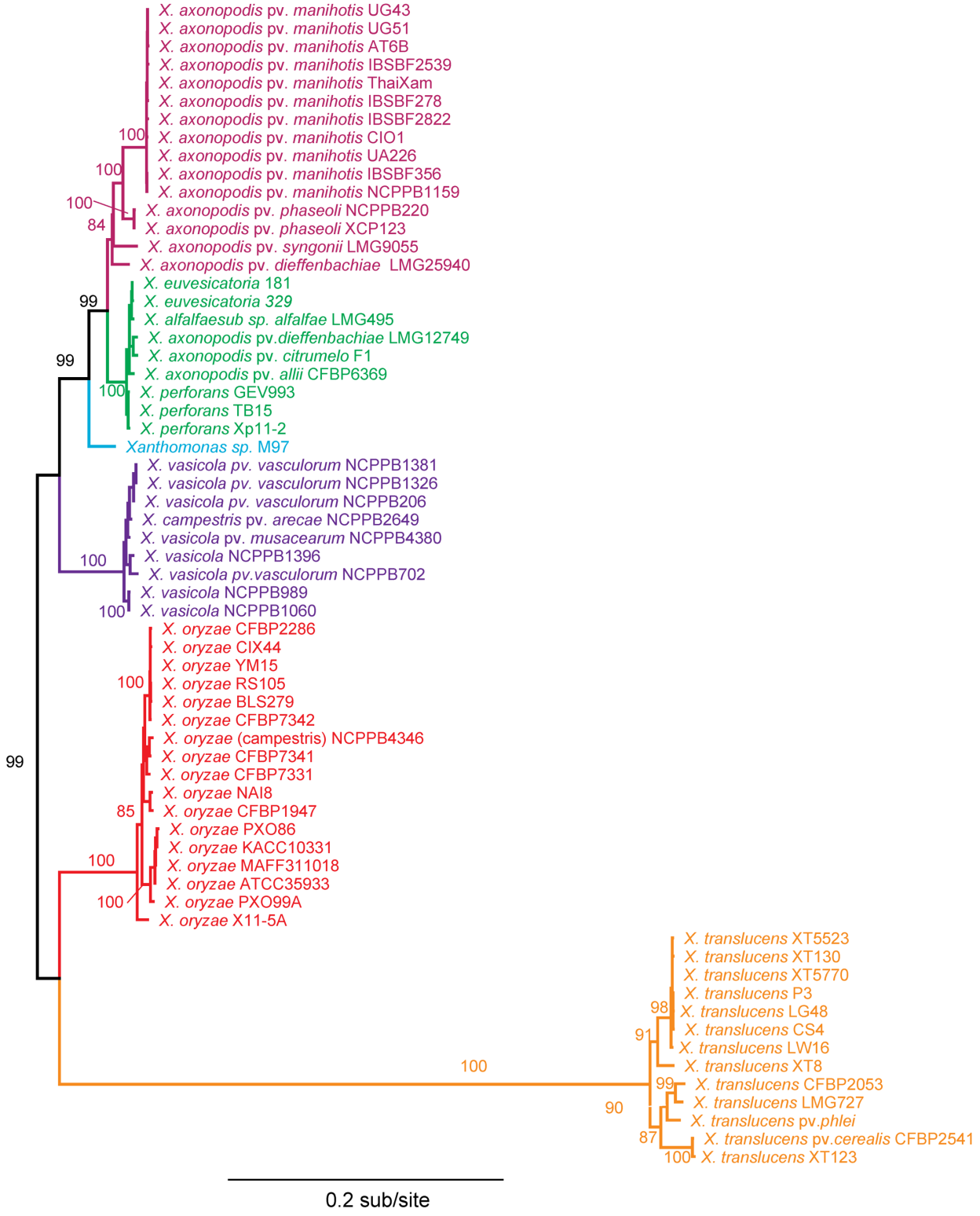

RaxX16
RaxX13
40 44 48 52 55 60
Species Strain Group
X. oryzae pv. oryzae PXO99A A DYPPPGANPKHDP•••PPRNPGHH
X. oryzae pv. oryzae AXO1947 B2 DYPPSGPNPKHDP•••PPRNNGHH
X. oryzae pv. oryzae IXO685 D DYPPSGPNTKHDP•••PPPNNGHH
X. euvesicatoria 85-10 E DYPPPGANTKHDP••••PPKNGHH
X. campestris pv. musacearum NCPPB 4394 G DYAPPGSNDRHDP•PKPPPKKGNP
X. vasicola pv. vasculorum NCPPB 206 H DYPPPGSNNRHDP•TQPPPKKGKP
X. translucens DAR61454 K DYPPPSSNGRHDP••••••PGHHH
X. oryzae pv. oryzae NAI8 B1 DYPPSGPNPKHDPPHPPPRNNGHH
X. oryzae X8-1A B3 DYPPSGPNPKHDP••••PGNNGHH
X. oryzae X11-5A C DYPPAGPSTKHDP•••••••••HH
X. maliensis M97 F DYPPPGANDRHNP•PPPPSNGGHH
1 5 9 13 18
PSY Homolog
Arabidopsis thaliana AtPSY1 DYGDPSANPKHDP•••GVPPS
Oryza sativa OsPSY1a DYPAPGANPRHNP•••KRPPG

752 **Fig. 3.**

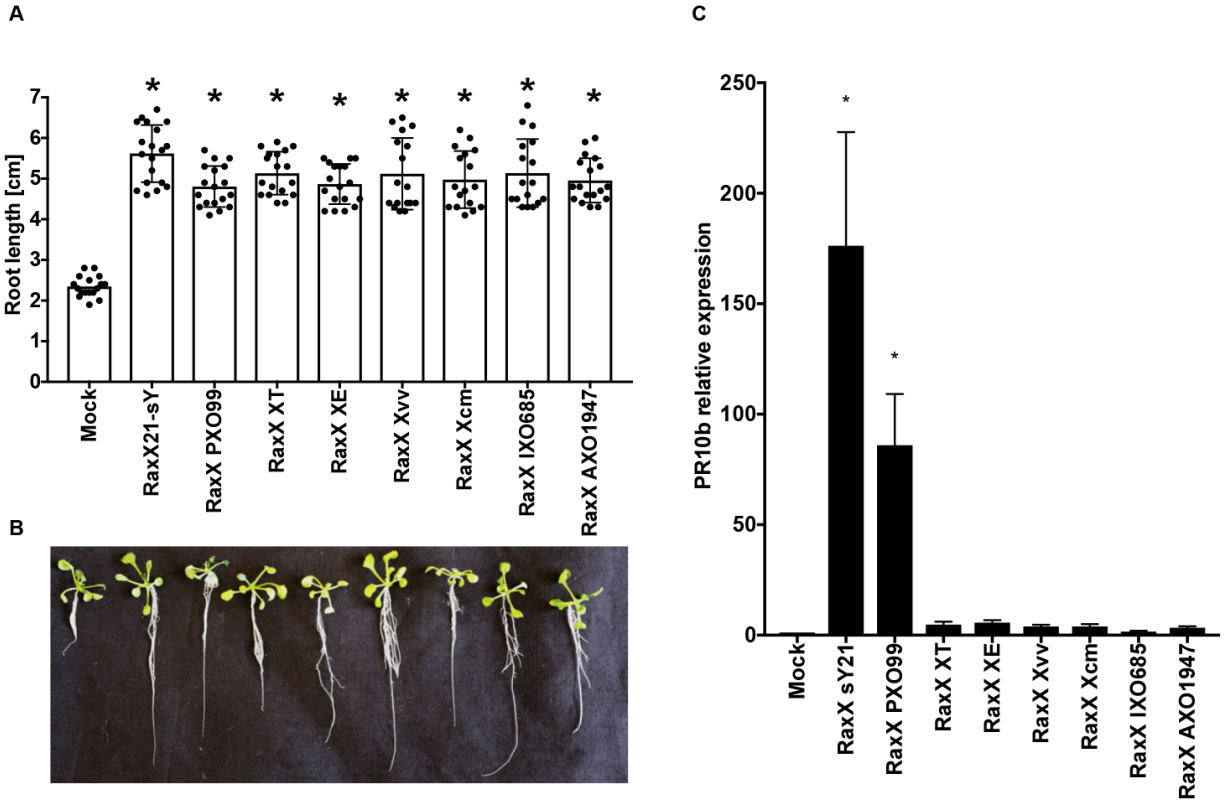
753


1 S. maltophilia
3b X. albilineans
1 X. sacchari
1 X. hyacinthi
1 X. translucens
early-branching
DAR61454
B
2b X. maliensis
M97
C
1 X. fragariae
3a X. campestris
1 X. vesicatoria
2a X. arboricola
1 X. cannabis
2a X. vasicola
NCPPB 4394
NCPPB 206
(7 strains)
A
2a X. oryzae
2a X. citri
D
1 X. axonopodis
1 X. euvesicatoria
85-10

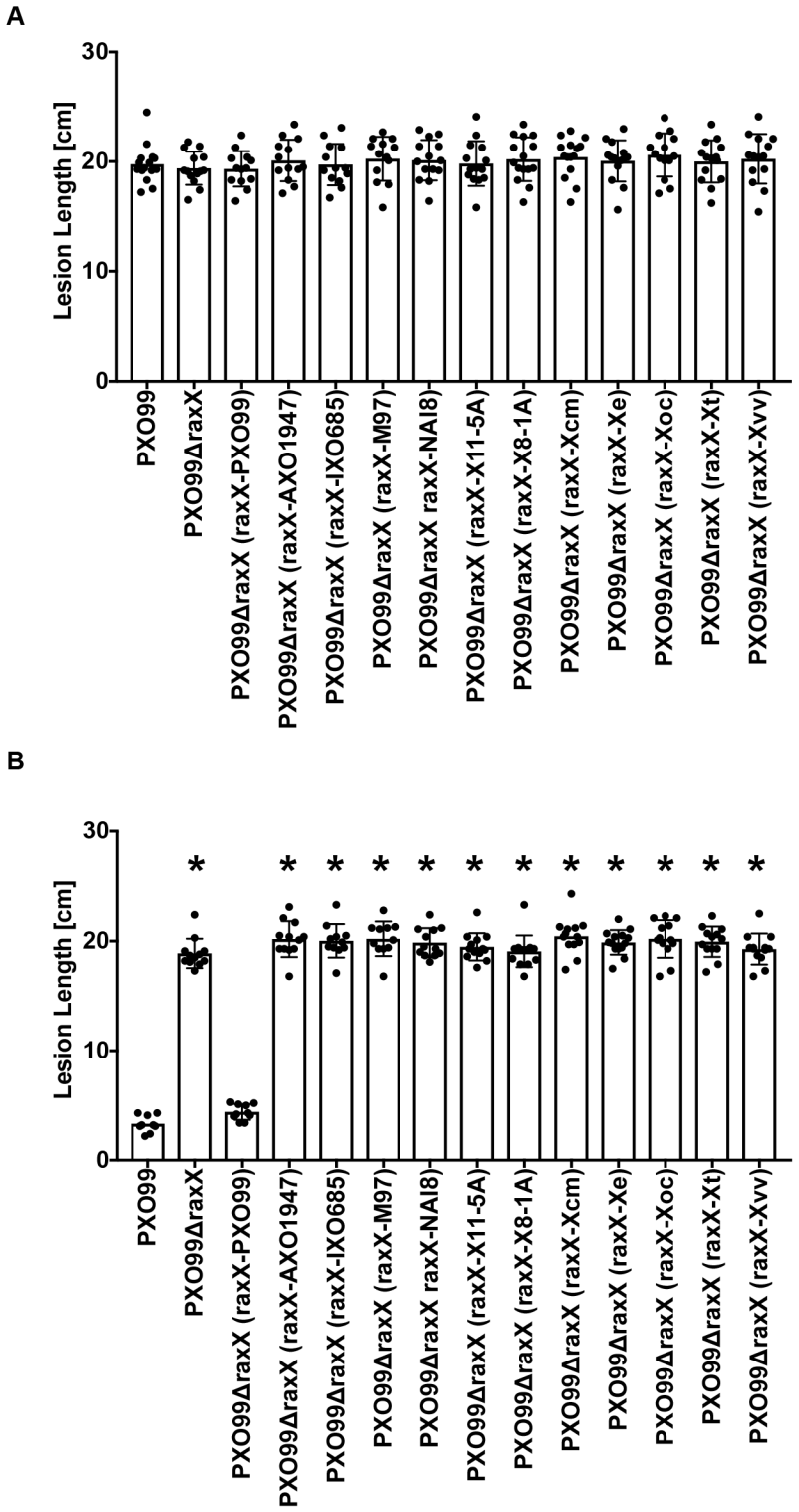
780 **Fig. 4.**


X. axonopodis pv. manihotis UG43
X. axonopodis pv. manihotis UG51
X. axonopodis pv. manihotis AT6B
X. axonopodis pv. manihotis IBSBF2539
X. axonopodis pv. manihotis ThaiXam
X. axonopodis pv. manihotis IBSBF278
X. axonopodis pv. manihotis IBSBF2822
X. axonopodis pv. manihotis CIO1
X. axonopodis pv. manihotis UA226
X. axonopodis pv. manihotis IBSBF356
X. axonopodis pv. manihotis NCPPB1159
X. axonopodis pv. phaseoli NCPPB220
X. axonopodis pv. phaseoli XCP123
X. axonopodis pv. syngonii LMG9055
X. axonopodis pv. dieffenbachiae LMG25940
X. euvesicatoria 181
X. euvesicatoria 329
X. alfalfaesub sp. alfalfae LMG495
X. axonopodis pv.dieffenbachiae LMG12749
X. axonopodis pv. citrumelo F1
X. axonopodis pv. allii CFBP6369
X. perforans GEV993
X. perforans TB15
X. perforans Xp11-2
Xanthomonas sp. M97
X. vasicola pv. vasculorum NCPPB1381
X. vasicola pv. vasculorum NCPPB1326
X. vasicola pv. vasculorum NCPPB206
X. campestris pv. arecae NCPPB2649
X. vasicola pv. musacearum NCPPB4380
X. vasicola NCPPB1396
X. vasicola pv.vasculorum NCPPB702
X. vasicola NCPPB989
X. vasicola NCPPB1060
X. oryzae CFBP2286
X. oryzae CIX44
X. oryzae YM15
X. oryzae RS105
X. oryzae BLS279
X. oryzae CFBP7342
X. oryzae (campestris) NCPPB4346
X. oryzae CFBP7341
X. oryzae CFBP7331
X. oryzae NAI8
X. oryzae CFBP1947
X. oryzae PXO86
X. oryzae KACC10331
X. oryzae MAFF311018
X. oryzae ATCC35933
X. oryzae PXO99A
X. oryzae X11-5A
X. translucens XT5523
X. translucens XT130
X. translucens XT5770
X. translucens P3
X. translucens LG48
X. translucens CS4
X. translucens LW16
X. translucens XT8
X. translucens CFBP2053
X. translucens LMG727
X. translucens pv.phlei
X. translucens pv.cerealis CFBP2541
X. translucens XT123
0.2 sub/site

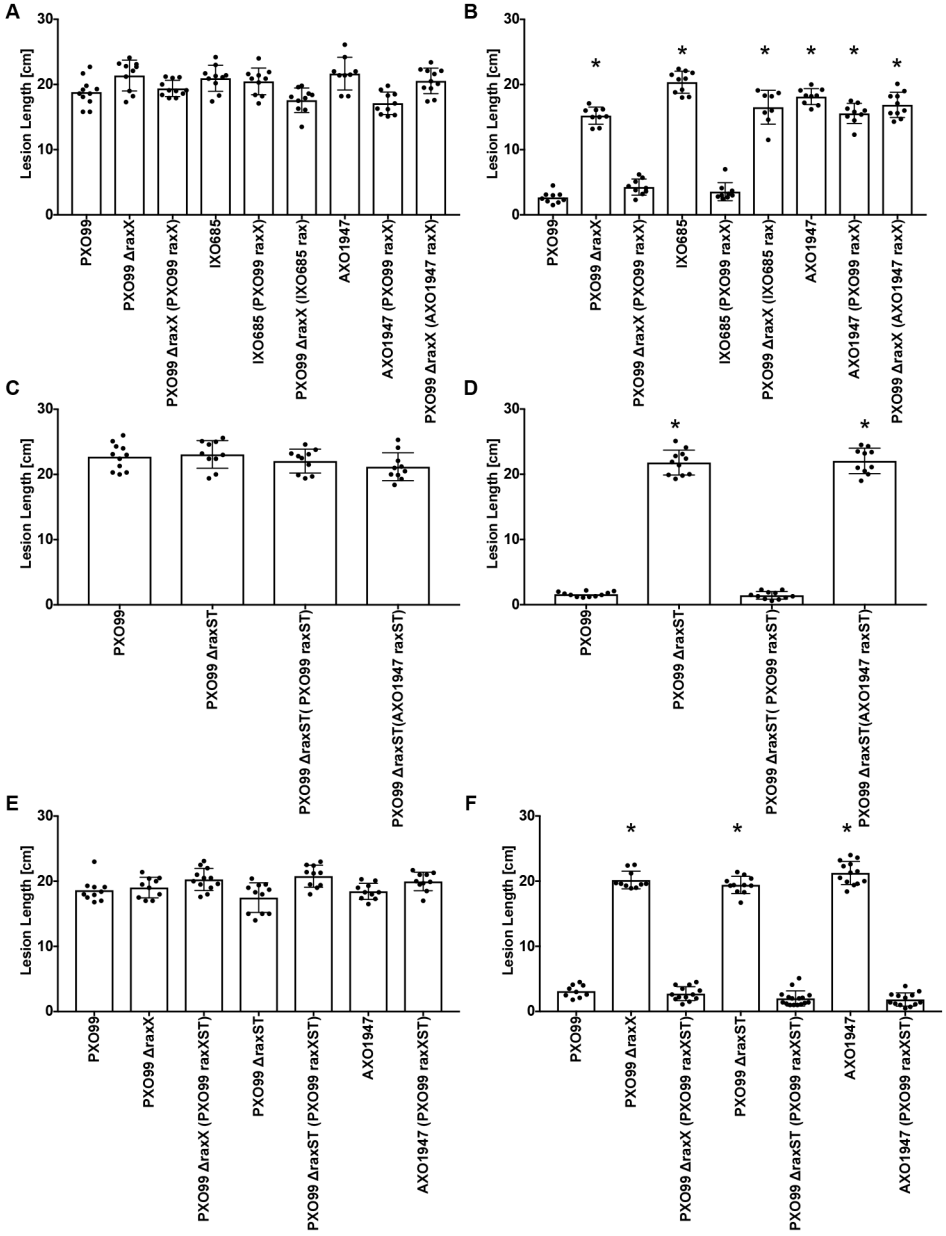
781 **Fig. 5.**


A
Root length [cm]
0
1
2
3
4
5
6
7
Mock
RaxX21-sY
RaxX PXO99
RaxX XT
RaxX XE
RaxX Xvv
RaxX Xcm
RaxX IXO685
RaxX AXO1947
B
C
PR10b relative expression
0
50
100
150
200
250
Mock
RaxX sY21
RaxX PXO99
RaxX XT
RaxX XE
RaxX Xvv
RaxX Xcm
RaxX IXO685
RaxX AXO1947

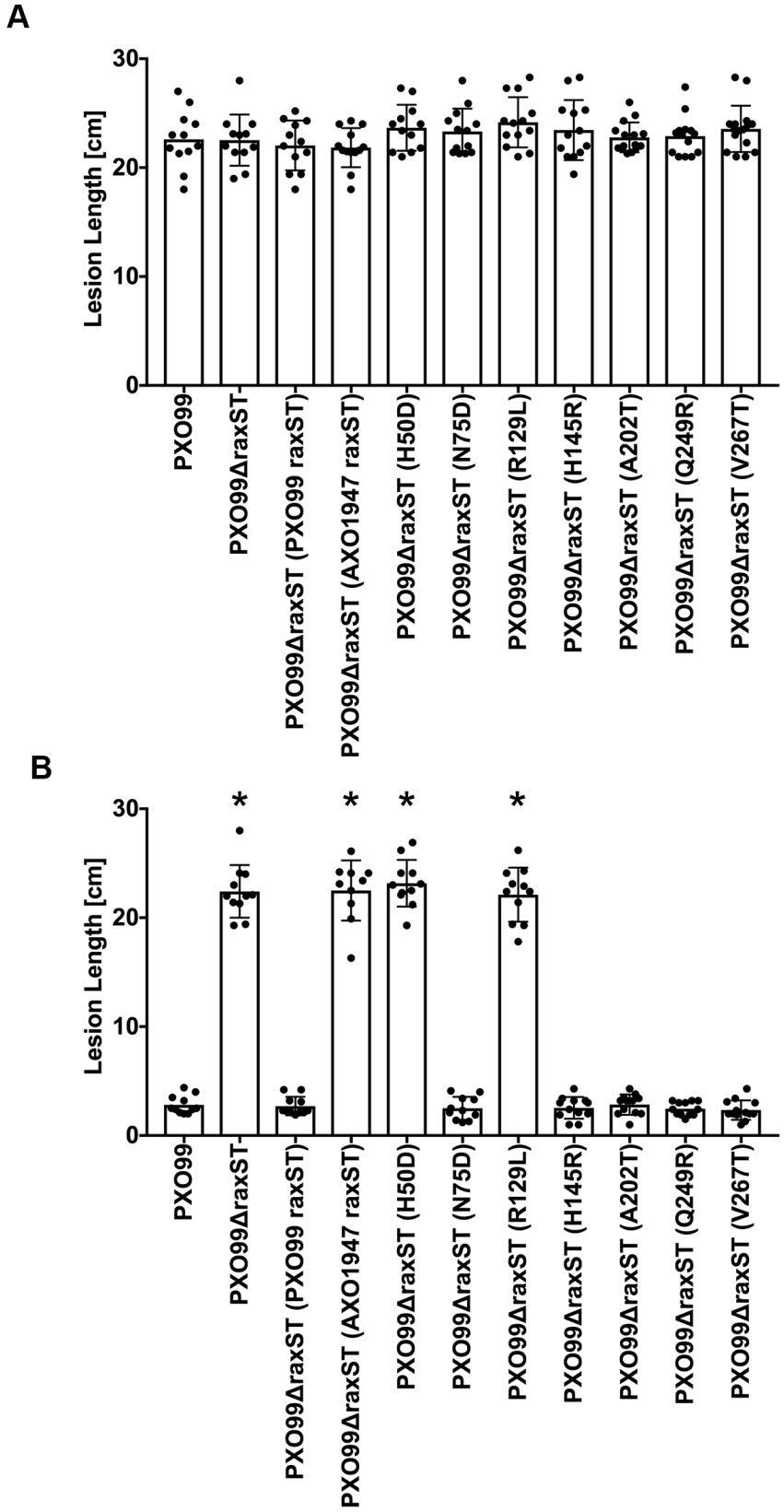
782 **Fig. 6.**


A
Lesion Length [cm]
30
20
10
0
PXO99
PXO99ΔraxX
PXO99ΔraxX (raxX-PXO99)
PXO99ΔraxX (raxX-AXO1947)
PXO99ΔraxX (raxX-IXO685)
PXO99ΔraxX (raxX-M97)
PXO99ΔraxX raxX-NAI8)
PXO99ΔraxX (raxX-X11-5A)
PXO99ΔraxX (raxX-X8-1A)
PXO99ΔraxX (raxX-Xcm)
PXO99ΔraxX (raxX-Xe)
PXO99ΔraxX (raxX-Xoc)
PXO99ΔraxX (raxX-Xt)
PXO99ΔraxX (raxX-Xvv)
B
Lesion Length [cm]
30
20
10
0
*
PXO99
PXO99ΔraxX
PXO99ΔraxX (raxX-PXO99)
PXO99ΔraxX (raxX-AXO1947)
PXO99ΔraxX (raxX-IXO685)
PXO99ΔraxX (raxX-M97)
PXO99ΔraxX raxX-NAI8)
PXO99ΔraxX (raxX-X11-5A)
PXO99ΔraxX (raxX-X8-1A)
PXO99ΔraxX (raxX-Xcm)
PXO99ΔraxX (raxX-Xe)
PXO99ΔraxX (raxX-Xoc)
PXO99ΔraxX (raxX-Xt)
PXO99ΔraxX (raxX-Xvv)

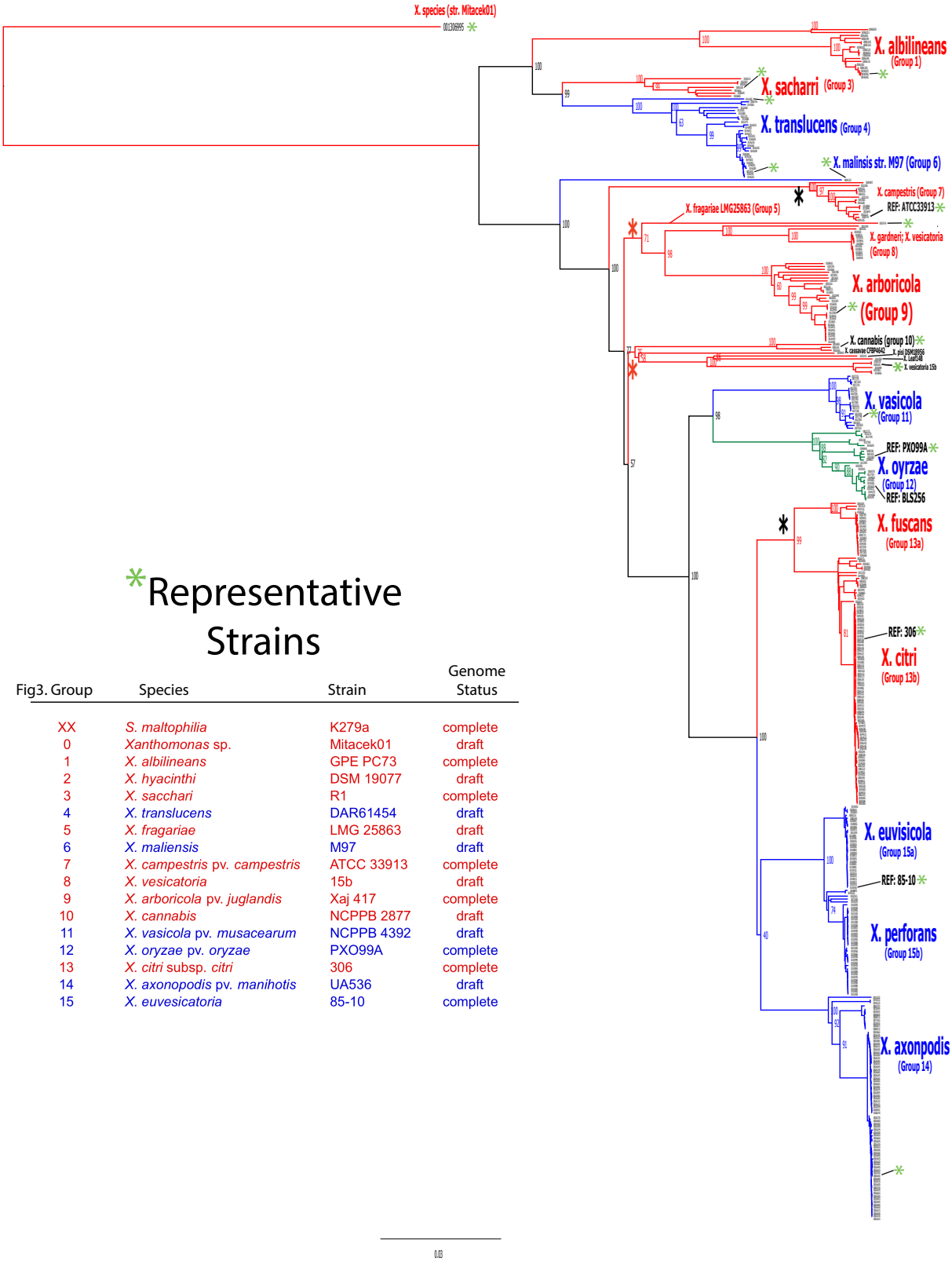
783 **Fig. 7.**


A
B
C
D
E
F
Lesion Length [cm]
0
10
20
30
PXO99
PXO99 ΔraxX
PXO99 ΔraxX (PXO99 raxX)
IXO685
IXO685 (PXO99 raxX)
PXO99 ΔraxX (IXO685 rax)
AXO1947
AXO1947 (PXO99 raxX)
PXO99 ΔraxX (AXO1947 raxX)
PXO99 ΔraxST
PXO99 ΔraxST( PXO99 raxST)
PXO99 ΔraxST(AXO1947 raxST)
PXO99 ΔraxX (PXO99 raxXST)
PXO99 ΔraxST
PXO99 ΔraxST (PXO99 raxXST)
AXO1947 (PXO99 raxXST)
*

801 **Fig. 8.**


A
Lesion Length [cm]
30
20
10
0
PXO99
PXO99ΔraxST
PXO99ΔraxST (PXO99 raxST)
PXO99ΔraxST (AXO1947 raxST)
PXO99ΔraxST (H50D)
PXO99ΔraxST (N75D)
PXO99ΔraxST (R129L)
PXO99ΔraxST (H145R)
PXO99ΔraxST (A202T)
PXO99ΔraxST (Q249R)
PXO99ΔraxST (V267T)
B
Lesion Length [cm]
30
20
10
0
*
*
*
*
PXO99
PXO99ΔraxST
PXO99ΔraxST (PXO99 raxST)
PXO99ΔraxST (AXO1947 raxST)
PXO99ΔraxST (H50D)
PXO99ΔraxST (N75D)
PXO99ΔraxST (R129L)
PXO99ΔraxST (H145R)
PXO99ΔraxST (A202T)
PXO99ΔraxST (Q249R)
PXO99ΔraxST (V267T)

820 **Fig. S1.**




*Representative Strains

| Fig3. Group | Species | Strain | Genome Status |
|---|---|---|---|
| XX | *S. maltophilia* | K279a | complete |
| 0 | *Xanthomonas* sp. | Mitacek01 | draft |
| 1 | *X. albilineans* | GPE PC73 | complete |
| 2 | *X. hyacinthi* | DSM 19077 | draft |
| 3 | *X. sacchari* | R1 | complete |
| 4 | *X. translucens* | DAR61454 | draft |
| 5 | *X. fragariae* | LMG 25863 | draft |
| 6 | *X. maliensis* | M97 | draft |
| 7 | *X. campestris* pv. *campestris* | ATCC 33913 | complete |
| 8 | *X. vesicatoria* | 15b | draft |
| 9 | *X. arboricola* pv. *juglandis* | Xaj 417 | complete |
| 10 | *X. cannabis* | NCPPB 2877 | draft |
| 11 | *X. vasicola* pv. *musacearum* | NCPPB 4392 | draft |
| 12 | *X. oryzae* pv. *oryzae* | PXO99A | complete |
| 13 | *X. citri* subsp. *citri* | 306 | complete |
| 14 | *X. axonopodis* pv. *manihotis* | UA536 | draft |
| 15 | *X. euvesicatoria* | 85-10 | complete |

**Fig. S2.**

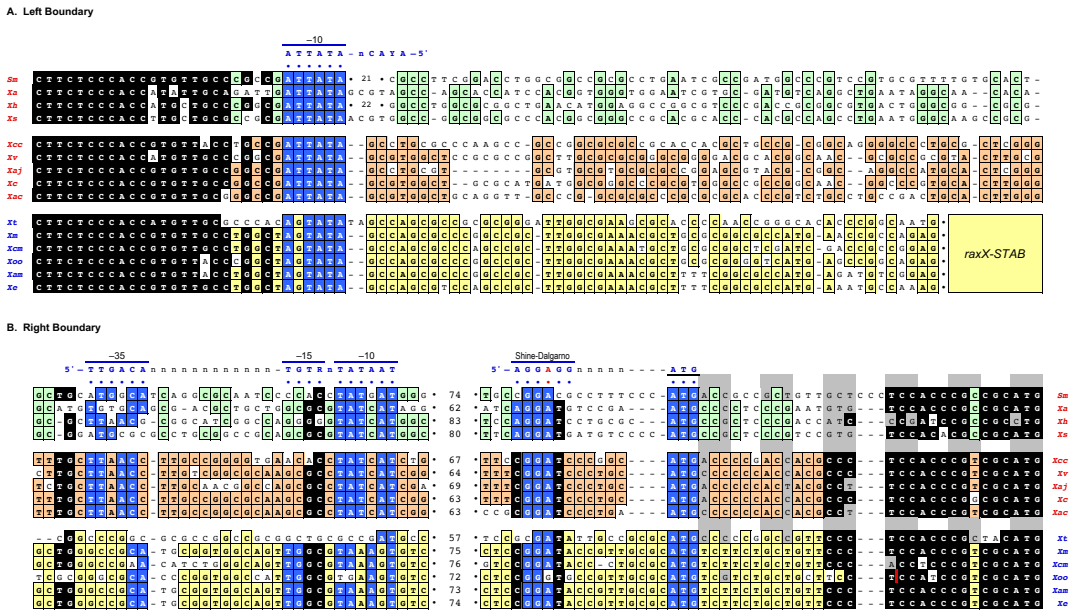
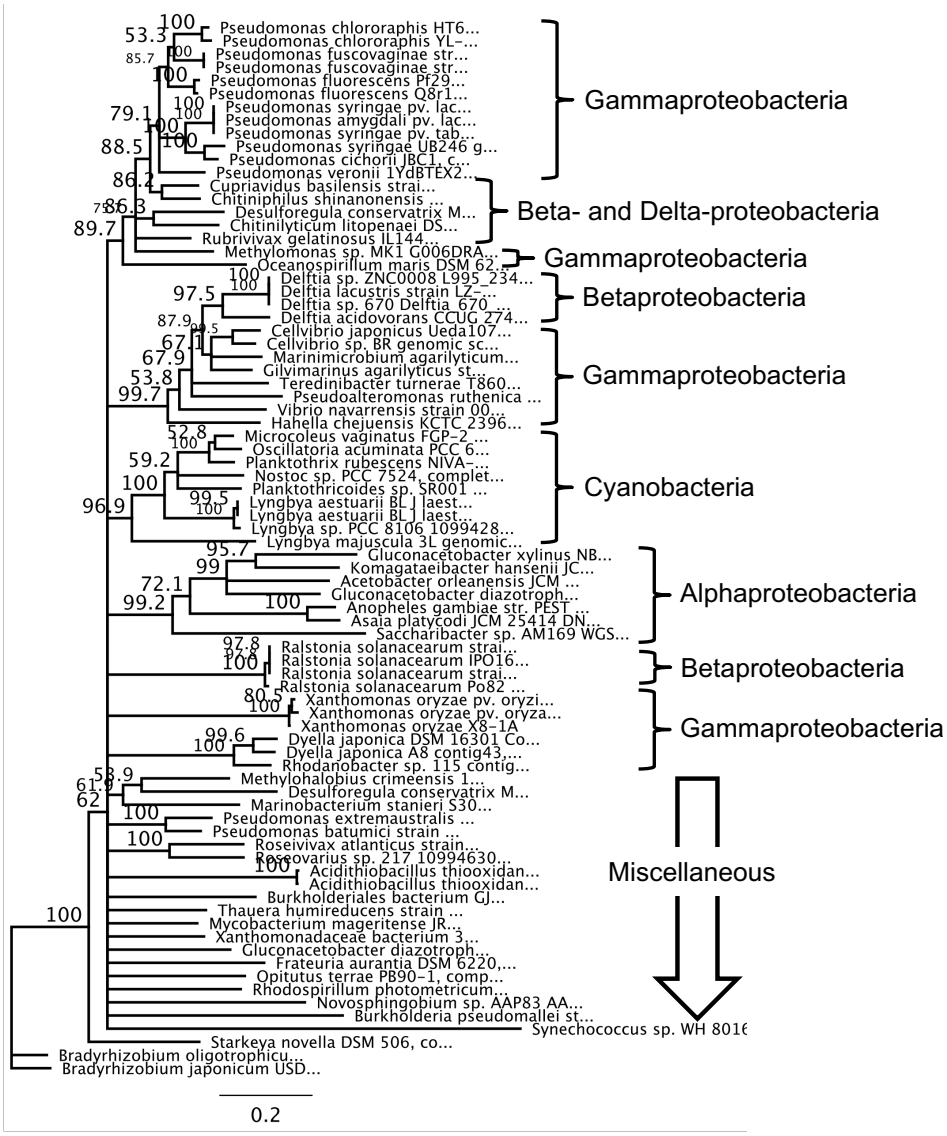

A. Left Boundary
−10
ATTATA-nCAYA-5'
raxX-STAB
B. Right Boundary
−35
−15
−10
5'-TTGACA
TGTRnTATAAT
Shine-Dalgarno
5'-AGGAGG
ATG

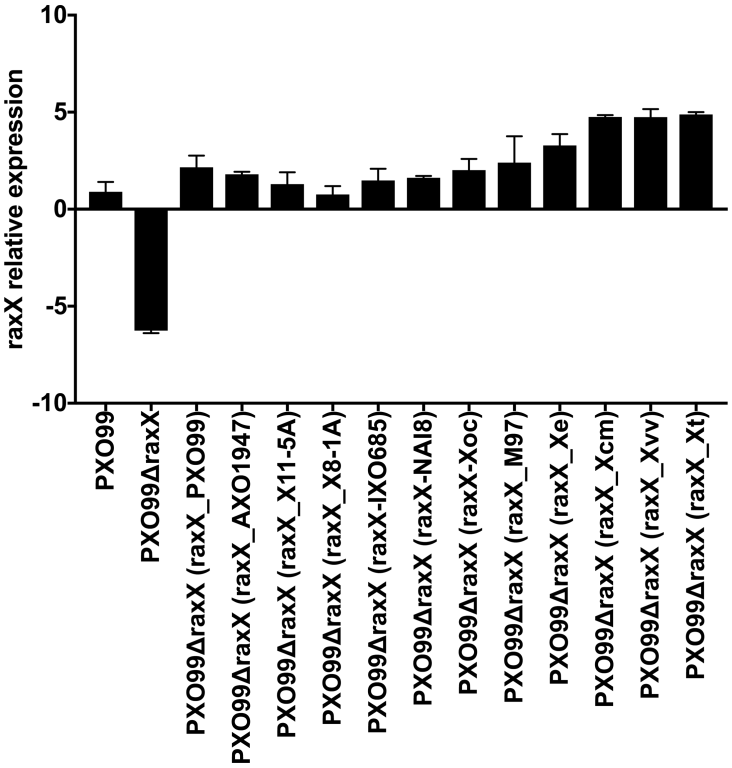
**Fig. S3.**

```
Xaj   2a   GVGPCAVKSHLAPYLPRAGI----HAGEGQTAAIHGGGLNSESGSGHSSRIGGMVSAAAYGSASILPISWM
Xcm   2a   GVGPCAVKSHLAPYLPRAGI----HAGEGQDVAAHGGGLNSESGAAGSLRTGGMVSAAAYGSASILPISWM
Xoo   2a   GVGPCAVKSHLAPFLPRAGL----HAGEGQTAAIHGGGFNSGSGSGHSSRIGGMVSAAAYGSASILPISWM
Xac   2a   GVGPCAVKSHLAPYLPRAGI----HAGEGQTAAIHGGGFNSESGNGHSSRIGGMVSAAAYGSASILPISWM
Xm    2b   GVGPCAVKSHLAPYLPRAGI--------------HGGGFNSESGSGHSSRIGGMVSAAAYGSASILPISWM
Xoc   2b   GVGPCAVKSHLAPFLPRAGL--------------HAGGFNSESGSGHSSRIGGMVSAAAYGSASILPISWM
Xcc   3a   GVGPCAVKSHLAPFLPKTLPNAGIRAGENQKAAIHGSGSNF--GEGE----VGMVSAASYGSASILPISWM
Xa    3b   GVGPCAVKAHLAPYLPMTLPN----AGEAQKAA----------GEGV----VGMVSAASFGSASILPISWM
Sm    1    GVGPCAVKEHLAPFLPGKLG-----------------------DNGP----VGMVSAASFGSASILPISWM
Xh    1    GVGPCAVKSHLAPYLPKTLG-----------------------GEGD----VGMVSAASFGSASILPISWM
Xs    1    GVGPCAVKAHLAPYLPKTLG-----------------------GDGE----VGMVSAASFGSASILPISWM
Xt    1    GVGPCAVKSHLAPYLPKTLG-----------------------GEGD----VGMVSAASFGSASILPISWM
Xf    1    GVGPCAVKSHLAPFLPRTLG-----------------------SEGD----VGMVSAASYGSASILPISWM
Xv    1    GVGPCAVKSHLAPFLPKTLG-----------------------GEGD----VGMVSAASYGSASILPISWM
Xc    1    GVGPCAVKSHLAPFLPRTLG-----------------------GEGD----VGMVSAASYGSASILPISWM
Xam   1    GVGPCAVKSHLAPFLPRTLG-----------------------GEGD----VGMVSAASYGSASILPISWM
Xe    1    GVGPCAVKSHLAPYLPKTLG-----------------------GEGD----VGMVSAASYGSASILPISWM
                                 ↑                                ↑
                                Gly-733                          Val-738
```

823 **Fig. S4.**


Pseudomonas chlororaphis HT6...
Pseudomonas chlororaphis YL-...
Pseudomonas fuscovaginae str...
Pseudomonas fuscovaginae str...
Pseudomonas fluorescens Pf29...
Pseudomonas fluorescens Q8r1...
Pseudomonas syringae pv. lac...
Pseudomonas amygdali pv. lac...
Pseudomonas syringae pv. tab...
Pseudomonas syringae UB246 g...
Pseudomonas cichorii JBC1, c...
Pseudomonas veronii 1YdBTEX2...
Cupriavidus basilensis strai...
Chitiniphilus shinanonensis ...
Desulforegula conservatrix M...
Chitinilyticum litopenaei DS...
Rubrivivax gelatinosus IL144...
Methylomonas sp. MK1 G006DRA...
Oceanospirillum maris DSM 62...
Delftia sp. ZNC0008 L995_234...
Delftia lacustris strain LZ-...
Delftia sp. 670 Delftia_670 ...
Delftia acidovorans CCUG 274...
Cellvibrio japonicus Ueda107...
Cellvibrio sp. BR genomic sc...
Marinimicrobium agarilyticum...
Gilvimarinus agarilyticus st...
Teredinibacter turnerae T860...
Pseudoalteromonas ruthenica ...
Vibrio navarrensis strain 00...
Hahella chejuensis KCTC 2396...
Microcoleus vaginatus FGP-2 ...
Oscillatoria acuminata PCC 6...
Planktothrix rubescens NIVA-...
Nostoc sp. PCC 7524, complet...
Planktothricoides sp. SR001 ...
Lyngbya aestuarii BL J laest...
Lyngbya aestuarii BL J laest...
Lyngbya sp. PCC 8106 1099428...
Lyngbya majuscula 3L genomic...
Gluconacetobacter xylinus NB...
Komagataeibacter hansenii JC...
Acetobacter orleanensis JCM ...
Gluconacetobacter diazotroph...
Anopheles gambiae str. PEST ...
Asaia platycodi JCM 25414 DN...
Saccharibacter sp. AM169 WGS...
Ralstonia solanacearum strai...
Ralstonia solanacearum IPO16...
Ralstonia solanacearum strai...
Ralstonia solanacearum Po82 ...
Xanthomonas oryzae pv. oryzi...
Xanthomonas oryzae pv. oryza...
Xanthomonas oryzae X8-1A
Dyella japonica DSM 16301 Co...
Dyella japonica A8 contig43,...
Rhodanobacter sp. 115 contig...
Methylohalobius crimeensis 1...
Desulforegula conservatrix M...
Marinobacterium stanieri S30...
Pseudomonas extremaustralis ...
Pseudomonas batumici strain ...
Roseivivax atlanticus strain...
Roseovarius sp. 217 1094630...
Acidithiobacillus thiooxidan...
Acidithiobacillus thiooxidan...
Burkholderiales bacterium GJ...
Thauera humireducens strain ...
Mycobacterium mageritense JR...
Xanthomonadaceae bacterium 3...
Gluconacetobacter diazotroph...
Frateuria aurantia DSM 6220,...
Opitutus terrae PB90-1, comp...
Rhodospirillum photometricum...
Novosphingobium sp. AAP83 AA...
Burkholderia pseudomallei st...
Synechococcus sp. WH 8016
Starkeya novella DSM 506, co...
Bradyrhizobium oligotrophicu...
Bradyrhizobium japonicum USD...
Gammaproteobacteria
Beta- and Delta-proteobacteria
Gammaproteobacteria
Betaproteobacteria
Gammaproteobacteria
Cyanobacteria
Alphaproteobacteria
Betaproteobacteria
Gammaproteobacteria
Miscellaneous
0.2

824 **Fig. S5.**


raxX relative expression
10
5
0
-5
-10
PXO99
PXO99ΔraxX
PXO99ΔraxX (raxX_PXO99)
PXO99ΔraxX (raxX_AXO1947)
PXO99ΔraxX (raxX_X11-5A)
PXO99ΔraxX (raxX_X8-1A)
PXO99ΔraxX (raxX-IXO685)
PXO99ΔraxX (raxX-NAI8)
PXO99ΔraxX (raxX-Xoc)
PXO99ΔraxX (raxX_M97)
PXO99ΔraxX (raxX_Xe)
PXO99ΔraxX (raxX_Xcm)
PXO99ΔraxX (raxX_Xvv)
PXO99ΔraxX (raxX_Xt)

838 **Fig. S6.**

```
          11                                        50
           •                                         •
VDYHFISGLPRAGSSLLAALLRQNPQLHADVTSPVARLYAAMLMGMSEEHPSNVQIDDAQ  60
         5'-PSB                                     D

              75                                          118
               •                                           •
RVRLLRAVFDAYYQNRQELGTVFDTNRAWCSRLTGLARLFPRSRMICCVRDVGWIVDSFE  120
              D                                      3'-PB

       129             145
        •               •
RLAQSQPLRLSALFGYDPEDSVSMHADLLTAPRGVVGYALDGLRQAFYGDHADRLLLLRY  180
        L               R

                    202
                     •
DTLAQRPAQAMEQVYAFLQLPAFAHDYAGVQAEAERFDAALQMPGLHRVRRGVHYVPRRS  240
                     T

       249               267
        •                 •
VLPPALFDQLQELAFWESAPSHGALLV                                   267
        R                 I
```

839 **Fig. S7.**

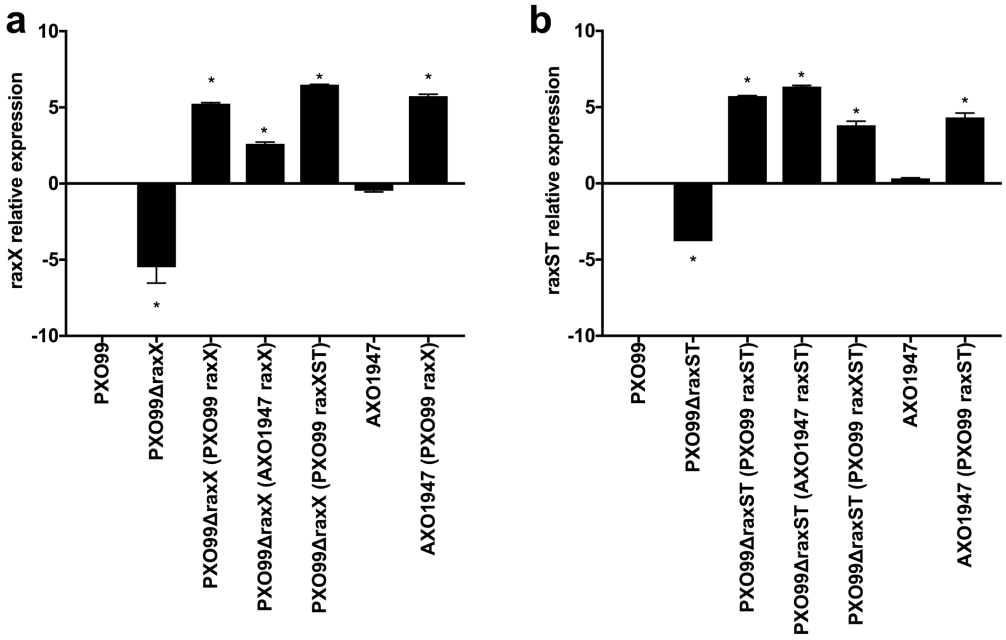
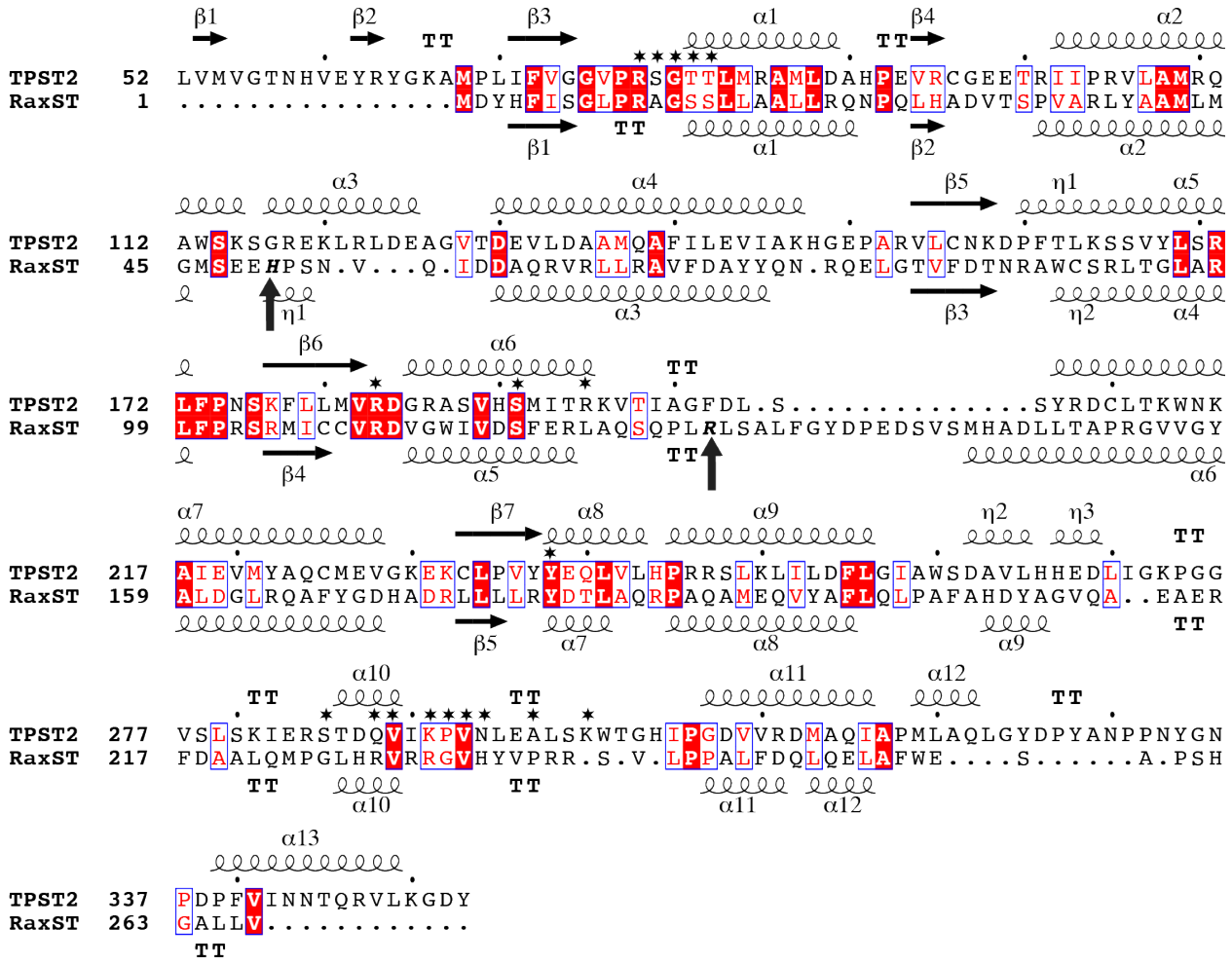
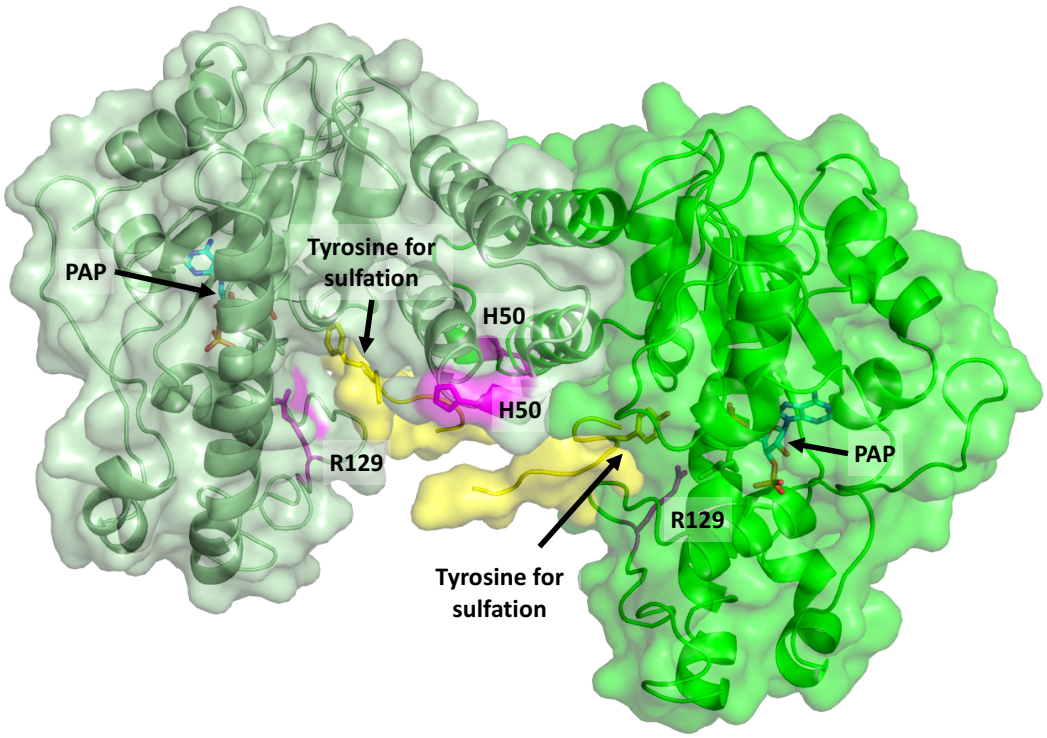

a
raxX relative expression
10
5
0
-5
-10
PXO99
PXO99ΔraxX
PXO99ΔraxX (PXO99 raxX)
PXO99ΔraxX (AXO1947 raxX)
PXO99ΔraxX (PXO99 raxXST)
AXO1947
AXO1947 (PXO99 raxX)
b
raxST relative expression
10
5
0
-5
-10
PXO99
PXO99ΔraxST
PXO99ΔraxST (PXO99 raxST)
PXO99ΔraxST (AXO1947 raxST)
PXO99ΔraxST (PXO99 raxXST)
AXO1947
AXO1947 (PXO99 raxST)

**Fig. S8.**


β1 β2 TT β3 α1 β4 α2
TPST2 52 LVMVGTNHVEYRYGKAMPLIFVGGVPRSGTTLMRAMLDAHPEVRCGEETRIIPRVLAMRQ
RaxST 1 .....MDYHFISGLPRAGSSLLAALLRQNPQLHADVTSPVARLYAAMLM
TT β1 α1 β2 α2
α3 α4 β5 η1 α5
TPST2 112 AWSKSGREKLRLDEAGVTDEVLDAAMQAFILEVIAKHGEPARVLCNKDPFTLKSSVYLSR
RaxST 45 GMSEEHPSN.V...Q.IDDAQRVRLLRAVFDAYYQN.RQELGTVFDTNRAWCSRLTGLAR
η1 α3 β3 η2 α4
β6 α6 TT α6
TPST2 172 LFPNSKFLLMVRDGRASVHSMITRKVTIAGFDL.S.....SYRDCLTKWNK
RaxST 99 LFPRSRMICCVRDVGWIVDSFERLAQSQPLRLSALFGYDPEDSVSMHADLLTAPRGVVGY
β4 α5 TT
α7 β7 α8 α9 η2 η3 TT
TPST2 217 AIEVMYAQCMEVGKEKCLPVYYEQLVLHPRRSLKLILDFLGIAWSDAVLHHEDLIGKPGG
RaxST 159 ALDGLRQAFYGDHADRLLLLRYDTLAQRPAQAMEQVYAFLQLPAFAHDYAGVQA..EAER
β5 α7 α8 α9 TT
TT α10 TT α11 α12 TT
TPST2 277 VSLSKIERSTDQVIKPVNLEALSKWTGHIPGDVVRDMAQIAPMLAQLGYDPYANPPNYGN
RaxST 217 FDAALQMPGLHRVRRGVHYVPRR.S.V.LPPALFDQLQELAFWE....S.....A.PSH
TT α10 TT α11 α12
α13
TPST2 337 PDPFVINNTQRVLKGDY
RaxST 263 GALLV.....
TT

841 **Fig. S9.**


PAP
Tyrosine for sulfation
H50
H50
R129
PAP
R129
Tyrosine for sulfation