Distribution and inheritance of a gene cluster encoding a sulfated tyrosine peptide in *Xanthomonas* spp.

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

Megan McDonald,*[,1] Benjamin Schwessinger,*[,1,2]
Valley Stewart,†[,3] and Pamela C. Ronald†[,2,4]

* These authors contributed equally to this work

The authors declare no conflict of interest. Supported by grants from NIH and NSF.

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

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

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

[4] Joint Bioenergy Institute, Emeryville, California

† corresponding authors:

Valley Stewart
Dpt. Microbiology & Molecular Genetics
University of California, Davis
One Shields Avenue
Davis CA 95616-8665 USA
Email: vjstewart@ucdavis.edu
Office: (530) 754-7994
Fax: (530) 754-8973

Pamela C. Ronald
Department of Plant Pathology
University of California, Davis
One Shields Avenue
Davis CA 95616-8665 USA
Email: pcronald@ucdavis.edu
Office: (530) 752-1654
Fax: (530) 752-5674

ABSTRACT

Tyrosine sulfation is a post-translational modification that influences interaction specificity between certain receptors and their protein ligands in diverse biological processes. For example, rice XA21 receptor-mediated recognition of the sulfated bacterial protein RaxX activates an immune response and triggers resistance to the phytopathogen *Xanthomonas oryzae* pv. *oryzae* (*Xoo*). A five kb *raxX-raxSTAB* gene cluster of *Xoo* encodes RaxX, the RaxST tyrosylprotein sulfotransferasea and the RaxA and RaxB components of a predicted proteolytic maturation and ATP-dependent peptide secretion complex. The complete *raxX-raxSTAB* gene cluster was found only in *Xanthomonas* species, and its distribution is consistent with multiple gain and loss events during *Xanthomonas* speciation. Homologs of the *raxST* gene are present in genome sequences of diverse bacterial species. Together, these results establish a foundation for investigating biological roles for tyrosine sulfation in bacteria.

ABSTRACT = 132 words (maximum 200)

INTRODUCTION

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

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

30 Tyrosylprotein sulfotransferase is confined to the Golgi complex in both plants and
31 animals (Moore 2009). Thus, this post-translational modification is targeted to a subset
32 of cell surface and secreted proteins that influence a variety of eukaryotic physiological
33 processes (Matsubayashi 2014; Stone *et al.,* 2009). For example, tyrosine sulfation of
34 the chemokine receptors CCR5 and CXCR4 is necessary for high-affinity binding not
35 only to chemokines, but also to the HIV-1 surface glycoprotein (Farzan *et al.,* 1999;

Kleist *et al.,* 2016). In plants, sulfated tyrosine peptides influence xylem development, root growth, and/or plant immune signaling (Matsubayashi 2014; Zhou *et al.,* 2017).

RaxST sulfation of RaxX residue Tyr-41 is the only example of tyrosine sulfation reported in bacteria (da Silva *et al.,* 2004; Pruitt *et al.,* 2015). Strikingly, RaxX residues 40-52 share sequence similarity with mature active **p**lant peptide containing **s**ulfated t**y**rosine (PSY) hormones (Amano *et al.,* 2007; Pruitt *et al.,* 2015; Pruitt *et al.,* 2017). Indeed, RaxX, like PSY1, can enhance root growth in diverse plant species (Pruitt *et al.,* 2017). Moreover, RaxX also contributes to *Xoo* virulence in the absence of the XA21 immune receptor (Pruitt *et al.,* 2017). This apparent hormone mimickry by RaxX therefore may serve broad functions in *Xoo* pathogenesis. To further elucidate the biological role of bacterial tyrosine sulfation, we sought to identify the species distribution and possible origin of genes in the *raxX-raxSTAB* gene cluster.

Here we show that the *raxX-raxSTAB* gene cluster is confined to a subset of *Xanthomonas* species. In all cases examined, the *raxX-raxSTAB* gene cluster lies between two core (housekeeping) genes, *gcvP* encoding a subunit of glycine dehydrogenase, and a gene encoding a major facilitator subfamily transporter (*"mfsX"*). Examination of nucleotide sequence conservation across the *raxX-raxSTAB* gene cluster, and at its boundaries with the *gcvP* and *"mfsX"* genes, suggests that the *raxX-raxSTAB* gene cluster was acquired through lateral transfer by *X. translucens*, a pathogen of diverse cereal species (Langlois *et al.,* 2017), and separately by *X. maliensis*, associated with but nonpathogenic for rice (Triplett *et al.,* 2015). Finally, genes homologous to *raxST* are present in bacterial genomes from a wide range of species, raising the possibility that RaxST-catalyzed tyrosine sulfation may occur in other genomic and biological contexts in addition to RaxX.

RESULTS

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

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

Xanthomonas taxonomy has undergone substantial changes over the years (Vauterin *et al.,* 2000; Young 2008); see (Midha and Patil 2014) for a representative example). At one point, many strains were denoted as pathovars of either *X. campestris* or *X. axonopodis*, but today over 20 species are distinguished, many 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 phylogenetic tree in order to organize these sequences by relatedness (**Fig. S1**).

The phylogenetic relationships among *Xanthomonas* spp. was assessed using the entire genome assembly with Andi v0.10 (Haubold *et al.,* 2015; Klotzl and Haubold 2016). We compared our topology with several other *Xanthomonas* phylogenetic trees published previously (Ferreira-Tonin *et al.,* 2012; Gardiner *et al.,* 2014; Hauben *et al.,* 1997; Midha and Patil 2014; Parkinson *et al.,* 2007; Parkinson *et al.,* 2009; Rademaker *et al.,* 2005; Triplett *et al.,* 2015; Young *et al.,* 2008). Most share broad similarity with each other and with the whole-genome tree presented here in defining relationships

between well-sampled species. To facilitate discussion, we represent our phylogenetic tree as a cladogram (**Fig. 2**).

We detected the *raxX-raxSTAB* gene cluster in six lineages that consistently are identified as being distinct from one another (Rademaker *et al.,* 2005; Vauterin *et al.,* 1995) (**Fig. 2**). One lineage includes the two *X. oryzae* pathovars, *oryzae* and *oryzicola*, pathogenic on rice (Niño-Liu *et al.,* 2006). A second lineage includes *X. vasicola*, strains of which are pathogenic on sugarcane, sorghum or maize, together with strains denoted as *X. campestris* pv. *musacearum*, pathogenic on banana (Aritua *et al.,* 2008). The third lineage includes *X. euvesicatoria* and *X. perforans*, pathogenic on pepper and tomato (Potnis *et al.,* 2015), together with strains denoted as *X. alfalfa* subsp. *citrumelonis* (pathogenic on citrus) and *X. dieffenbachiae* (anthuriums) (Rademaker group 9.2; (Barak *et al.,* 2016; Rademaker *et al.,* 2005). The fourth lineage includes strains denoted as *X. axonopodis* pathovars *manihotis* (pathogenic on cassava) and *phaseoli* (bean) (Rademaker group 9.4; (Mhedbi-Hajri *et al.,* 2013; Rademaker *et al.,* 2005).

The fifth lineage includes *X. translucens*, different strains of which are pathogenic on one or more cereal crops such as wheat and barley, and/or non-cereal forage and turfgrass species (Langlois *et al.,* 2017). *X. translucens* is 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 nonpathogenic on rice (Triplett *et al.,* 2011); phylogenetic analyses place this species between the "early-branching" species and the remainder (Triplett *et al.,* 2015).

103 Notably, the *raxX-raxSTAB* gene cluster is absent from the *X. citri* pathovar group,
104 pathogenic on a range of dicots including citrus. This group, which includes certain
105 strains denoted as *X. axonopodis* or *X. campestris* (Bansal *et al.,* 2017), clusters
106 phylogenetically among the *X. oryzae*, *X. euvesicatoria* and *X. axonopodis* pv.
107 *manihotis* groups (Midha and Patil 2014; Rademaker *et al.,* 2005; Vauterin *et al.,* 1995)
108 (**Fig. 2**).

109 Together, these observations suggest that the *raxX-raxSTAB* gene cluster experienced
110 multiple gains and/or losses during *Xanthomonas* speciation.

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

111 From the initial analysis described above, we selected 15 species, representing the
112 phylogenetic range of *Xanthomonas*, for more detailed analyses of *rax* gene cluster
113 composition, organization, and inheritance (**Fig. 2; Table 1**). The corresponding
114 genome sequences are accompanied by published descriptions (**Table 1**). The close
115 relative *Stenotrophomonas maltophilia*, which does not contain the *raxX-raxSTAB* gene
116 cluster, serves as a reference (Moore *et al.,* 1997).

117 Both the organization and size of the *raxX-raxSTAB* gene cluster are conserved across
118 all six lineages in which it resides. To address hypotheses for patterns of *raxX-*
119 *raxSTAB* gene cluster inheritance, we compared individual phylogenetic trees for each
120 of the four *rax* genes to the overall *Xanthomona*s phylogenetic tree (**Fig. 3**) (Kuo and
121 Ochman 2009). For all four genes, sequences in *X. translucens*, in the early-branching
122 group, cluster separately from sequences in the other lineages. This finding is
123 congruent with the hypothesis, that *X. translucens* acquired the *raxX-raxSTAB* gene

124 cluster relatively early during *Xanthomonas* speciation. For *X. maliensis*, the *raxX-*
125 *raxSTAB* genes assort among those from *X. euvesicatoria* and the *X. axonopodis*
126 pathovars *manihotis* and *phaseoli* (**Fig. 3**), even though the *X. maliensis* genome
127 sequence itself is more distantly related (**Fig. 2**). This finding suggests that *X. maliensis*
128 acquired the *raxX-raxSTAB* gene cluster relatively late during *Xanthomonas* speciation.

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

136 We further examined phylogenetic relationships by comparing nucleotide sequence
137 identity across the *gcvP, raxX, raxST, raxA, raxB* and *"mfsX"* coding regions, each from
138 the initiation through termination codon (**Table 2**). For comparison, values are
139 presented also for **g**enome-wide **a**verage **n**ucleotide **i**dentity (gANI) as well as the
140 **a**lignment **f**raction (AF), which estimates the fraction of orthologous genes (Varghese *et*
141 *al.,* 2015). For context, a widely-used criterion assigns 95% **a**verage **n**ucleotide **i**dentity
142 (ANI) as the cut-off point for species delineation (Goris *et al.,* 2007). Sequence from *X.*
143 *euvesicatoria* is the reference.

144 The *raxST*, *raxA* and *raxB* coding sequences from *X. axonopodis* pv. *manihotis* and *X.*
145 *maliensis* display the highest identity to those from *X. euvesicatoria*, at least 95% in
146 each case. Sequences from *Xoo* and *X. vasicola* (also known as *X. campestris* pv.
147 *malvacearum*) are about 90% identical, and those from *X. translucens* roughly 75%

148 identical (**Table 2**). The *raxX* coding sequences are more divergent, with identity to the
149 *X. euvesicatoria* sequence ranging from almost 90% for *X. maliensis* and *Xoo* to only
150 63% for *X. translucens* (**Table 2**).

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

151 Comparison of the *gcvP* - *[raxX-raxSTAB]* - "*mfsX*" region from all 16 reference species
152 reveals sharp boundaries flanking the position of the *raxX-raxSTAB* gene cluster. On
153 the left flank, substantial nucleotide identity spans the *gcvP* gene, the *gcvR* riboswitch,
154 and a presumptive promoter –10 element (Mitchell *et al.,* 2003) (**Fig. 1b**). On the right
155 flank, identity begins shortly after the "*mfsX*" initiation codon. Accordingly, upstream
156 signals for "*mfsX*" gene transcription (Mitchell *et al.,* 2003) and translation (Ma *et al.,*
157 2002) are conserved within, but not between, *raxX-raxSTAB*-positive and -negative
158 sequences (**Fig. 1b**).

159 Between these boundaries in *raxX-raxSTAB* gene cluster-negative species, the
160 compact (≤ 200 nt) *gcvP*-"*mfsX*" intergenic sequence is modestly conserved in most
161 genomes (about 60-80% overall identity) (**Fig. 1b**). Much of this identity comes from the
162 "*mfsX*" potential transcription and translation initiation sequences described above. The
163 overall intergenic sequence is less conserved in the early-branching species (*X.*
164 *albilineans*, *X. hyacinthi* and *X. sacchari*), displaying about 50-65% overall identity.

165 In *raxX-raxSTAB* gene cluster-positive genomes, sequence flanking these boundaries
166 appears unrelated to the *gcvP*-"*mfsX*" intergenic sequence from *raxX-raxSTAB* gene
167 cluster-negative genomes (**Fig. 1b**). Rather, it is well-conserved even in the early-

168 branching species, *X. translucens*. These results suggest that anomolous *raxX-*
169 *raxSTAB* gene cluster phylogenetic distribution results from lateral gene transfer.

170 Sequences of the adjacent *gcvP* gene display length polymorphisms (**Fig. 4**) that do not
171 align with overall *Xanthomonas* species relationships (**Fig. 2**). Polymorphisms of this
172 type are unusual, and indicate recombination (Nelson *et al.,* 1997). Their occurrence in
173 a gene adjacent to the *raxX-raxSTAB* gene cluster independently supports the model,
174 that this genomic region evolves through lateral gene transfer.

raxST homologs are present in genomes of diverse bacterial species

175 As we searched genome sequences available through GenBank for evidence of the
176 *raxX-raxSTAB* gene cluster outside of *Xanthomonas* spp., we identified sequences
177 encoding proteins with about 40% identity to, and approximately the same length as, the
178 *Xoo* RaxST protein. Sequence identity is high in residues that form the binding pocket
179 for the cofactor, 3'-phosphoadenosine 5'-phosphosulfate (da Silva *et al.,* 2004; Kakuta
180 *et al.,* 1998), consistent with assignment of these encoded proteins as
181 sulfotransferases. It is not known if these genes encoding tyrosylprotein
182 sulfotransferases, as there are no defined sequence features that distinguish such
183 enzymes from other sulfotransferases that have non-protein substrates (Dong *et al.,*
184 2012; Teramoto *et al.,* 2013).

185 These *raxST* homologs are in a range of bacterial phyla including Proteobacteria and
186 Cyanobacteria (**Fig. 5**). Nevertheless, for most species represented by multiple
187 genome sequences, the *raxST* homolog was detected in a minority of individuals, so it
188 is not part of the core genome in these strains. Moreover, relationships between
189 species in a *raxST* gene phylogenetic tree bear no resemblance to those in the overall

190 tree of bacterial species. For example, in the *raxST* gene tree, sequences from
191 Cyanobacteria are flanked on both sides by sequences from Proteobacteria (**Fig. 5**).
192 Together, these findings provide evidence for lateral transfer of *raxST* homologs
193 transfer (Kuo and Ochman 2009).

DISCUSSION

194 We hypothesize that the *raxX-raxSTAB* gene cluster originated in an ancestor to the
195 lineage containing *X. oryzae*, *X. euvesicatoria*, and related species, with further gains or
196 loss through lateral transfer as described below (**Fig. 2**). Analysis suggests that
197 relatively few events were necessary to form the *raxX-raxSTAB* gene cluster. The
198 *raxAB* genes are homologous to those encoding proteolytic maturation and ATP-
199 dependent peptide secretion complexes (da Silva *et al.,* 2004; Lin *et al.,* 2015), related
200 to type 1 secretion systems but specialized for secreting small peptides such as
201 bacteriocins and peptide pheromones (Holland *et al.,* 2016). Frequently, the gene
202 encoding the secreted substrate is adjacent to genes encoding components of the
203 secretion complex (Dirix *et al.,* 2004). The *raxX* gene therefore might have evolved
204 from the gene for the seccreted peptide substrate of the RaxAB ancestor. Finally, as
205 we show here, homologs for the *raxST* gene are distributed broadly (**Fig. 5**).

206 The *raxX-raxSTAB* gene cluster does not exhibit features, such as a gene for a site-
207 specific recombinase, characteristic of self-mobile genomic islands (Hacker *et al.,*
208 1997). Moreover, variant-length alleles of the adjacent *gcvP* gene (**Fig. 4**) provide
209 evidence for general recombination in the vicinity (Nelson *et al.,* 1997). Thus, the
210 simplest model for *raxX-STAB* gene cluster lateral transfer is that it occurred through
211 general recombination between genes flanking each side of the *raxX-STAB* gene
212 cluster (**Fig. 1b**).

213 Three examples provide further evidence for lateral transfer. First, the *raxX-raxSTAB*
214 gene cluster from the early branching species *X. translucens* has essentially the same
215 size, composition and structure as the others. However, the *X. translucens raxX-*
216 *raxSTAB* sequences are more divergent (**Table 2**). This predicts that the *raxX-raxSTAB*

217 gene cluster has been part of the *X. translucens* genome sufficiently long for sequence
218 alterations to accumulate (Kuo and Ochman 2009).

219 In the second example of evidence for lateral transfer, the *X. maliensis raxSTAB*
220 sequences share strong similarity to those of *X. euvesicatoria* and the *X. axonopodis*
221 pathovars *manihotis* and *phaseoli*, whereas their genome sequences are more
222 divergent (**Table 2**). This suggests that *X. maliensis* acquired the *raxX-raxSTAB* gene
223 cluster relatively recently (Kuo and Ochman 2009).

224 The final example of evidence for lateral transfer considers apparent loss of the *raxX-*
225 *raxSTAB* gene cluster during differentiation of *X. citri* from the large group including
226 *Xoo, X. euvesicatoria* and related species (Midha and Patil 2014; Rademaker *et al.,*
227 2005; Vauterin *et al.,* 1995) (**Fig. 2**). Sequence in the *gcvP*-"*mfsX*" intergenic region is
228 conserved among *raxX-raxSTAB* gene cluster-negative species, including *X. citri* (**Fig.**
229 **1b**). This is consistent with loss from the *X. citri* lineage mediated by recombination,
230 with the *gcvP*-"*mfsX*" region from a *raxX-raxSTAB* gene cluster-negative species. The
231 result of this recombination would be replacement of the *raxX-raxSTAB* gene cluster
232 with a conserved *gcvP*-"*mfsX*" region.

233 In an alternative scenario, where the *X. citri raxX-raxSTAB* gene cluster was lost
234 through deletion, the remaining sequence in the intergenic region would more closely
235 resemble the *raxX-raxSTAB* gene cluster-positive boundary sequence. This conclusion
236 is not supported by the intergenic region found in *X. citri*. A second alternative scenario,
237 that the *raxX-raxSTAB* gene cluster formed after *X. citri* speciation, is not supported by
238 the analysis of *X. translucens* sequences described above.

239 Broad phylogenetic distribution of the *raxX-raxSTAB* gene cluster implies that its
240 associated phenotypes can contribute to host interactions with diverse *Xanthomonas*
241 spp. The *raxX-raxSTAB* gene cluster was identified in the context of the rice XA21-
242 mediated immune response (da Silva *et al.,* 2004; Pruitt *et al.,* 2015), but the sequence
243 and functional similarities between the bacterial RaxX and the plant PSY sulfopeptides
244 suggests that RaxX may mimic PSY phytohormone activies to facilitate infection (Pruitt
245 *et al.,* 2017). Indeed, *Xoo* strains that cannot synthesize sulfated RaxX exhibit reduced
246 virulence (Pruitt *et al.,* 2017).

247 Evidence for lateral transfer to *X. translucens* and *X. maliensis* suggests that *raxX-*
248 *raxSTAB* gene cluster acquisition may contribute to emergence of new species or
249 pathovars. The potentially useful phenotype of PSY hormone mimickry conceivably
250 could introduce a particular strain to previously inaccessible hosts or niches. On the
251 other hand, loss of the *raxX-raxSTAB* gene cluster apparently occurred during formation
252 of the *X. citri* lineage (**Fig. 2**), perhaps indicating that the *raxX-raxSTAB* gene cluster did
253 not enhance fitness in this case.

254 Pathovar phenotypes that differentiate bacterium-plant interactions, characterized
255 extensively in members of the genus *Xanthomonas* (Jacques *et al.,* 2016), are not
256 predicted by the presence or absence of the *raxX-raxSTAB* gene cluster. Some
257 species that infect monocots exclusively contain the *raxX-raxSTAB* gene cluster (e.g.,
258 *X. oryzae*, *X. translucens*), whereas others do not (e.g., *X. arboricola*, *X. hyacinthi*).
259 Likewise, some species that infect dicots exclusively contain the *raxX-raxSTAB* gene
260 cluster (e.g., *X. euvesicatoria*), whereas others do not (e.g., *X. campestris* pv.
261 *campestris*; *X. citri*). Similarly, there is no association with tissue specificity; for
262 example, a single species contains both vascular (*X. oryzae* pv. *oryzae*) and
263 nonvascular (*X. oryzae* pv. *oryzicola*) pathogens.

MATERIALS AND METHODS

Survey of *raxX-STAB* gene clusters in *Xanthomonas* spp.

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

279 Whole genome phylogenies were generated using the entire genome assembly with the
280 program Andi v0.10 (Haubold *et al.,* 2015; Klotzl and Haubold 2016). These distance
281 matrices were plotted as neighbor-joining tree using Phylip v3.695 (Felsenstein 1981).
282 Numbers on the branches represent the proportion (0-100) that the branch appeared in
283 the “bootstrapped” distance matrices using Andi.

Sequence analyses

Nucleotide and deduced amino acid sequences were edited and analyzed with the programs EditSeq™ (version 14.1.0), MegAlign™ (version 14.1.0) and SeqBuilder™ (version 14.1.0), DNASTAR, Madison, WI. The Integrated Microbial Genomes interface (Chen *et al.,* 2017) was used to compare genome segments from different species, and also to extract values for genome-wide Average Nucleotide Identity and genome Alignment Fraction (Varghese *et al.,* 2015).

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGMENTS

Research reported in this publication was supported by the National Institute of General Medical Sciences of the National Institutes of Health under award number R01GM59962, and by the Division of Integrative Organismal Systems of the National Science Foundation under award number IOS-1237975. B.S. is supported by a Discovery Early Career Research Award from the Australian Research Council under award number DE150101897.

REFERENCES

Amano Y, Tsubouchi H, Shinohara H, Ogawa M, Matsubayashi Y (2007). Tyrosine-sulfated glycopeptide involved in cellular proliferation and expansion in *Arabidopsis*. *Proc Natl Acad Sci U S A* **104:** 18333-18338.

Aritua V, Parkinson N, Thwaites R, Heeney JV, Jones DR, Tushemereirwe W *et al* (2008). Characterization of the *Xanthomonas* sp. causing wilt of enset and banana and its proposed reclassification as a strain of *X. vasicola*. *Plant Pathol* **57:** 170-177.

Bansal K, Midha S, Kumar S, Patil PB (2017). Ecological and evolutionary insights into pathovar diversity of Xanthomonas citri. *Appl Environ Microbiol* **83:** e02993-02916.

Barak JD, Vancheva T, Lefeuvre P, Jones JB, Timilsina S, Minsavage GV *et al* (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 EJ, Shybut M, Petriello A *et al* (2012). High-throughput genomic sequencing of cassava bacterial blight strains identifies conserved effectors to target for durable resistance. *Proc Natl Acad Sci U S A* **109:** E1972-1979.

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K *et al* (2009). BLAST+: architecture and applications. *BMC Bioinformatics* **10:** 421.

Chen IA, Markowitz VM, Chu K, Palaniappan K, Szeto E, Pillay M *et al* (2017). IMG/M: integrated genome and metagenome comparative data analysis system. *Nucleic Acids Res* **45:** D507-D516.

Crossman LC, Gould VC, Dow JM, Vernikos GS, Okazaki A, Sebaihia M *et al* (2008). The complete genome, comparative and functional analysis of *Stenotrophomonas maltophilia* reveals an organism heavily shielded by drug resistance determinants. *Genome Biol* **9:** R74.

da Silva AC, Ferro JA, Reinach FC, Farah CS, Furlan LR, Quaggio RB *et al* (2002). Comparison of the genomes of two *Xanthomonas* pathogens with differing host specificities. *Nature* **417:** 459-463.

da Silva FG, Shen Y, Dardick C, Burdman S, Yadav RC, de Leon AL *et al* (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.

Dirix G, Monsieurs P, Dombrecht B, Daniels R, Marchal K, Vanderleyden J *et al* (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.

Dong D, Ako R, Wu B (2012). Crystal structures of human sulfotransferases: insights into the mechanisms of action and substrate selectivity. *Expert Opin Drug Metab Toxicol* **8:** 635-646.

Edgar RC (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 NP *et al* (1999). Tyrosine sulfation of the amino terminus of CCR5 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, Destefano SA (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 DM, Upadhyaya NM, Stiller J, Ellis JG, Dodds PN, Kazan K *et al* (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.

Goris J, Konstantinidis KT, Klappenbach JA, Coenye T, Vandamme P, Tiedje JM (2007). DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. *Int J Syst Evol Microbiol* **57:** 81-91.

Hacker J, Blum-Oehler G, Mühldorfer I, Tschäpe H (1997). Pathogenicity islands of virulent bacteria: structure, function and impact on microbial evolution. *Mol Microbiol* **23:** 1089-1097.

Han SW, Lee SW, Bahar O, Schwessinger B, Robinson MR, Shaw JB *et al* (2012). Tyrosine sulfation in a Gram-negative bacterium. *Nat Commun* **3:** 1153.

Hauben L, Vauterin L, Swings J, Moore ER (1997). Comparison of 16S ribosomal DNA sequences of all *Xanthomonas* species. *Int J Syst Bacteriol* **47:** 328-335.

Haubold B, Klotzl F, Pfaffelhuber P (2015). andi: fast and accurate estimation of evolutionary distances between closely related genomes. *Bioinformatics* **31:** 1169-1175.

Holland IB, Peherstorfer S, Kanonenberg K, Lenders M, Reimann S, Schmitt L (2016). Type I protein secretion-deceptively simple yet with a wide range of mechanistic variability across the family. *EcoSal Plus* **7:** doi:10.1128/ecosalplus.ESP-0019-2015.

Jacobs JM, Pesce C, Lefeuvre P, Koebnik R (2015). Comparative genomics of a cannabis pathogen reveals insight into the evolution of pathogenicity in *Xanthomonas*. *Front Plant Sci* **6:** 431.

Jacques MA, Arlat M, Boulanger A, Boureau T, Carrere S, Cesbron S *et al* (2016). Using ecology, physiology, and genomics to understand host specificity in *Xanthomonas*. *Annu Rev Phytopathol* **54:** 163-187.

Kakuta Y, Pedersen LG, Pedersen LC, Negishi M (1998). Conserved structural motifs in the sulfotransferase family. *Trends Biochem Sci* **23:** 129-130.

Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S *et al* (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28:** 1647-1649.

Kleist AB, Getschman AE, Ziarek JJ, Nevins AM, Gauthier PA, Chevigne A *et al* (2016). New paradigms in chemokine receptor signal transduction: Moving beyond the two-site model. *Biochem Pharmacol* **114:** 53-68.

Klotzl F, Haubold B (2016). Support values for genome phylogenies. *Life (Basel)* **6:** 11.

Kuo CH, Ochman H (2009). The fate of new bacterial genes. *FEMS Microbiol Rev* **33:** 38-43.

Langlois PA, Snelling J, Hamilton JP, Bragard C, Koebnik R, Verdier V *et al* (2017). Characterization of the *Xanthomonas translucen*s complex using draft genomes, comparative genomics, phylogenetic analysis, and diagnostic LAMP assays. *Phytopathology* **107:** 519-527.

Lin DY, Huang S, Chen J (2015). Crystal structures of a polypeptide processing and secretion transporter. *Nature* **523:** 425-430.

Ma J, Campbell A, 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 JE, Weinberg Z, Emilsson GM, Ruzzo WL *et al* (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 *et al* (2013). Evolutionary history of the plant pathogenic bacterium *Xanthomonas axonopodis*. *PLoS One* **8:** e58474.

Midha S, Patil PB (2014). Genomic insights into the evolutionary origin of *Xanthomonas axonopodis* pv. *citri* and its ecological relatives. *Appl Environ Microbiol* **80:** 6266-6279.

Mitchell JE, Zheng D, Busby SJ, Minchin SD (2003). Identification and analysis of 'extended -10' promoters in *Escherichia coli*. *Nucleic Acids Res* **31:** 4689-4695.

Moore ER, Kruger AS, Hauben L, Seal SE, Daniels MJ, De Baere R *et al* (1997). 16S rRNA gene sequence analyses and inter- and intrageneric relationships of *Xanthomonas* species and *Stenotrophomonas maltophilia*. *FEMS Microbiol Lett* **151:** 145-153.

Moore KL (2009). Protein tyrosine sulfation: a critical posttranslation modification in plants and animals. *Proc Natl Acad Sci U S A* **106:** 14741-14742.

Naushad S, Adeolu M, Wong S, Sohail M, Schellhorn HE, Gupta RS (2015). A phylogenomic and molecular marker based taxonomic framework for the order *Xanthomonadales*: proposal to transfer the families *Algiphilacea*e 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.

Nelson K, Wang FS, Boyd EF, Selander RK (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.

Niño-Liu DO, Ronald PC, Bogdanove AJ (2006). *Xanthomonas oryzae* pathovars: model pathogens of a model crop. *Mol Plant Pathol* **7:** 303-324.

O'Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, McVeigh R *et al* (2016). Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res* **44:** D733-745.

Parkinson N, Aritua V, Heeney J, Cowie C, Bew J, 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, Stead D (2009). Phylogenetic structure of *Xanthomonas* determined by comparison of *gyrB* sequences. *Int J Syst Evol Microbiol* **59:** 264-274.

Pereira UP, Gouran H, Nascimento R, Adaskaveg JE, Goulart LR, Dandekar AM (2015). Complete genome sequence of *Xanthomonas arboricola* pv. *juglandis* 417, a copper-resistant strain isolated from *Juglans regia* L. *Genome Announc* **3:** e01126-01115.

Pieretti I, Cociancich S, Bolot S, Carrere S, Morisset A, Rott P *et al* (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.

Potnis N, Timilsina S, Strayer A, Shantharaj D, Barak JD, Paret ML *et al* (2015). Bacterial spot of tomato and pepper: diverse *Xanthomonas* species with a wide variety of virulence factors posing a worldwide challenge. *Mol Plant Pathol* **16:** 907-920.

Pruitt RN, Schwessinger B, Joe A, Thomas N, Liu F, Albert M *et al* (2015). The rice immune receptor XA21 recognizes a tyrosine-sulfated protein from a Gram-negative bacterium. *Sci Adv* **1:** e1500245.

Pruitt RN, Joe A, Zhang W, Feng W, Stewart V, Schwessinger B *et al* (2017). A microbially derived tyrosine-sulfated peptide mimics a plant peptide hormone. *New Phytol* **215:** in press.

Rademaker JLW, Louws FJ, Schultz MH, Rossbach U, Vauterin L, Swings J *et al* (2005). A comprehensive species to strain taxonomic framework for Xanthomonas. *Phytopathology* **95:** 1098-1111.

Ronald PC, Beutler B (2010). Plant and animal sensors of conserved microbial signatures. *Science* **330:** 1061-1064.

Salzberg SL, Sommer DD, Schatz MC, Phillippy AM, Rabinowicz PD, Tsuge S *et al* (2008). Genome sequence and rapid evolution of the rice pathogen Xanthomonas oryzae pv. oryzae PXO99A. *BMC Genomics* **9:** 204.

Shen Y, Sharma P, da Silva FG, 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.

Song WY, Wang GL, Chen LL, Kim HS, Pi LY, Holsten T *et al* (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 MJ, Chuang S, Hou X, Shoham M, Zhu JZ (2009). Tyrosine sulfation: an increasingly recognised post-translational modification of secreted proteins. *N Biotechnol* **25:** 299-317.

Studholme DJ, Wasukira A, Paszkiewicz K, Aritua V, Thwaites R, Smith J *et al* (2011). Draft genome sequences of Xanthomonas sacchari and two banana-associated Xanthomonads reveal insights into the *Xanthomonas* group 1 clade. *Genes (Basel)* **2:** 1050-1065.

Teramoto T, Fujikawa Y, Kawaguchi Y, Kurogi K, Soejima M, Adachi R *et al* (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 *et al* (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.

Triplett LR, Hamilton JP, Buell CR, Tisserat NA, Verdier V, Zink F *et al* (2011). Genomic analysis of *Xanthomonas oryzae* isolates from rice grown in the United States reveals substantial divergence from known *X. oryzae* pathovars. *Appl Environ Microbiol* **77:** 3930-3937.

Triplett LR, Verdier V, Campillo T, Van Malderghem C, Cleenwerck I, Maes M *et al* (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, Koebnik R (2015). Draft genome sequences of two *Xanthomonas vesicatoria* strains from the balkan peninsula. *Genome Announc* **3:** e01558-01514.

Vandroemme J, Cottyn B, Baeyen S, De Vos P, Maes M (2013). Draft genome sequence of *Xanthomonas fragariae* reveals reductive evolution and distinct virulence-related gene content. *BMC Genomics* **14:** 829.

Varghese NJ, Mukherjee S, Ivanova N, Konstantinidis KT, Mavrommatis K, Kyrpides NC *et al* (2015). Microbial species delineation using whole genome sequences. *Nucleic Acids Res* **43:** 6761-6771.

Vauterin L, Hoste B, Kersters K, Swings J (1995). Reclassification of *Xanthomonas*. *Int J Syst Bacteriol* **45**.

Vauterin L, Rademaker J, Swings J (2000). Synopsis on the taxonomy of the genus *Xanthomonas*. *Phytopathology* **90:** 677-682.

Wang GL, Song WY, Ruan DL, Sideris S, Ronald PC (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 *et al* (2012). Genome-wide sequencing reveals two major sub-lineages in the genetically monomorphic pathogen *Xanthomonas campestris* pathovar *musacearum*. *Genes (Basel)* **3:** 361-377.

Young JM (2008). An overview of bacterial nomenclature with special reference to plant pathogens. *Syst Appl Microbiol* **31:** 405-424.

Young JM, Park DC, Shearman HM, Fargier E (2008). A multilocus sequence analysis of the genus *Xanthomonas*. *Syst Appl Microbiol* **31:** 366-377.

Zhou JM, Tang D, Wang G (2017). Receptor kinases in plant pathogen interactions: more than pattern recognition. *Plant Cell* **29:** 618-637.

Figure 1. The *raxX-raxSTAB* gene cluster. A. The *rax* genetic region, drawn to scale. B. Boundary sequences. Sequences conserved within a group but different from other groups are colored green, brown, or yellow. Black sequence is conserved in all lineages, and blue sequence represents matches to consensi for transcription and translation initiation sequences. An "*mfsX*" +1 frameshift in *Xoo* sequences is indicated by the vertical red line. Abbreviations: *S. maltophilia, Sm; X. albilineans, Xa; X. arboricola pv. juglandis, Xaj; X. axonopodis pv. manihotis, Xam; X. campestris pv. campestris, Xcc; X. campestris pv. musacearum, Xcm; X. cannabis, Xc; X. citri subsp. citri, Xac; X. euvesicatoria, Xe; X. fragariae, Xf; X. hyacinthi, Xh; X. maliensis, Xm; X. oryzae pv. oryzae,Xoo; X. sacchari,; Xs X. translucens, Xt; X. vesicatoria, Xv.*

Figure 2. Model for r*axX-raxSTAB* inheritance during *Xanthomonas* speciation. The *Xanthomonas* spp. cladogram is based on published phylogenetic trees; see text for references. Gray lines depict lineages for strains that lack the *raxX-raxSTAB* gene cluster, whereas black lines depict those that carry the cluster. Numbers indicate *gcvP* length polymorphism in each species (see **Fig. 4**). Hypothetical events are: A, formation of the *raxX-raxSTAB* gene cluster; B, lateral transfer to *X. translucens*; C, lateral transfer to *X. maliensis*; D, loss from *X. citri*.

Figure 3. Phylogenetic trees for *rax* gene nucleotide sequences. The best scoring maximum likelihood trees for (A) *raxA*, (B) *raxB*, (C) *raxX* and (D) *raxST* in *Xanthomona*s spp. Numbers shown on 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, nor for branches too short to easily distinguish.

Figure 4. GcvP length polymorphisms. 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 lines carry the cluster. Numbers denote different allelic types. The positions of residues Gly-733 and Val-738 (numbering for allelic type 1) are indicated. Abbreviations are as in Fig. 1b.

Figure 5. Phylogenetic tree for *raxST* homologs in diverse bacterial genera. Distribution of *raxST* homolgs across bacterial genera. The tree shown was constructed by neighbor-joining with 1000 bootstrap replicates; branches with < 50% bootstrap support are not drawn. The *raxST* sequence from *Xoo* strain PXO99^{A} was used as query for tBLASTn.

Figure S1. Whole genome-based *Xanthomonas* phylogenetic tree. Phylogenetic tree constructed from comparison of whole genome sequences; see text for details. See **Fig. 2** for the corresponding cladogram. Red lines depict lineages for strains that lack the *raxX-raxSTAB* gene cluster, whereas blue lines depict those that carry the cluster.

Figure S1.

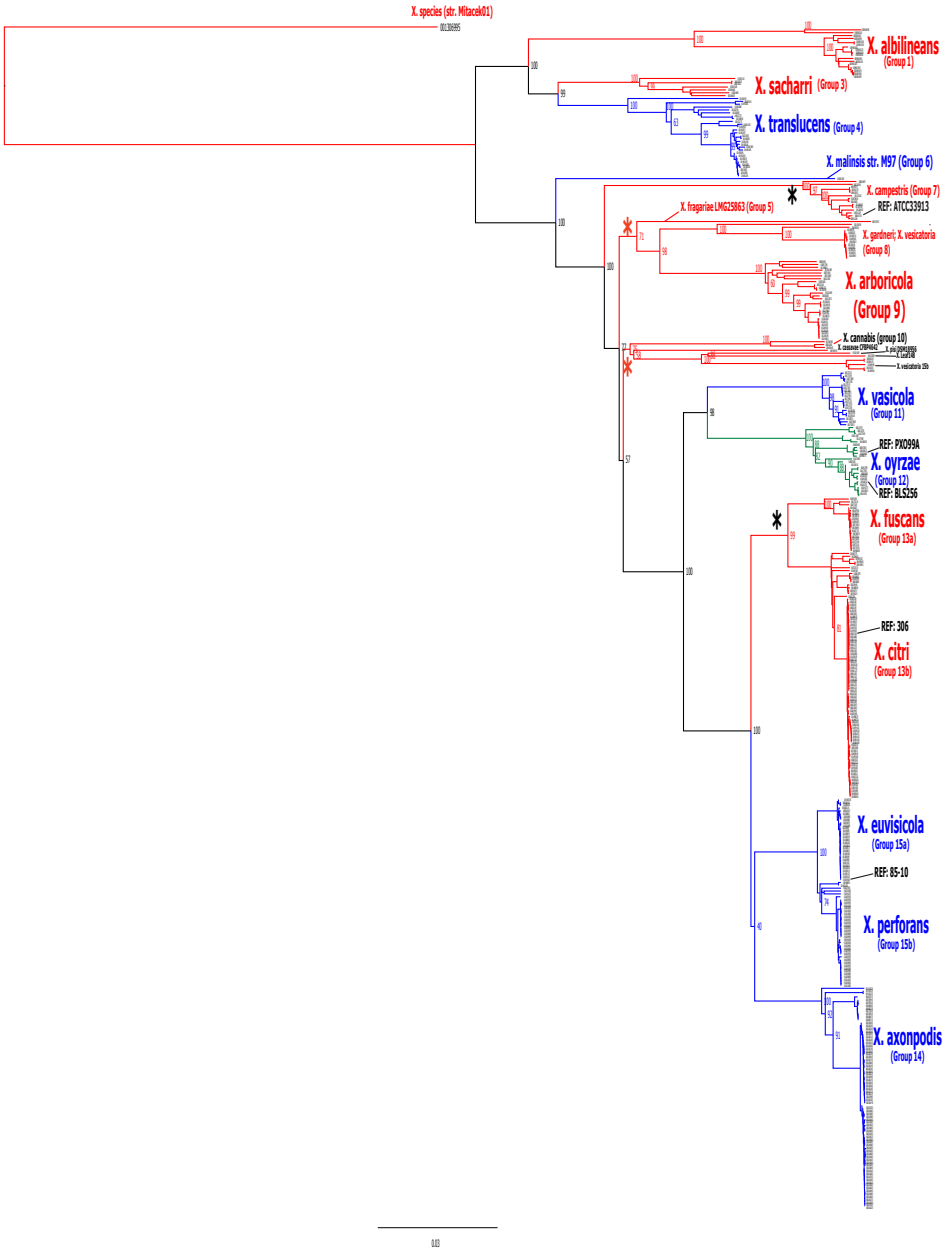

X. species (str. Mitacek01)
X. albilineans
(Group 1)
X. sacharri (Group 3)
X. translucens (Group 4)
X. malinsis str. M97 (Group 6)
X. campestris (Group 7)
REF: ATCC33913
X. fragariae LMG25863 (Group 5)
X. gardneri; X. vesicatoria
(Group 8)
X. arboricola
(Group 9)
X. cannabis (group 10)
X. vasicola
(Group 11)
REF: PXO99A
X. oyrzae
(Group 12)
REF: BLS256
X. fuscans
(Group 13a)
REF: 306
X. citri
(Group 13b)
X. euvisicola
(Group 15a)
REF: 85-10
X. perforans
(Group 15b)
X. axonpodis
(Group 14)
0.03

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	–	NZ_JSXZ00000000.1	(Vancheva *et al.,* 2015)

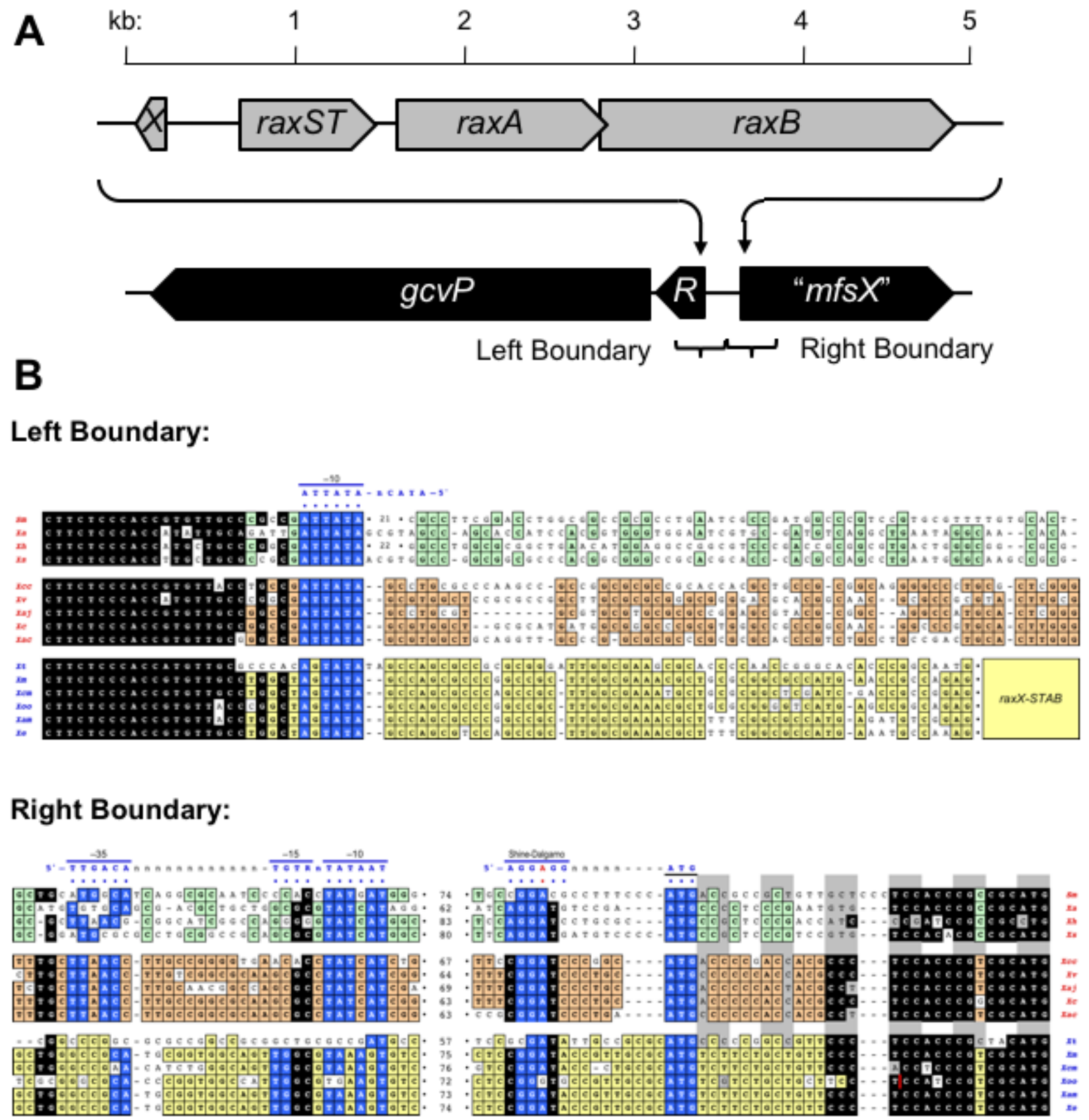
Table 2. Nucleotide sequence identity between *rax* genes.

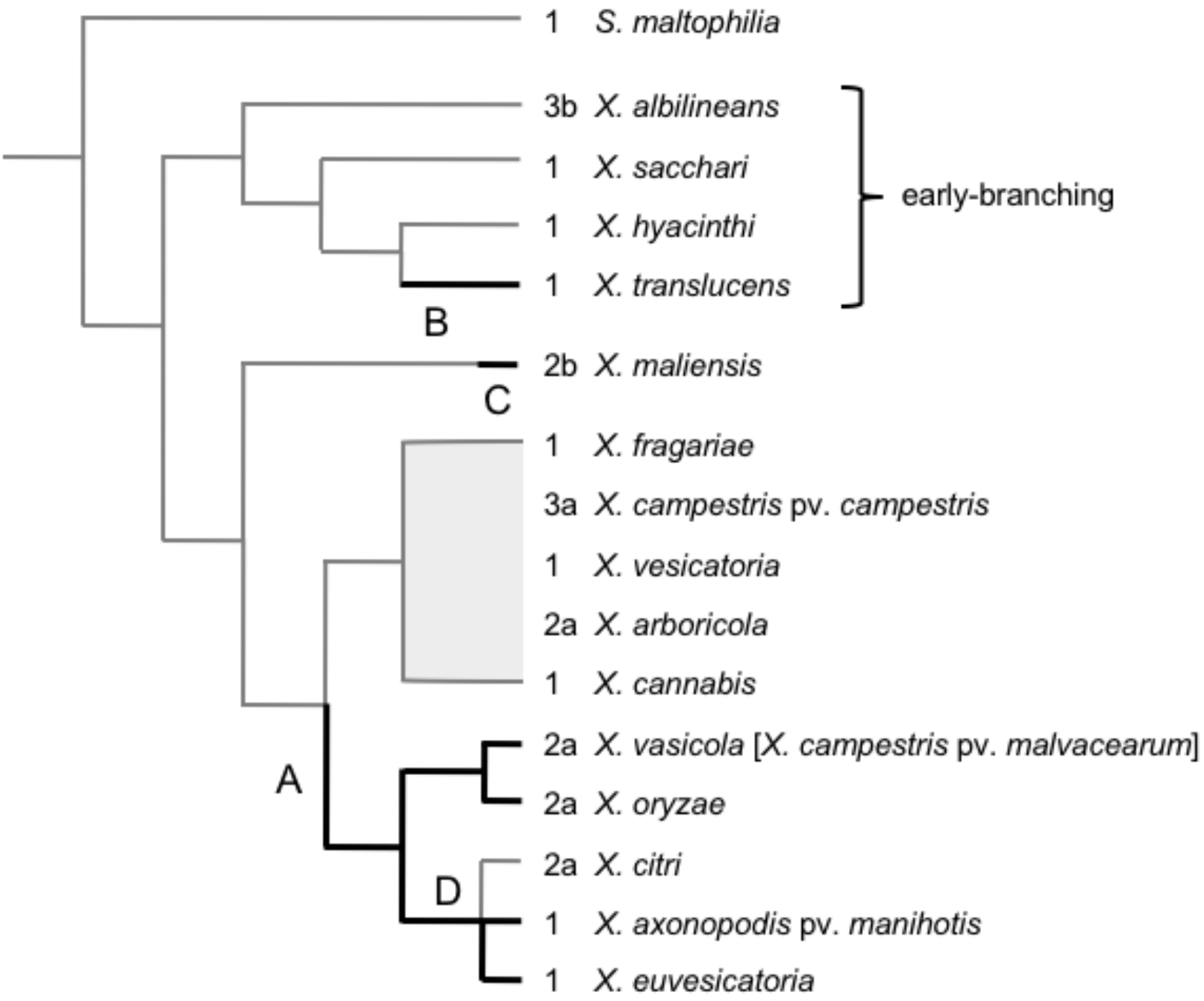
	Comparison to *X. euvesicatoria* sequence							
	% Nucleotide identity						Genome	
Species	*gcvP*	*raxX*	*raxST*	*raxA*	*raxB*	*"mfsX"*	gANI [a]	AF [b]
X. citri subsp. *citri*	94.8	— [c]	—	—	—	97.3	94.9	0.80
X. axonopodis pv. *manihotis*	94.1	99.0	96.1	97.6	95.8	97.3	95.0	0.79
X. maliensis	93.5	89.5	96.4	95.0	96.4	**85.0**	**83.5**	**0.66**
X. campestris pv. *malvacearum*	91.3	82.2	91.4	90.9	93.2	93.6	91.2	0.72
X. oryzae pv. *oryzae*	92.8	88.5	91.4	87.6	89.5	91.6	91.2	0.61
X. translucens	88.7	63.1	80.7	71.0	77.3	80.4	80.4	0.54

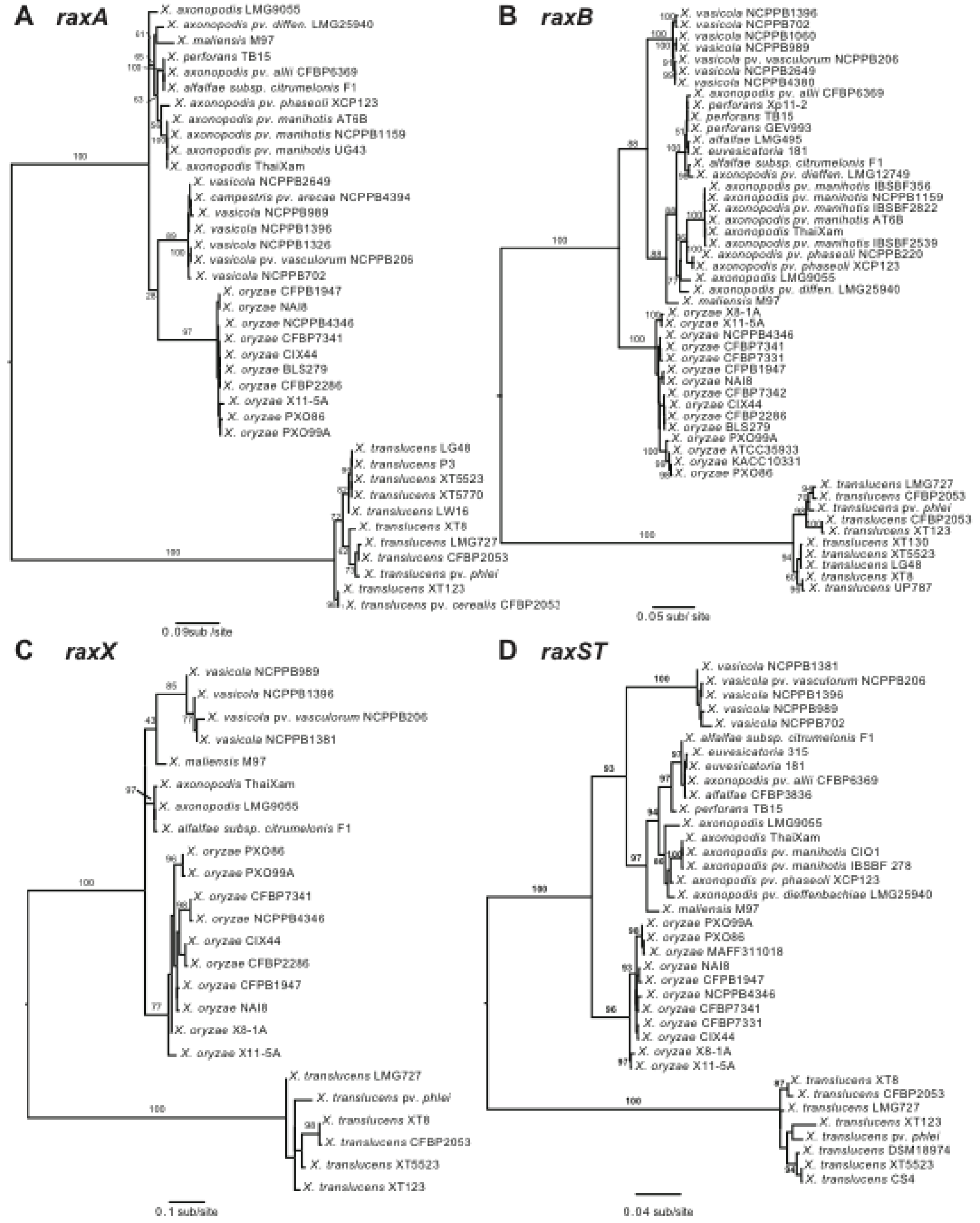
[a] genome-wide Average Nucleotide Identity (Varghese *et al.,* 2015).

[b] genome Alignment Fraction, from *X. euvesicatoria* to subject species (Varghese *et al.,* 2015).

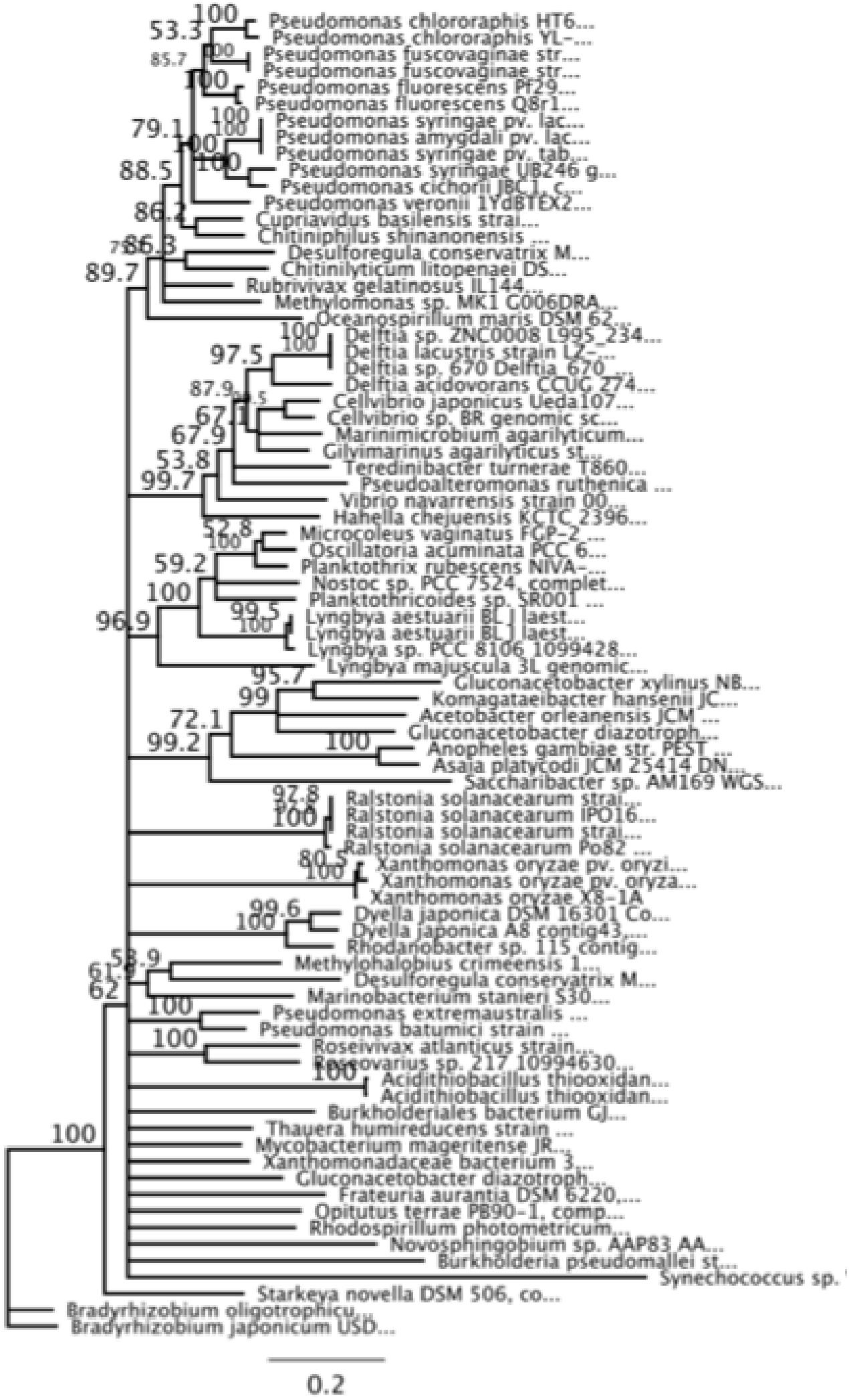
[c] —, gene not present


A
kb:
1
2
3
4
5
X
raxST
raxA
raxB
gcvP
R
"mfsX"
Left Boundary
Right Boundary
B
Left Boundary:
raxX-STAB
Right Boundary:
Shine-Dalgarno


1 S. maltophilia
3b X. albilineans
1 X. sacchari
1 X. hyacinthi
1 X. translucens
early-branching
B
2b X. maliensis
C
1 X. fragariae
3a X. campestris pv. campestris
1 X. vesicatoria
2a X. arboricola
1 X. cannabis
2a X. vasicola [X. campestris pv. malvacearum]
A
2a X. oryzae
2a X. citri
D
1 X. axonopodis pv. manihotis
1 X. euvesicatoria


A raxA
X. axonopodis LMG9055
X. axonopodis pv. diffen. LMG25940
X. maliensis M97
X. perforans TB15
X. axonopodis pv. alfi CFBP6369
X. alfalfae subsp. citrumelonis F1
X. axonopodis pv. phaseoli XCP123
X. axonopodis pv. manihotis AT6B
X. axonopodis pv. manihotis NCPPB1159
X. axonopodis pv. manihotis UG43
X. axonopodis ThaiXam
X. vasicola NCPPB2649
X. campestris pv. arecae NCPPB4394
X. vasicola NCPPB989
X. vasicola NCPPB1396
X. vasicola NCPPB1326
X. vasicola pv. vasculorum NCPPB206
X. vasicola NCPPB702
X. oryzae CFPB1947
X. oryzae NAI8
X. oryzae NCPPB4346
X. oryzae CFBP7341
X. oryzae CIX44
X. oryzae BLS279
X. oryzae CFBP2286
X. oryzae X11-5A
X. oryzae PXO86
X. oryzae PXO99A
X. translucens LG48
X. translucens P3
X. translucens XT5523
X. translucens XT5770
X. translucens LW16
X. translucens XT8
X. translucens LMG727
X. translucens CFBP2053
X. translucens pv. phlei
X. translucens XT123
X. translucens pv. cerealis CFBP2052
0.09sub /site
B raxB
X. vasicola NCPPB1396
X. vasicola NCPPB702
X. vasicola NCPPB1060
X. vasicola NCPPB989
X. vasicola pv. vasculorum NCPPB206
X. vasicola NCPPB2649
X. vasicola NCPPB4380
X. axonopodis pv. alfi CFBP6369
X. perforans Xp11-2
X. perforans TB15
X. perforans GEV993
X. alfalfae LMG495
X. euvesicatoria 181
X. alfalfae subsp. citrumelonis F1
X. axonopodis pv. diffen. LMG12749
X. axonopodis pv. manihotis IBSBF356
X. axonopodis pv. manihotis NCPPB1159
X. axonopodis pv. manihotis IBSBF2822
X. axonopodis pv. manihotis AT6B
X. axonopodis ThaiXam
X. axonopodis pv. manihotis IBSBF2539
X. axonopodis pv. phaseoli NCPPB220
X. axonopodis pv. phaseoli XCP123
X. axonopodis LMG9055
X. axonopodis pv. diffen. LMG25940
X. maliensis M97
X. oryzae X8-1A
X. oryzae X11-5A
X. oryzae NCPPB4346
X. oryzae CFBP7341
X. oryzae CFBP7331
X. oryzae CFPB1947
X. oryzae NAI8
X. oryzae CFBP7342
X. oryzae CIX44
X. oryzae CFBP2286
X. oryzae BLS279
X. oryzae PXO99A
X. oryzae ATCC35933
X. oryzae KACC10331
X. oryzae PXO86
X. translucens LMG727
X. translucens CFBP2053
X. translucens pv. phlei
X. translucens CFBP2053
X. translucens XT123
X. translucens XT130
X. translucens XT5523
X. translucens LG48
X. translucens XT8
X. translucens UP787
0.05 sub/ site
C raxX
X. vasicola NCPPB989
X. vasicola NCPPB1396
X. vasicola pv. vasculorum NCPPB206
X. vasicola NCPPB1381
X. maliensis M97
X. axonopodis ThaiXam
X. axonopodis LMG9055
X. alfalfae subsp. citrumelonis F1
X. oryzae PXO86
X. oryzae PXO99A
X. oryzae CFBP7341
X. oryzae NCPPB4346
X. oryzae CIX44
X. oryzae CFBP2286
X. oryzae CFPB1947
X. oryzae NAI8
X. oryzae X8-1A
X. oryzae X11-5A
X. translucens LMG727
X. translucens pv. phlei
X. translucens XT8
X. translucens CFBP2053
X. translucens XT5523
X. translucens XT123
0.1 sub/site
D raxST
X. vasicola NCPPB1381
X. vasicola pv. vasculorum NCPPB206
X. vasicola NCPPB1396
X. vasicola NCPPB989
X. vasicola NCPPB702
X. alfalfae subsp. citrumelonis F1
X. euvesicatoria 315
X. euvesicatoria 181
X. axonopodis pv. alfi CFBP6369
X. alfalfae CFBP3836
X. perforans TB15
X. axonopodis LMG9055
X. axonopodis ThaiXam
X. axonopodis pv. manihotis CIO1
X. axonopodis pv. manihotis IBSBF 278
X. axonopodis pv. phaseoli XCP123
X. axonopodis pv. dieffenbachiae LMG25940
X. maliensis M97
X. oryzae PXO99A
X. oryzae PXO86
X. oryzae MAFF311018
X. oryzae NAI8
X. oryzae CFPB1947
X. oryzae NCPPB4346
X. oryzae CFBP7341
X. oryzae CFBP7331
X. oryzae CIX44
X. oryzae X8-1A
X. oryzae X11-5A
X. translucens XT8
X. translucens CFBP2053
X. translucens LMG727
X. translucens XT123
X. translucens pv. phlei
X. translucens DSM18974
X. translucens XT5523
X. translucens CS4
0.04 sub/site

```
Xaj   2a   GVGPCAVKSHLAPYLPRAGI----HAGEGQTAAIHGGGLNSESGSGHSSRIGGMVSAAAYGSASILPISWM
Xcm   2a   GVGPCAVKSHLAPYLPRAGI----HAGEGQDVAAHGGGLNSESGAAGSLRTGGMVSAAAYGSASILPISWM
Xoo   2a   GVGPCAVKSHLAPFLPRAGL----HAGEGQTAAIHGGGFNSGSGSGHSSRIGGMVSAAAYGSASILPISWM
Xc    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
```


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 ...
Asaja 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 10994630...
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.
Starkeya novella DSM 506, co...
Bradyrhizobium oligotrophicu...
Bradyrhizobium japonicum USD...
0.2