Megan McDonald et al.

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 Pamela C. Ronald
Dpt. Microbiology & Molecular Genetics Department of Plant Pathology
University of California, Davis University of California, Davis
One Shields Avenue One Shields Avenue
Davis CA 95616-8665 USA Davis CA 95616-8665 USA
Email: vjstewart@ucdavis.edu Email: pcronald@ucdavis.edu
Office: (530) 754-7994 Office: (530) 752-1654
Fax: (530) 754-8973 Fax: (530) 752-5674
Megan McDonald et al.

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

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

Several Xoo rax genes are required for activation of XA21-mediated immunity (Fig. 1a). The raxX-raxSTAB gene cluster encodes the 60-residue RaxX predicted precursor protein that undergoes sulfation by the RaxST tyrosylprotein sulfotransferase at residue Tyr-41 (Pruitt et al., 2015). We hypothesize that the RaxABC proteolytic maturation and ATP-dependent peptide secretion complex (da Silva et al., 2004) further processes the sulfated RaxX precursor by removing its double-glycine leader peptide prior to secretion (Holland et al., 2016). Located outside the raxX-raxSTAB gene cluster, the raxC gene, an ortholog of the tolC gene, encodes the predicted outer membrane channel for this complex (da Silva et al., 2004). Finally, the raxPQ genes encode enzymes to assimilate sulfate into 3’-phosphoadenosine 5’-phosphosulfate (Shen et al., 2002), the sulfodonor for the RaxST tyrosylprotein sulfotransferase (Han et al., 2012).

Tyrosylprotein sulfotransferase is confined to the Golgi complex in both plants and animals (Moore 2009). Thus, this post-translational modification is targeted to a subset of cell surface and secreted proteins that influence a variety of eukaryotic physiological processes (Matsubayashi 2014; Stone et al., 2009). For example, tyrosine sulfation of the chemokine receptors CCR5 and CXCR4 is necessary for high-affinity binding not 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 plant peptide containing sulfated tyrosine (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. citrulononis (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).
Notably, the raxX-raxSTAB gene cluster is absent from the X. citri pathovar group, pathogenic on a range of dicots including citrus. This group, which includes certain strains denoted as X. axonopodis or X. campestris (Bansal et al., 2017), clusters phylogenetically among the X. oryzae, X. euvesicatoria and X. axonopodis pv. manihotis groups (Midha and Patil 2014; Rademaker et al., 2005; Vauterin et al., 1995) (Fig. 2).

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

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

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

Both the organization and size of the raxX-raxSTAB gene cluster are conserved across all six lineages in which it resides. To address hypotheses for patterns of raxX-raxSTAB gene cluster inheritance, we compared individual phylogenetic trees for each of the four rax genes to the overall Xanthomonas phylogenetic tree (Fig. 3) (Kuo and Ochman 2009). For all four genes, sequences in X. translucens, in the early-branching group, cluster separately from sequences in the other lineages. This finding is congruent with the hypothesis, that X. translucens acquired the raxX-raxSTAB gene
cluster relatively early during *Xanthomonas* speciation. For *X. maliensis*, the raxX-raxSTAB genes assort among those from *X. euvesicatoria* and the *X. axonopodis* pathovars *manihotis* and *phaseoli* (Fig. 3), even though the *X. maliensis* genome sequence itself is more distantly related (Fig. 2). This finding suggests that *X. maliensis* acquired the raxX-raxSTAB gene cluster relatively late during *Xanthomonas* speciation.

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

We further examined phylogenetic relationships by comparing nucleotide sequence identity across the *gcvP, raxX, raxST, raxA, raxB* and "*mfsX*" coding regions, each from the initiation through termination codon (Table 2). For comparison, values are presented also for genome-wide average nucleotide identity (gANI) as well as the alignment fraction (AF), which estimates the fraction of orthologous genes (Varghese *et al.*, 2015). For context, a widely-used criterion assigns 95% average nucleotide identity (ANI) as the cut-off point for species delineation (Goris *et al.*, 2007). Sequence from *X. euvesicatoria* is the reference.

The raxST, raxA and raxB coding sequences from *X. axonopodis* pv. *manihotis* and *X. maliensis* display the highest identity to those from *X. euvesicatoria*, at least 95% in each case. Sequences from *Xoo* and *X. vasicola* (also known as *X. campestris* pv. *malvacearum*) are about 90% identical, and those from *X. translucens* roughly 75%
identical (Table 2). The raxX coding sequences are more divergent, with identity to the X. euvesicatoria sequence ranging from almost 90% for X. maliensis and Xoo to only 63% for X. translucens (Table 2).

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

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

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

In raxX-raxSTAB gene cluster-positive genomes, sequence flanking these boundaries appears unrelated to the gcvP-"mfsX" intergenic sequence from raxX-raxSTAB gene cluster-negative genomes (Fig. 1b). Rather, it is well-conserved even in the early-
branching species, *X. translucens*. These results suggest that anomalous *raxX-raxSTAB* gene cluster phylogenetic distribution results from lateral gene transfer.

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

**raxST homologs are present in genomes of diverse bacterial species**

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

These *raxST* homologs are in a range of bacterial phyla including Proteobacteria and Cyanobacteria (Fig. 5). Nevertheless, for most species represented by multiple genome sequences, the *raxST* homolog was detected in a minority of individuals, so it is not part of the core genome in these strains. Moreover, relationships between species in a *raxST* gene phylogenetic tree bear no resemblance to those in the overall
tree of bacterial species. For example, in the raxST gene tree, sequences from Cyanobacteria are flanked on both sides by sequences from Proteobacteria (Fig. 5). Together, these findings provide evidence for lateral transfer of raxST homologs transfer (Kuo and Ochman 2009).
DISCUSSION

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

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

Three examples provide further evidence for lateral transfer. First, the raxX-raxSTAB gene cluster from the early branching species X. translucens has essentially the same size, composition and structure as the others. However, the X. translucens raxX-raxSTAB sequences are more divergent (Table 2). This predicts that the raxX-raxSTAB
gene cluster has been part of the *X. translucens* genome sufficiently long for sequence alterations to accumulate (Kuo and Ochman 2009).

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

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

In an alternative scenario, where the *X. citri* *raxX-raxSTAB* gene cluster was lost through deletion, the remaining sequence in the intergenic region would more closely resemble the *raxX-raxSTAB* gene cluster-positive boundary sequence. This conclusion is not supported by the intergenic region found in *X. citri*. A second alternative scenario, that the *raxX-raxSTAB* gene cluster formed after *X. citri* speciation, is not supported by the analysis of *X. translucens* sequences described above.
Broad phylogenetic distribution of the $\text{raxX}$-$\text{raxSTAB}$ gene cluster implies that its associated phenotypes can contribute to host interactions with diverse $\text{Xanthomonas}$ spp. The $\text{raxX}$-$\text{raxSTAB}$ gene cluster was identified in the context of the rice $\text{XA21}$-mediated immune response (da Silva et al., 2004; Pruitt et al., 2015), but the sequence and functional similarities between the bacterial $\text{RaxX}$ and the plant $\text{PSY}$ sulfopeptides suggests that $\text{RaxX}$ may mimic $\text{PSY}$ phytohormone activities to facilitate infection (Pruitt et al., 2017). Indeed, $\text{Xoo}$ strains that cannot synthesize sulfated $\text{RaxX}$ exhibit reduced virulence (Pruitt et al., 2017).

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

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

Survey of raxX-STAB gene clusters in Xanthomonas spp.

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

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

**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).
Megan McDonald et al.

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


Megan McDonald et al.

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.
Megan McDonald et al.

Gardiner DM, Upadhyaya NM, Stiller J, Ellis JG, Dodds PN, Kazan K et al (2014).


Megan McDonald et al.


Megan McDonald et al.


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 Algiphilaceae and Solimonadaceae to the order Nevskiales ord. nov. and to create a new family within the order Xanthomonadales, the family Rhodanobacteraceae fam. nov., containing the genus Rhodanobacter and its closest relatives. Antonie Van Leeuwenhoek 107: 467-485.


Megan McDonald et al.


Megan McDonald et al.


Megan McDonald et al.


Megan McDonald et al.


Megan McDonald et al.

**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 *raxX-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 *Xanthomonas* 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 homologs 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 PXO99A 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.
Megan McDonald et al.

Figure S1.
Table 1. Reference strains for sequence comparisons.

<table>
<thead>
<tr>
<th>Species</th>
<th>Strain</th>
<th>raxSTAB</th>
<th>Accession</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>S. maltophilia</td>
<td>K279a</td>
<td>–</td>
<td>NC_010943.1</td>
<td>(Crossman et al., 2008)</td>
</tr>
<tr>
<td>X. albilineans</td>
<td>GPE PC73</td>
<td>–</td>
<td>NC_013722.1</td>
<td>(Pieretti et al., 2015)</td>
</tr>
<tr>
<td>X. arboricola pv. juglandis</td>
<td>Xaj 417</td>
<td>–</td>
<td>NZ_CP012251.1</td>
<td>(Pereira et al., 2015)</td>
</tr>
<tr>
<td>X. axonopodis pv. manihotis</td>
<td>UA536</td>
<td>+</td>
<td>NZ_AKEQ00000000.1</td>
<td>(Bart et al., 2012)</td>
</tr>
<tr>
<td>X. campestris pv. campestris</td>
<td>ATCC 33913</td>
<td>–</td>
<td>NC_003902.1</td>
<td>(da Silva et al., 2002)</td>
</tr>
<tr>
<td>X. campestris pv. musacearum</td>
<td>NCPPB 4392</td>
<td>+</td>
<td>NZ_AKB100000000.1</td>
<td>(Wasukira et al., 2012)</td>
</tr>
<tr>
<td>X. cannabis</td>
<td>NCPPB 2877</td>
<td>–</td>
<td>NZ_JSZE00000000.1</td>
<td>(Jacobs et al., 2015)</td>
</tr>
<tr>
<td>X. citri subsp. citri</td>
<td>306</td>
<td>–</td>
<td>NC_003919.1</td>
<td>(da Silva et al., 2002)</td>
</tr>
<tr>
<td>X. euvesicatoria</td>
<td>85-10</td>
<td>+</td>
<td>NZ_CP017190.1</td>
<td>(Thieme et al., 2005)</td>
</tr>
<tr>
<td>X. fragariae</td>
<td>LMG 25863</td>
<td>–</td>
<td>NZ_AJRZ00000000.1</td>
<td>(Vandroemme et al., 2013)</td>
</tr>
<tr>
<td>X. hyacinthi</td>
<td>DSM 19077</td>
<td>–</td>
<td>JPLD00000000.1</td>
<td>(Naushad et al., 2015)</td>
</tr>
<tr>
<td>X. maliensis</td>
<td>M97</td>
<td>+</td>
<td>NZ_AQPR00000000.1</td>
<td>(Triplett et al., 2015)</td>
</tr>
<tr>
<td>X. oryzae pv. oryzae</td>
<td>PXO99A</td>
<td>+</td>
<td>NC_010717.2</td>
<td>(Salzberg et al., 2008)</td>
</tr>
<tr>
<td>X. sacchari</td>
<td>R1</td>
<td>–</td>
<td>NZ_CP010409.1</td>
<td>(Studholme et al., 2011)</td>
</tr>
<tr>
<td>X. translucens</td>
<td>DAR61454</td>
<td>+</td>
<td>GCA_000334075.1</td>
<td>(Gardiner et al., 2014)</td>
</tr>
<tr>
<td>X. vesicatoria</td>
<td>15b</td>
<td>–</td>
<td>NZ_JSXZ00000000.1</td>
<td>(Vancheva et al., 2015)</td>
</tr>
</tbody>
</table>
Megan McDonald et al.

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

<table>
<thead>
<tr>
<th>Species</th>
<th>% Nucleotide identity</th>
<th>Genome</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>X. citri</em> subsp. <em>citri</em></td>
<td>94.8</td>
<td>94.9, 0.80</td>
</tr>
<tr>
<td><em>X. axonomopodis</em> pv. <em>manihotis</em></td>
<td>94.1, 99.0, 96.1, 97.6, 95.8, 97.3</td>
<td>95.0, 0.79</td>
</tr>
<tr>
<td><em>X. maliensis</em></td>
<td>93.5, 89.5, 96.4, 95.0, 96.4, 85.0</td>
<td>83.5, 0.66</td>
</tr>
<tr>
<td><em>X. campestris</em> pv. <em>malvacearum</em></td>
<td>91.3, 82.2, 91.4, 90.9, 93.2, 93.6</td>
<td>91.2, 0.72</td>
</tr>
<tr>
<td><em>X. oryzae</em> pv. <em>oryzae</em></td>
<td>92.8, 88.5, 91.4, 87.6, 89.5, 91.6</td>
<td>91.2, 0.61</td>
</tr>
<tr>
<td><em>X. translucens</em></td>
<td>88.7, 63.1, 80.7, 71.0, 77.3, 80.4</td>
<td>80.4, 0.54</td>
</tr>
</tbody>
</table>

* 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
- raxST → raxA → raxB
- gcVP
- R
- “mfsX”
- Left Boundary
- Right Boundary

B

Left Boundary:

- ATGATA-ACATA
- Sequence alignment

Right Boundary:

- -35
- -15
- -10
- Shine-Dalgarno

Sequence alignment
<table>
<thead>
<tr>
<th></th>
<th>Sequence</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Xaj</td>
<td>GVPCKSAPLAPYPRAGI----HAGEGQTAAIHGGGLNSESNGSHSSRIGGMVASAAYGSASILPISWM</td>
<td></td>
</tr>
<tr>
<td>Xcm</td>
<td>GVPCKSAPLAPYPRAGI----HAGEGQDVAHGGGLNSESNGAAGSLRTGGMVASAAYGSASILPISWM</td>
<td></td>
</tr>
<tr>
<td>Xoo</td>
<td>GVPCKSAPLAPLPRAGI----HAGEGQTAAIHGGGFNSGSNGSHSSRIGGMVASAAYGSASILPISWM</td>
<td></td>
</tr>
<tr>
<td>Xc</td>
<td>GVPCKSAPLAPLPRAGI----HAGEGQTAAIHGGGNSESNGHSSRIGGMVASAAYGSASILPISWM</td>
<td></td>
</tr>
<tr>
<td>Xm</td>
<td>GVPCKSAPLAPYPRAGI----HGGGFNSGSNGSHSSRIGGMVASAAYGSASILPISWM</td>
<td></td>
</tr>
<tr>
<td>Xoc</td>
<td>GVPCKSAPLAPLPRAGI----HAGGFNSGSNGSHSSRIGGMVASAAYGSASILPISWM</td>
<td></td>
</tr>
<tr>
<td>Xcc</td>
<td>GVPCKSAPLKLPLPPNAGRAGNAGAHLGGSNSF---GEG---VGMVASASYGSASILPISWM</td>
<td></td>
</tr>
<tr>
<td>Xa</td>
<td>GVPCKSAPLAPLPMTPN----AGEAQQAA----GEGV---VGMVASAASFSGASILPISWM</td>
<td></td>
</tr>
<tr>
<td>Sm</td>
<td>GVPCKSAPLAPLPGKLG---DNGP---VGMVASAASFSGASILPISWM</td>
<td></td>
</tr>
<tr>
<td>Xh</td>
<td>GVPCKSAPLAPLPTLG----GEGD---VGMVASAASFSGASILPISWM</td>
<td></td>
</tr>
<tr>
<td>Xs</td>
<td>GVPCKSAPLAPLPTLG----GDRE---VGMVASAASFSGASILPISWM</td>
<td></td>
</tr>
<tr>
<td>Xt</td>
<td>GVPCKSAPLAPLPTLG----GEGD---VGMVASAASFSGASILPISWM</td>
<td></td>
</tr>
<tr>
<td>Xf</td>
<td>GVPCKSAPLAPLPTLG----SEGD---VGMVASAASYGSASILPISWM</td>
<td></td>
</tr>
<tr>
<td>Xv</td>
<td>GVPCKSAPLAPLPTLG----SEGD---VGMVASAASYGSASILPISWM</td>
<td></td>
</tr>
<tr>
<td>Xc</td>
<td>GVPCKSAPLAPLPTLG----GEGD---VGMVASAASYGSASILPISWM</td>
<td></td>
</tr>
<tr>
<td>Xam</td>
<td>GVPCKSAPLAPLPTLG----GEGD---VGMVASAASYGSASILPISWM</td>
<td></td>
</tr>
<tr>
<td>Xe</td>
<td>GVPCKSAPLAPLPTLG----GEGD---VGMVASAASYGSASILPISWM</td>
<td></td>
</tr>
</tbody>
</table>

Gly-733  Val-738