bioRxiv preprint doi: https://doi.org/10.1101/180901; this version posted August 25, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1 **Genome shuffling in a globalized bacterial plant pathogen: Recombination-mediated**
2 **evolution in *Xanthomonas euvesicatoria* and *X. perforans***
3
4 Mustafa O. Jibrin[1,2,3], Neha Potnis[4], Sujan Timilsina[1], Gerald V. Minsavage[1], Gary E. Vallad[1,5],
5 Pamela D. Roberts[1,2,], Jeffrey B. Jones[1] and Erica M. Goss[1,6#]
6 [1]Department of Plant Pathology, University of Florida, Gainesville, Florida, USA.
7 [2]Southwest Research and Education Center, University of Florida, Immokalee, Florida, USA.
8 [3]Department of Crop Protection, Ahmadu Bello University, Zaria, Nigeria.
9 [4]Department of Entomology and Plant Pathology, Auburn University, Auburn, Alabama, USA.
10 [5]Gulf Coast Research and Education Center, University of Florida, Wimauma, Florida, USA.
11 [6]Emerging Pathogen Institute, University of Florida, Gainesville, Florida, USA.
12
13 Running Head: Genomic Recombination in Xanthomonads
14 #Address correspondence to Erica M. Goss, emgoss@ufl.edu.
15
16
17
18
19
20
21
22
23

bioRxiv preprint doi: https://doi.org/10.1101/180901; this version posted August 25, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

24 **ABSTRACT**

25 Bacterial recombination and clonality underly the evolution and epidemiology of pathogenic
26 lineages as well as their cosmopolitan spread. While the spread of stable clonal bacterial
27 pathotypes drives disease epidemics, recombination leads to the evolution of new bacterial
28 lineages. Recombinant lineages of plant bacterial pathogens are typically associated with
29 colonization of novel hosts and emergence of new diseases. Here, we show that recombination
30 between evolutionarily and phenotypically distinct plant pathogenic lineages has generated new
31 recombinant lineages with unique combinations of pathogenicity and virulence factors. *X.*
32 *euvesicatoria* (*Xe*) and *X. perforans* (*Xp*) are two closely related monophyletic species causing
33 bacterial spot disease on tomato and pepper worldwide. We sequenced the genomes of strains
34 representing populations on tomato in Nigeria and found shuffling of secretion systems and
35 effectors such that these strains contain genes from both *Xe* and *Xp*. Multiple strains, from
36 populations in Nigeria, Italy, and Florida, USA, exhibited extensive genomewide homologous
37 recombination and both species exhibited dynamic open pangenomes. Our results show that
38 recombination is generating new lineages of bacterial spot pathogens on tomato with
39 consequences for disease management strategies.

40

41

42 ***Importance***

43 The *Xanthomonas* pathogens that cause bacterial spot of tomato and pepper have been model
44 systems for plant-microbe interactions. Two of these pathogens, *X. euvesicatoria* and *X.*
45 *perforans*, are very closely related. Genome sequences of bacterial spot field strains from
46 Nigeria, Italy, and the United States showed varying levels of homologous recombination that
47 changed the amino sequence of effectors, secretion systems, and other proteins. This shuffling of

bioRxiv preprint doi: https://doi.org/10.1101/180901; this version posted August 25, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

48 genome content occurred between *X*. *euvesicatoria* and *X*. *perforans*, while a Nigerian lineage
49 also contained the lipopolysaccharide cluster of a distantly related *Xanthomonas* species. Gene
50 content varied among strains and the affected genes are important in the establishment of
51 disease, therefore our findings point to global variation in the host-pathogen interaction driven by
52 gene exchange among evolutionarily distinct lineages.

bioRxiv preprint doi: https://doi.org/10.1101/180901; this version posted August 25, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

53 **Introduction**

54 While what defines a bacterial species remains contentious, our understanding the evolution of
55 bacterial lineages is dependent on the elucidation of the various forms of adaptive divergence
56 they experience (1-3). Bacterial population structure may change over time, such that an
57 historically clonal bacterial population can experience recombination to the point of panmixis
58 and return to clonality as ecologically stable forms emerge and increase in frequency by positive
59 selection (4-6). Processes that lead to ecologically stable populations include the cessation of
60 panmixis accompanied by continued accumulation of mutations in what is now described as the
61 ecotype model (7-8). Clonality maintains adaptive gene complexes and produces the epidemic
62 population structure described by Maynard Smith and colleagues (4). This conceptual model has
63 been used to describe the emergence, spread and cosmopolitan fitness of virulent clones of the
64 O1 and O139 serotypes of *Vibrio cholera*, and evolutionary processes leading to the divergence
65 of clades of *Bacillus simplex* and *Mycobacterium tuberculosis* (6; 9-11). Among plant pathogens,
66 populations of rapidly spreading bacteria such as *Pseudomonas syringae* pv. *actinidae* and *P.*
67 *syringae* pv. *aesculi* causing the kiwifruit canker epidemic in Europe/New Zealand and bleeding
68 canker disease of horse chestnut in Europe, respectively, are viewed as emergent clonal
69 pathotypes that have colonized a new ecological niche (12-13).

70 In the ecotype model, non-homologous and homologous recombination are major drivers
71 of evolution in bacterial populations (14-17; 6). Horizontal gene transfer (HGT, also called non-
72 homologous recombination or lateral gene transfer), occurs between and within species (18).
73 Transformation, transduction and plasmid-mediated gene transfer are all mechanisms through
74 which HGT is known to occur in bacteria (19). The rate of HGT can be high and produces
75 substantial variation in gene content even among members of a taxonomic group (20, 14; 21).

bioRxiv preprint doi: https://doi.org/10.1101/180901; this version posted August 25, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

76 Recombination can spread beneficial mutations (otherwise at risk of elimination due to clonal
77 interference) and bring together independently evolved beneficial mutations (15). The result is
78 that DNA sequences of new bacterial lineages are often mosaics of several closely related
79 bacterial sequences (22). While HGT involves the direct transfer and receipt of genes between
80 strains, homologous recombination requires some level of sequence homology. Previously, the
81 prevailing model was that a high degree of sequence identity was important and that the
82 efficiency of recombination decreased greatly as identity decreased (23). However, it has been
83 shown that the level of sequence identity does not need to be constant across the gene length,
84 rather, it is most important within minimum efficiently processed segments (MEPS) which are
85 present in the inserted DNA's flanking regions (16). In fact, the regions between MEPS may not
86 share homology. The need for some level of homology means that recombination is highest
87 among close relatives and across regions of the 'core' genome where homology is retained (24,
88 25).

89 Advancements in genome sequencing technology has led to genome-based analyses of
90 bacterial recombination. These studies have shown the importance of recombination in the
91 evolution of major bacterial pathogens of humans (26, 24-25, 27-28). Genomic studies on the
92 relative role of recombination in plant pathogenic bacterial evolution remain limited.
93 Recombination has played a role in host shifts by xanthomonads infecting Brassicaceae and
94 citrus (29-30). Adaptation to agricultural crops by *Pseudomonas* pathogens has also been shown
95 to be influenced by recombination (31). Plant disease management strategies could be aided by
96 improved understanding of the contribution of recombination to plant bacterial pathogen
97 evolution.

bioRxiv preprint doi: https://doi.org/10.1101/180901; this version posted August 25, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

98 The genus *Xanthomonas*, which consists of more than twenty-seven described pathogenic
99 species infecting over four hundred plant hosts, is one of the most important groups of bacteria
100 causing plant disease (32). Bacterial spot of tomato and pepper is found around the world and is
101 caused by four *Xanthomonas* species: *X. euvesicatoria* (*Xe*), *X. vesicatoria* (*Xv*), *X. perforans*
102 (*Xp*) and *X. gardneri* (*Xg*) (33-35). Strains belonging to *Xe* and *Xp* are closely related (Potnis *et*
103 *al.*, 2011) and may comprise a single species (36). Yet, *Xe* and *Xp* possess different sets of
104 secreted effectors that elicit hypersensitive reactions on tomato differential lines (37, 35). Type
105 III secretion effectors (T3SEs) are essential virulence factors that are secreted through the highly
106 conserved type III secretion system (T3SS) and translocated into host plants where they interfere
107 with host immunity (38). Within *Xp*, MLSA and core genome analyses have shown distinct
108 evolutionary groups: 1A, 1B and 2 (39-40). A core genome phylogeny of *Xp* strains from Florida
109 showed that Group 1A formed a monophyletic lineage of strains isolated in 2012. The reference
110 strain for *Xp*, Xp91-118, represented one of multiple lineages in group 1B together with strains
111 from 2006. Group 2 strains formed another distinct lineage and represented the bulk of all
112 sequenced strains, and included strains from 2006, 2010, 2012 and 2013 (40). Irrespective of
113 these groupings, there are effectors that have been consistently associated with either *Xe* or *Xp*
114 and are described as *Xe*-specific and *Xp*-specific effectors (35, 40). Other studies have found
115 unique bacterial spot strains in Europe and Africa, suggesting the possibility of additional
116 evolutionarily distinct lineages (41-44).

117 We previously reported the occurrence in Nigeria of atypical bacterial spot of tomato
118 strains (represented by strain NI1). These strains, collected in 2014, were identified as *Xp*
119 because differential reactions on tomato genotypes indicated the presence of the *Xp* effector
120 protein AvrXv3, and the highly conserved gene *hrcN* showed 100% sequence identity to all

bioRxiv preprint doi: https://doi.org/10.1101/180901; this version posted August 25, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

121 previously isolated *Xp* strains (44). However, multi-locus sequence analysis using six
122 housekeeping genes clustered these Nigerian strains with *Xe*. Specifically, the *gltA*, *lacF*, and
123 *gapA* sequences were identical or nearly identical to those of *Xe*, yet genes *fusA* and *gyrB* were
124 identical to alleles found in *Xp* group 1 strains and the *lepA* allele was distinct from both *Xe* and
125 *Xp* (39). Strains like NI1 were isolated from the same tomato field in 2015, but we also identified
126 a second group of strains. Differential reactions indicated that these strains contained the *Xe*
127 effector protein AvrRxv. However, the *hrcN* gene sequence placed the strains in *Xp*. This was
128 surprising because no *Xp* field strain has been reported to contain *avrRxv*. The presence of this
129 gene was confirmed in representative strain NI38 by Sanger sequencing (Jibrin, Roberts, Jones
130 and Minsavage, *unpublished*). In summary, genotypic and phenotypic tests typically used to
131 assign bacterial spot strains to species could not assign the Nigerian strains to a *Xanthomonas*
132 species due to conflicting race, MLSA, and/or *hrcN* data.

133 The objective of this study was to examine the evolutionary processes that led to the
134 conflicting assignments of the Nigerian strains. We hypothesized that recombination had
135 generated lineages of bacterial spot strains with unique T3SS and T3SE gene profiles. We are
136 interested in T3SS effectors, because they are often targets of resistance breeding and are used to
137 monitor pathogen population shifts (37). Representatives of the Nigerian strains, NI1 and NI38,
138 were sequenced and compared with 63 previously sequenced strains, focusing mostly but not
139 exclusively on *Xe* and *Xp*. Our results show that recombination is driving the evolution of new
140 lineages of bacterial spot pathogens and thereby affecting management decisions.

141

142 Results

143 *Average nucleotide identity distinguishes NI1 from Xe and Xp*

bioRxiv preprint doi: https://doi.org/10.1101/180901; this version posted August 25, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

144 Whole genome sequences were obtained for NI1 and NI38 as representative strains of Nigerian
145 bacterial spot populations. Average nucleotide identity (ANI) statistics showed >99% identity
146 among strains within species (within *Xp* and within *Xe*) and <99% identity in *Xp* to *Xe*
147 comparisons (Table 1; Supplementary File 1). In contrast, the NI1 strain showed only 98.9% and
148 98.5% identity in pairwise comparisons to *Xp* and *Xe* reference genomes, respectively (Table 1)
149 and <99% identity to all other *Xp* and *Xe* strains (Supplementary File 1). The second strain
150 representing Nigerian populations, NI38, showed ANIs typical of an *Xe* strain: ANIs were >99%
151 in comparisons against *Xe* genomes and <99% against *Xp* genomes. The ANI results also showed
152 that strains belonging to *Xp* group 1A had more conserved ANI (≥99.97) while more variability
153 was found within groups 1B and 2 (Supplementary File 1).
154
155 ***Xe and Xp have open pangenomes***
156 Analysis of the *Xe* and *Xp* pan and core genomes showed heterogeneity in genome composition.
157 We observed a gradual decrease in core genome size with the addition of strains, and a bimodal
158 distribution in core genome size depending on the genomes re-sampled (Supplementary Figure
159 S1). Likewise, the pangenome size was highly dependent on the strains sampled. For *Xe*,
160 addition of the NI38 genome produced only a small increase in pangenome size (Supplementary
161 Figure S1). For *Xp*, addition of the Italian strain Xp_4P1S2 and NI1 produced a noticeable uptick
162 in the size of the *Xp* pangenome. The increasing size of the pangenomes as more strains were
163 sampled indicates open pangenomes for both *Xe* and *Xp*.
164 To examine the differences in gene content between NI1, NI38 and other *Xp* and *Xe*
165 strains, we produced a pangenome tree using Gower's similarity coefficient, which is 0 when
166 genomes have identical gene content and increases up to a maximum of 1 with increasing

bioRxiv preprint doi: https://doi.org/10.1101/180901; this version posted August 25, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

167 dissimilarity. The Gower's coefficient values clearly showed NI1 to be intermediate between
168 *Xe* and *Xp* (Figure 1), which is in agreement with the ANI results. NI38 clustered with other *Xe*
169 strains.
170

171 ***Phylogenetic reconstruction groups NI1 with Xp and NI38 with Xe***

172 The maximum likelihood phylogeny for concatenated core genes showed *Xe* and *Xp* as diverged
173 monophyletic groups, consistent with previous studies (Figure 2; 39-40). NI1 and 4P1S2 formed
174 distinct lineages of *Xp*, while NI38 is an *Xe* lineage. NI1 is substantially diverged from the other
175 *Xp* strains, but the genetic distance between NI1 and other *Xp* strains was not as great as the
176 interspecific distance between *Xp* and *Xe*.
177

178 ***Extensive recombination in lineages of Xp***

179 ClonalFrameML analyses showed different rates of homologous recombination in *Xe* and *Xp*
180 (Table 2). The ratios of recombination to mutation for *Xe* showed that homologous
181 recombination of imported DNA occurred at less than half the rate of mutation (R/θ) across all
182 *Xe* strains. For *Xp*, estimated rates of recombination to mutation were dependent on the
183 population used for analysis. For all *Xp* strains, recombination and mutation approached the
184 same rate (R/θ=0.85). When strains from the Florida population were considered exclusively
185 (NI1 and 4P1S2 excluded), recombination reached 1.5 times the rate of mutation. In all cases, the
186 overall impact of recombination on nucleotide variation was greater than that of mutation (R/m >
187 1).

188 The extensive homologous recombination in lineages of *Xe* and *Xp* is visualized in Figure
189 3. The number of recombination events is generally high in ancestral strains of each lineage, with

bioRxiv preprint doi: https://doi.org/10.1101/180901; this version posted August 25, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

190 a mostly clonal genealogy within each lineage. Strain NI1 had a particularly high number of
191 inferred recombination events, and large chromosomal replacements, shown in Figure 3 by long
192 continuous dark blue lines. Strains NI38, 4P1S2, Xp17-12, Xp5-6 and Xp4-20, which represent
193 independent lineages, also had high numbers of recombination events. Regions of homoplasy
194 were more common in strains of *Xe* than strains of *Xp,* suggesting that there may be more
195 undetected recombination events in *Xe*.
196

197 ***NI1 LPS cluster resembles the LPS cluster of X. translucens pv. translucens***

198 Lipopolysaccharides are highly antigenic and often act as pathogen associated molecular
199 patterns, virulence factors and defense response elicitors (45-47). The lipopolysaccharide (LPS)
200 gene clusters in the genus *Xanthomonas* usually consist of 15 genes and are flanked by two
201 conserved housekeeping genes, cystathionine gamma lyase (*metB)* and electron transport
202 flavoprotein *(etfA)* (48). A previous comparison of the LPS clusters of bacterial spot
203 xanthomonads using reference strains showed that *Xp* has a unique LPS cluster while *Xe, Xv* and
204 *Xg* have similar clusters (35). We found high gene conservation and little variation among *Xp*
205 strains (Figure 4b). In contrast, the NI1 LPS cluster contains several additional genes. For
206 example, NI1 has two glycosyltransferase genes which belong to a cluster of orthologous genes
207 (COG1216, GT2 family) that are absent in other *Xp* LPS clusters. BLAST searches of genes
208 unique to the NI1 LPS cluster revealed similarity with the LPS cluster of *Xanthomonas*
209 *translucens pv. translucens* (*Xtt)*, causal agent of bacterial wilt (also known as black chaff) of
210 barley (*Hordeum vulgare*). Out of two available *Xtt* genome sequences on the IMG database, the
211 NI1 LPS cluster was most similar to the *Xtt* type strain, DSM 18974, which was isolated from
212 barley in the United States (49; Figure 4a).

bioRxiv preprint doi: https://doi.org/10.1101/180901; this version posted August 25, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

213

214 ***Recombination affected type 3 and type 4 secretion system genes***

215 The Type 3 Secretion System (T3SS) is a delivery pathway for secreted effectors and avirulence
216 proteins, which are effectors that are recognized by the plant thus triggering a resistance
217 response. NI1 generally contains the expected *Xp* alleles for T3SS genes, but the *Xp* strain 4P1S2
218 contained a mixture of *Xe* and *Xp* alleles (Table 3). NI38 also exhibited a mix of *Xe* and *Xp* T3SS
219 alleles (Table 3; Supplementary Table S3).

220 Two Type 4 Secretion Systems (T4SS) were reported in bacterial spot xanthomonads, the
221 Vir and Dot/Icm systems (35). The reference strain for *Xe* has both systems and the type strain of
222 *Xp* has only the Vir system. The T4SS is important for horizontal gene transfer between bacteria
223 and delivery of effectors into hosts. Both types of T4SS were found in all *Xe* strains, including
224 NI38, and also in NI1. With the exception of strain 4P1S2, all other *Xp* strains have a complete
225 Vir system and only the *FimT* and *PilC* genes of the Dot/Icm system, lacking *PilE*, *PilV*, *PilW*,
226 *PilX* and *PilY1* (Supplementary Table S4). Strain 4P1S2 is unique among these in having the
227 *PilE* gene.

228 We did not identify variation in gene composition in the T2SS, T5SS or T6SS among
229 strains. Both NI1 and NI38 had the type 1 and 3 T6SS clusters that were previously identified in
230 bacterial spot xanthomonads (35).

231

232 ***NI1 and NI38 contain a mix of Xe-specific and Xp-specific effectors***

233 The mix of genes in the T3SS of NI1 and NI38 suggested the possibility of variability in T3SS
234 effectors, which are often primary targets of resistance breeding (35, 40). The NI1 genome
235 possesses all 11 previously described genes for core effectors, as well as *xopE2*, which was

bioRxiv preprint doi: https://doi.org/10.1101/180901; this version posted August 25, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

236 recently added to the core effectors of bacterial spot xanthomonads (Supplementary Table S5;
237 35, 40). Among the groups of effectors that are shared by *Xe* and *Xp* but absent in other bacterial
238 spot xanthomonads, NI1 possesses all but *xopF2* (Table 4). Two of the four copies of *xopP* in
239 NI1, together with *xopAK*, have 100% amino acid identity with the corresponding *Xp* effectors,
240 but the *xopI* effector has 100% amino acid identity to the *Xe* allele of *xopI*. Some effectors have
241 been previously described as species-specific among the bacterial spot xanthomonads. NI1 has
242 *xopAJ* specific to *Xe*; *xopC2*, *xopAE* and *xopAF* specific to *Xp*; and *xopAQ* and *xvrHah1* specific
243 to *X. gardneri* (*Xg*). While the *xopAJ* of NI1 shares 98.7% similarity to the *Xe* allele, it is 100%
244 identical to a copy of *xopAJ* found in *X. axonopodis pv. poinsenttiicola*. NI1 lacked *xopJ4* which
245 was found to be conserved among *Xp* strains in Florida (39). Consistent with this finding, NI1
246 did not cause a hypersensitive response on tomato plants containing the *RXopJ4* resistance gene
247 (Figure 5). Strain 4P1S2 is similar to NI1 in that it contains *avrHah1* and *xopAQ*, however, it is
248 similar to other *Xp* in having *xopJ4*.

249 NI38 has the 11 bacterial spot core effector genes, with two copies each of *avrBs2* and
250 *xopD*, and four copies of *xopAD* (Supplementary Table S6). NI38 has two copies each of *xopE1*
251 and *xopP*, which are effector genes found exclusively in *Xe* and *Xp* among the bacterial spot
252 species. Of the species-specific effectors, NI38 has all the *Xe*-specific effectors except *avrBs1*,
253 and two copies each of *xopC2* and *xopAE* specific to *Xp*. Like other *Xe* strains, NI38 has *xopB*,
254 which is the only effector shared by *Xe* and *Xg* with 100% amino acid identity. The *xopAJ* gene
255 in NI38 is identical to a copy of *xopAJ* in *X. axonopodis pv. poinsenttiicola,* but a different copy
256 of the gene from the *xopAJ* in NI1. The only *Xe* effector missing in NI38 was *xopG*, which is an
257 effector common to *Xe*, *Xv* and *Xg*.

258

bioRxiv preprint doi: https://doi.org/10.1101/180901; this version posted August 25, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

259 **Discussion**

260 ***Core genome phylogenies resolved taxonomic assignments of NI1 and NI38***

261 Whole genome sequencing resolved the previous conflicting multilocus sequence analyses and
262 race differentiation tests of strains NI1 and NI38, which are representative of 2014 and 2015
263 tomato field collections from Nigeria. We have identified these strains as novel lineages of *X. perforans* and *X. euvesicatoria,* respectively, using whole genome phylogenetic analysis. They
265 were notably diverged from other *Xp* and *Xe* populations as a result of horizontal gene transfer
266 and homologous recombination that has affected housekeeping genes, lipopolysaccharide
267 clusters, secretion systems and effectors, among other genomic regions.

268 The close relatedness between strains of *X. euvesicatoria* and *X. perforans* is reflected in
269 high values of average nucleotide identities in pairwise comparisons. These species were
270 previously classified into separate species based on DNA:DNA hybridization (34) and form
271 distinct monophyletic groups (40), but they show ANIs well above 95%. ANI identified NI1 as
272 divergent from both *Xp* and *Xe*, and could not be used to assign this strain to either taxon.

273

274 ***Recombination has shaped the evolution of X. euvesicatoria and X. perforans lineages***

275 Our results show that recombination is a major driver of evolution of both *Xe* and *Xp*. We found
276 evidence of both homologous and non-homologous recombination. Rates of homologous
277 recombination varied within and between species, with *Xp* showing more evidence of
278 recombination than *Xe*, which is perhaps most obvious in the distinct, recombinant lineages of
279 *Xp*. Our analysis also detected higher recombination rates within the Florida population of *Xp*.
280 The *Xe* strains were collected in the southeastern U.S., but over a longer span of time than the *Xp*
281 strains (40). Homologous recombination is more likely to be detected in single population

bioRxiv preprint doi: https://doi.org/10.1101/180901; this version posted August 25, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

282 samples than in a timeseries (50). Recombination could be especially frequent in Florida
283 populations, or our inclusion of single strains representing diverged lineages may have caused an
284 underestimation of recombination in *Xp* as a whole. Additional strains from populations
285 respresenting the diverged lineages will be required to determine if homologous recombination
286 rates are high in other regions.

287 Both *Xp* and *Xe* have open and highly dynamic pangenomes, in contrast to the stability in
288 gene composition implied when *Xanthomonas* bacterial spot pathogens are characterized by
289 single reference genomes. Comparisons of pangenomes showed NI1 to be intermediate between
290 *Xe* and *Xp*, whereas core genomes placed it closer to *Xp*. Our data indicate that obtaining
291 pangenomes for *Xanthomonas* populations will be important in understanding the ongoing
292 evolution of these pathogens.

293 Horizontal gene transfer and recombination likely contribute to the adaptive divergence
294 that triggers the evolution of new bacterial spot lineages, similar to other bacterial systems such
295 as *Bartonella henselae* and *Clostridium difficile* ST6 (51-52). The observation of extensive
296 chromosomal replacement in NI1 was unexpected and mirrors findings in *Staphylococcus aureus*
297 in which multiple chromosomal replacements were found in new strains compared to the
298 reference strain, MRSA252 (52). The difference here is that NI1 was collected from a different
299 continent than the reference strains of *X. euvesicatoria* and *X. perforans*. Nevertheless, our
300 results for NI1 from Nigeria and 4P1S2 from Italy suggest that *X. perforans* may have very
301 dynamic populations across the globe.

302

303 ***Recombination shuffled secretion systems and effectors***

bioRxiv preprint doi: https://doi.org/10.1101/180901; this version posted August 25, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

304 Recombination has specifically affected secretion system genes and effector proteins in NI1 and
305 NI38. Generally, characterizations of secretion systems in bacterial species have focused on
306 reference genomes (35, 53-55), and secretions systems tend to be viewed as conserved elements
307 of pathogenic bacteria. We found that recombination changes the gene and allelic content of
308 secretion systems, resulting in intraspecific variation in secretion systems. The significance of
309 the homogenization of secretion systems and effectors from two monophyletic groups in novel
310 lineages remains to be experimentally tested. Subsequent surveys in Nigeria recovered strains
311 like NI1 and NI38, indicating that these genotypes were stable in the short term. New
312 combinations of effectors could be important in adaptation to hosts and/or environment, which
313 means that recombination potentially plays an important role in the evolution of ecological
314 interactions for these xanthomonads.

315 Recombination in type three secretion system genes explains why NI38 was initially
316 identified as an *Xp* strain based on qPCR primers targeting the highly conserved *hrcN* gene. The
317 *hrcN* allele in NI38 is distinct from the *Xp* reference strain and identical to that of 4P1S2; they
318 both possess a single nucleotide substitution that is not present in other *Xp* strains. The
319 nucleotide sequences where the diagnostic primers anneal in *hrcN* are identical for NI38 and all
320 *Xp* but differ by two nucleotides from other *Xe* (73). This is the first time that the primers have
321 not correctly differentiated *Xe* and *Xp* strains, and points to the limitations of pathogen
322 identification using a single primer set.

323 The mix of type 3 secreted effectors in NI1, NI38 and 4P1S2 raises concerns regarding
324 which effectors to use as targets for durable resistance breeding. One strategy being used to
325 determine candidate targets for resistance breeding is to identify core, conserved effectors.
326 Recently, comparison of sequenced genomes of *Xp* strains from Florida identified XopJ4 and

bioRxiv preprint doi: https://doi.org/10.1101/180901; this version posted August 25, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

327 AvrBST as putative stable targets for resistance breeding (39). The lack of the *xopJ4* gene in NI1
328 and *avrBST* in 4P1S2 makes this a less viable long-term strategy due to the possibility of
329 introduction of these strains into Florida. Our findings suggest that bacterial spot xanthomonads
330 may be evolving locally such that targets vary across tomato production regions, and indicate the
331 need for global studies of effector gene content to better understand the variation in putative
332 resistance breeding targets.
333

334 ***Recombination as a cohesive and diversifying force for X. euvesicatoria and X. perforans***

335 The close genetic distance but distinct lineages of *Xe* and *Xp* is relevant to the on-going debates
336 on what defines a bacterial species. As observed by (7), allopatry (or at least, microallopatry) is
337 not required for bacterial speciation, because genetic exchange rarely hinders genetic divergence
338 in bacterial populations. Rather, recombination fosters the acquisition of novel genes and
339 operons that aids adaptation and promotes divergence (7, 17). A competing hypothesis suggests
340 that recombination is a cohesive force that prevents bacterial diversification and maintains
341 lineages by homogenizing populations (56-57). Both processes may be occurring in these
342 *Xanthomonas* populations. Thus far, most of the lineages of *Xe* and *Xp* can easily be assigned to
343 species, while recombination is simultaneously driving the emergence of new lineages within
344 species. However, NI1 is a more extreme case of a nearly intermediate lineage and may indicate
345 the potential for homogenization of *Xe* and *Xp* through recombination. Both NI1 and NI38 were
346 isolated from the same field, which is surprising because *Xp* strains are known to outcompete *Xe*
347 strains on tomato under field conditions (58-59). NI38 strains may have acquired factors that
348 allow co-existence with *Xp* strains on tomato.

bioRxiv preprint doi: https://doi.org/10.1101/180901; this version posted August 25, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

349 There have been recent suggestions that both *Xe* and *Xp* be classified as a single species
350 or as pathovars of one species (36, 60). However, these pathogens have thus far shown
351 phenotypic differences in the lab and field, and the term pathovar implies differences in host
352 range that is not the case with *Xe* and *Xp*. The ecotype model, which defines species by their
353 ecology (7), may be a more appropriate concept for *Xe* and *Xp*. The ecotype model essentially
354 defines ecotypes by their ability to coexist on the same or similar ecological resources and persist
355 through periodic selection events. Additional studies of the genetic diversity of locally evolved
356 lineages could provide insight into factors that drive genetic variation in effectors and other
357 adaptive genes.
358
359 **Materials and Methods**
360 ***Sequencing of strains and calculation of average nucletide identities***
361 After extraction of genomic DNA (61), the Nextera library preparation kit (Illumina Inc., San
362 Diego, CA) was used to prepare genomic libraries for strains NI1 and NI38, which were
363 subsequently sequenced using Illumina MiSeq platform. NI1 was sequenced at Kansas State
364 University while NI38 was sequenced at the Interdisciplinary Center for Biotechnology
365 Research, University of Florida. Draft genomes were *de novo* assembled using CLC Genomics
366 Workbench v5, with average coverage of 259X for NI1 and 17X for NI38. The assembled
367 sequences were annotated using the IMG/JGI platform (62). The NCBI accession numbers for
368 NI1 and NI38 are respectively NISG00000000 and NJID00000000. Strain names and Genbank
369 accession numbers for additional previously sequenced genomes used for this study are provided
370 in Supplementary Table S1. The genomes of these 63 strains were previously published (63, 35-
371 36, 40-41; 43). Pairwise Average Nucleotide Identity (ANI) based on blast was calculated using
372 jSpecies v1.2.1 (64).

bioRxiv preprint doi: https://doi.org/10.1101/180901; this version posted August 25, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

373

374 ***Pan- and core genome analysis***

375 Pan- and core genome inferences were carried out using the GET_HOMOLOGUE program (65),
376 using default settings. All-against-all BLASTP, OrthoMCL as well as cluster of orthologous
377 gene (COG) clustering were initially carried out to generate pan and core gene clusters before
378 other functions within the program were used to analyze pan and core genomes (66-67). Re-
379 sampling of genomes was performed to estimate the sizes of core and pangenomes for *X.*
380 *euvesicatoria* and for *X. perforans* in the orders shown in Supplementary Table S2 (20,65).
381 Fitted exponential decay functions were applied to resampled genomes as described by (65) and
382 (68). A pangenome matrix (gene presence/absence) using a subset of *Xe* and *Xp* strains was
383 generated to compare the genomes of NI1 and NI38 strains to selected *Xp* and *Xe* strains. Using
384 the pangenome matrix, a dendrogram was constructed and a heatmap generated showing
385 Gower's distance using the *hcluster_matrix.sh* function in *get_homologues.pl*.

386

387 ***Alignment and phylogenetic inference using core genomes***

388 To infer the phylogeny using models for DNA sequence evolution, core genes were extracted
389 using GET_HOMOLOGUES and filtered using an in-house python script to select for single
390 copy orthologous genes. Alignments were carried out for each gene individually before
391 concatenation using MUSCLE (69). We used the concatenated core genome alignment to infer a
392 maximum likelihood phylogeny using iQTree, which was set for automatic selection of the best
393 model (using the function –m TEST) (70). Based on the Bayesian information criterion and the
394 corrected Akaike information criterion, the general time reversible model with proportion of
395 invariant sites (GTR+I) was utilized as the model with the best fit out of 88 total models

bioRxiv preprint doi: https://doi.org/10.1101/180901; this version posted August 25, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

396 compared. Branch support was assessed with ultrafast bootstrap and SH-aLRT test using 1000
397 replicates (70). The maximum likelihood tree was visualized using FigTree v1.4.3
398 (http://tree.bio.ed.ac.uk/software/figtree/).
399
400 ***Analysis of recombination***
401 We carried out analysis of recombination using ClonalFrameML on *Xp* and *Xe* strains separately
402 and together. ClonalFrameML models recombination as imports from external populations and
403 uses a Hidden Markov Model to estimate the influence of recombination on nucleotide variation
404 (52). For ClonalFrameML analyses, we used maximum likelihood trees generated from iQTree
405 and core genome alignments as infiles. Transition/transversion ratios were determined by
406 PhyML (71).
407
408 ***Genome comparisons: lipopolysaccharide clusters, effectors and secretion systems***
409 Lipopolysaccharide clusters (LPS), effectors and secretion systems were compared among
410 strains at both the nucleotide and amino acid levels. Genes and proteins orthologous to known
411 LPS clusters, effectors and secretion systems were identified by BLAST on the IMG platform
412 (https://img.jgi.doe.gov/), EDGAR (edgar.computational.bio.uni-giessen.de) and from local
413 databases (62,66,72). Secretion systems and effector allele assignments followed previous
414 studies of *Xe* and *Xp* (35, 40).
415
416 **Acknowledgements:** This work was funded in part by USDA NIFA award number 2015-51181-
417 24312. We thank Xavier Didelot for additional help in interpreting the results of

bioRxiv preprint doi: https://doi.org/10.1101/180901; this version posted August 25, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

418 ClonalFrameML and Jose Huguet Tapia for some python scripts used in implementing
419 get_homologues.
420

421 **References**

422 1. Bank, C., Burger, R., & Hermisson, J. (2012). The limits to parapatric speciation:
423 Dobzhansky-Muller incompatibilities in a continent-island model. Genetics 191(3): 845-
424 863.
425 2. Monahan, W., Pereira, R., & Wake, D. (2012) Ring distributions leading to species
426 formation: a global topographic analysis of geographic barriers associated with ring
427 species. BMC Biology, 10: 1-22.
428 3. Van Leuven, JT., Meister, RC.; Simon, C. and McCutcheon, J.P. (2014). Sympatric
429 speciation in a bacterial endosymbiont results in two genomes with the functionality of
430 one. *Cell* 158, 1270–1280.
431 4. Maynard Smith J., Smith N., O'Rourke M., Spratt B., (1993). How clonal are
432 bacteria? Proc. Natl. Acad. Sci. USA 90: 4384–4388
433 5. Guttman DS, Dykhuizen DE (1994). Clonal divergence in *Escherichia coli* as a result of
434 recombination, not mutation. Science 266: 1380–1383
435 6. Shapiro BJ. (2016) How clonal are bacteria over time? *Curr Opin Microbiol* (31): 116–
436 123.
437 7. Cohan FM. (2002). What are bacterial species? Ann. Rev.Microb. 56:457e87.
438 8. Lassalle, F., Campillo, T., Vial, L., Baude, J., Costechareyre, D., Chapulliot, D., Shams
439 M, Abrouk D, Lavire C, Oger-Desfeux C, Hommais F, Guéguen L, Daubin V, Muller D,

bioRxiv preprint doi: https://doi.org/10.1101/180901; this version posted August 25, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

440 Nesme, X. (2011). Genomic Species Are Ecological Species as Revealed by Comparative
441 Genomics in *Agrobacterium tumefaciens*. *Genome Biol Evol 3:* 762–781.
442 9. Sikorski J, Nevo E. Adaptation and incipient sympatric speciation of *Bacillus*
443 *simplex* under microclimatic contrast at "Evolution Canyons" I and II,
444 Israel. *Proceedings of the National Academy of Sciences of the United States of America*.
445 2005;102(44):15924-15929.
446 10. Smith NH, Kremer K, Inwald J, Dale J, Driscoll JR, Gordon SV, van Soolingen
447 D, Hewinson RG, Smith JM. 2006. Ecotypes of the Mycobacterium tuberculosis
448 complex. Journal of Theor. Biol.: 239: 220–225
449 11. Keymer DP and Boehm AB. 2011. Recombination Shapes the Structure of an
450 Environmental *Vibrio cholerae* Population. App.and Envir. Microbio. 77(2): 537-544.
451 12. Green S, Studholme DJ, Laue BE, Dorati F, Lovell H, Arnold D, Cottrell JE, Bridgett S,
452 Blaxter M, Huitema E., Thwaites R, Sharp PM, Jackson RW, Kamoun S. 2010.
453 Comparative genome analysis provides insights into the evolution and adaptation of
454 *Pseudomonas syringae* pv. *aesculi* on *Aesculus hippocastanum*. PLoS ONE 5: e10224.
455 13. Mazzaglia A, Studholme DJ, Taratufolo MC, Cai R, Almeida NF, Goodman T, Guttman
456 DS, Vinatzer BA, Balestra GM. 2012. *Pseudomonas syringae* pv. *actinidiae* (PSA)
457 isolates from recent bacterial canker of kiwifruit outbreaks belong to the same genetic
458 lineage. PLoS ONE 7: e36518.
459 14. Didelot X, and Maiden MCJ. 2010. Impact of recombination on bacterial
460 evolution. Trends in Microb. 18(7): 315-322.

bioRxiv preprint doi: https://doi.org/10.1101/180901; this version posted August 25, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

461 15. Chaguza C, Cornick JE, Everett DB. 2015. Mechanisms and impact of genetic
462 recombination in the evolution of *Streptococcus pneumoniae*. Comp. and Struct. Biotech.
463 Jour. 13: 241-247.
464 16. Hanage, W.P. 2016. Not So Simple After All: Bacteria, Their Population Genetics, and
465 Recombination. Cold Spring Harbor Perspective in Biology: 8: a018069.
466 17. Melendrez, MC., Becraft, ED., Wood, JM., Olsen, MT., Bryant, DA., Heidelberg, JF.,
467 Rusch,, DB, Cohan, FM and Ward, DM. 2015. Recombination Does Not Hinder
468 Formation or Detection of Ecological Species of *Synechococcus* Inhabiting a Hot Spring
469 Cyanobacterial Mat. *Frontiers in Microbiology*, *6*, 1540.
470 18. Ambur OH, Engelstadter J, Johnsen PJ, Miller EL, Rozen DE. 2016. Steady at the wheel:
471 conservative sex and the benefits of bacterial transformation. Phil. Trans. R. Soc. B 371:
472 0528.
473 19. Thomas CM and Nielsen KM. 2005. Mechanisms of, and barriers to, horizontal gene
474 transfer between bacteria. Nat. Rev. Microbiol. 3: 711–721.
475 20. Tettelin H, Masignani V, Cieslewicz MJ, Donati C, Medini D, Ward NL, Angiuoli SV,
476 Crabtree J, Jones AL, Durkin AS, Deboy RT, Davidsen TM, Mora M, Scarselli
477 M, Margarit y Ros I, Peterson JD, Hauser CR, Sundaram JP, Nelson WC, Madupu
478 R, Brinkac LM, Dodson RJ, Rosovitz MJ, Sullivan SA, Daugherty SC, Haft
479 DH, Selengut J, Gwinn ML, Zhou L, Zafar N, Khouri H, Radune D, Dimitrov G, Watkins
480 K, O'Connor KJ, Smith S, Utterback TR, White O, Rubens CE, Grandi G, Madoff
481 LC, Kasper DL, Telford JL, Wessels MR, Rappuoli R, Fraser CM. 2005. Genome
482 analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the
483 microbial "pan-genome". Proc Natl Acad Sci USA. 2005, 102: 13950-13955.

bioRxiv preprint doi: https://doi.org/10.1101/180901; this version posted August 25, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

484 21. Vos, M , Hesselman, M.C., Beek, T.A., van Passel, M.W.J., Eyre-Walker, A. 2015. Rates
485 of lateral gene transfer in prokaryotes: High but why? Trends Microbiol 23 (10): 598-
486 605.
487 22. Marttinen P, Hanage WP, Croucher NJ, Connor TR, Harris, SR, Bentley, SD, Corander,
488 J. 2012. Detection of recombination events in bacterial genomes from large population
489 samples. Nuc. Acid. Res. 40(1): e6.
490 23. Majewski J, and Cohan FM. (1999). DNA sequence similarity requirements for
491 interspecific recombination in Bacillus. Genetics 153: 1525–1533.
492 24. Everitt RG, Didelot X, Batty EM, Miller RR, Knox K, Young BC, Bowden R, Auton A,
493 Votintseva A, Larner-Svensson H, Charlesworth J, Golubchik T, Ip CL, Godwin H, Fung
494 R, Peto TE, Walker AS, Crook DW, Wilson DJ. 2014. Mobile elements drive
495 recombination hotspots in the core genome of *Staphylococcus aureus*. Nat. Comm. 5:
496 3956.
497 25. Dixit, P.D.; Pang, T.Y., Studier, F.W. and Maslov, S .2015. Recombinant transfer in the
498 basic genome of *Escherichia coli*. Proc. Nat. Acad. of Sci. 112(29): 9070-9075
499 26. Didelot X, Bowden R, Street T, Golubchik T, Spencer C, McVean G, Sangal, V., Anjum,
500 MF, Achtman, M., Falush D. and Donelly, P. 2011. Recombination and Population
501 Structure in *Salmonella enterica*. PLoS Genet 7(7): e1002191.
502 27. Driebe EM, Sahl JW, Roe C, Bowers JR, Schupp JM, Gillece JD, Kelley E, Price LB,
503 Pearson TR, Hepp CM, Brzoska PM, Cummings CA, Furtado MR, Andersen PS, Stegger
504 M, Engelthaler DM, Keim PS. 2015. Using Whole Genome Analysis to Examine
505 Recombination across Diverse Sequence Types of *Staphylococcus aureus*. *PLoS ONE*.
506 10(7): e0130955.

bioRxiv preprint doi: https://doi.org/10.1101/180901; this version posted August 25, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

507 28. Yahara, K., Didelot, X., Jolley, K.A., Kobayashi, I., Maiden, MCJ., Shepard, S.K. and
508 Falush, D. 2015. The landscape of realized homologous recombination in pathogenic
509 bacteria. Molecular Biology and Evolution 33(2): 456-471.
510 29. Gordon, J.L., Lefeuvre, P., Escalon, A., Barbe, V., Cruveiller, S., Gagnevin, L and
511 Pruvost, O. 2015. Comparative genomics of 43 strains of *Xanthomonas*
512 *citri* pv. *citri* reveals the evolutionary events giving rise to pathotypes with different host
513 ranges. BMC Genomics 16: 1098.
514 30. Huang CL, Pu PH, Huang HJ, Sung HM, Liaw HJ, Chen YM, Chen CM, Huang MB,
515 Osada N, Gojobori T, Pai TW, Chen YT, Hwang CC and Chiang TY. 2015. Ecological
516 genomics in Xanthomonas: the nature of genetic adaptation with homologous
517 recombination and host shifts. *BMC Genomics* 16: 188.
518 31. Monteil, CL., Yahara, K., Studholme, DJ., Mageiros, L., Méric, G., Swingle, B., Morris,
519 CB, Vinatzer, B, Sheppard, SK. 2016. Population-genomic insights into emergence, crop
520 adaptation and dissemination of Pseudomonas syringae pathogens. Microbial Genomics
521 2(10): e000089.
522 32. Ryan R. P., Vorhölter F. J., Potnis N., Jones J. B., Van Sluys M. A., Bogdanove A. J.,
523 Dow, J.M. 2011. Pathogenomics of *Xanthomonas*: Understanding bacterium-plant
524 interactions. Nat. Rev. Microbiol. 9, 344–355.
525 33. Jones, J. B., Stall, R., and Bouzar, H. 1998. Diversity among xanthomonads pathogenic
526 on pepper and tomato. *Annual Review of Phytopathology* 36:41-58.
527

bioRxiv preprint doi: https://doi.org/10.1101/180901; this version posted August 25, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

528 34. Jones, J.B., Lacy, G.H., Bouzar, H., Stall, R.E., and Schaad, N.W. 2004. Reclassification
529 of the xanthomonads associated with bacterial spot disease of tomato and pepper. Syst.
530 and App. Microb. 27: 755-762.
531 35. Potnis, N., Krasileva, K., Chow, V., Almeida, N, Patil, P., Ryan, R., Sharlach, M., Behlau
532 F, Dow JM, Momol M, White FF, Preston JF, Vinatzer BA, Koebnik R, Setubal
533 JC, Norman DJ, Staskawicz BJ, Jones, J.B. 2011. Comparative genomics reveals
534 diversity among xanthomonads infecting tomato and pepper. *BMC Genomics*12:146.
535 36. Constantin, E. C., Cleenwerck, I. Maes, M., Baeyen, S., Van Malderghem, C., De Vos, P.
536 and Cottyn, B. 2016. Genetic characterization of strains named as Xanthomonas
537 axonopodis pv. dieffenbachiae leads to a taxonomic revision of the X. axonopodis species
538 complex. Plant Pathology 65: 792–806.
539 37. Stall, R.E., Jones, J.B., and Minsavage, G.V. 2009. Durability of resistance in tomato and
540 pepper to xanthomonads causing bacterial spot. *Annual Review* of *Phytopathology*
541 47:265-284.
542 38. White FF, Potnis N, Jones JB and Koebnik R. 2009. The type III Effectors of
543 Xanthomonas. Mol Plant Pathol. 10(6): 749-66
544 39. Timilsina, S., Jibrin, M.O., Potnis, N., Minsavage, G.V., Kebede, M., Schwartz, A., Bart,
545 R., Staskawicz, B.J., Boyer, C., Vallad, G.E., Pruvost, O., Jones, J.B., and Goss,E.M.
546 2015. Multilocus sequence analysis of xanthomonads causing bacterial spot of tomato
547 and pepper plants reveals strains generated by recombination among species and recent
548 global spread of *Xanthomonas gardneri*. Appl Environ Microbiol 81(4): 1520-1529.
549 40. Schwartz, A.R., Potnis, N., Timilsina, S., Wilson, M., Patané, J., Martins Jr., J.,
550 Minsavage, G.V., Dahlbeck, D., Akhunova, A., Almeida, N., Vallad, G.E., Barak, J.D.,

bioRxiv preprint doi: https://doi.org/10.1101/180901; this version posted August 25, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

551 White, F.F., Miller, S.A., Ritchie, D., Goss, E.M., Bart, R.S., Setubal, J.C., Jones, J.B.,
552 and Staskawicz, B.J. 2015. Phylogenomics of Xanthomonas field strains infecting pepper
553 and tomato reveals diversity in effector repertoires and identifies determinants of host
554 specificity. Frontiers in Microbiology 6: 519.

555 41. Vancheva, T., Lefeuvre, P., Bogatzevska, N., Moncheva, P., & Koebnik, R. 2015. Draft
556 genome sequences of two *Xanthomonas euvesicatoria* strains from the Balkan
557 Peninsula. *Genome Announcements*, *3*(1), e01528–14.

558 42. Mbega, E. R., Mabagala, R. B., Adriko, J., Lund, O. S., Wulff, E. G., & Mortensen, C.
559 N. 2012. Five species of xanthomonads associated with bacterial leaf spot symptoms in
560 tomato from Tanzania. *Plant Disease 96:* 760.

561 43. Torelli E, Aiello D, Polizzi G, Firrao G, Cirvilleri G. 2015. Draft genome of a
562 *Xanthomonas perforans* strain associated with pith necrosis. FEMS Microbiol
563 Lett. 362(4).

564 44. Jibrin, M.O., Timilsina, S., Potnis, n., Minsavage, G.V., Shenge, K.C., Akpa, A.D.,
565 Alegbejo, M.D., Beed, F., Vallad, G.E., and Jones, J.B. 2015. First report of atypical
566 *Xanthomonas euvesicatoria* strains causing bacterial spot of tomato in Nigeria. Plant Dis.
567 99: 415.

568 45. Dharmapuri S, Yashitola J, Vishnupriya MR, Sonti RV. 2001. Novel genomic locus with
569 atypical G+C content that is required for extracellular polysaccharide production and
570 virulence in *Xanthomonas oryzae pv. oryzae*. Mol. Plant-Microbe Interact. 14: 1335–
571 1339

bioRxiv preprint doi: https://doi.org/10.1101/180901; this version posted August 25, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

572 46. Keshavarzi M, Soylu S, Brown I, Bonas U, Nicole M, Rossiter J, Mansfield, J. 2004.
573 Basal defenses induced in pepper by lipopolysaccharides are suppressed by *Xanthomonas*
574 *campestris* pv. *vesicatoria*. Mol Plant Microbe Int. 17: 805−815
575 47. Newman M.-A., Dow J. M., Molinaro A., Parrilli M. 2007. Priming, induction and
576 modulation of plant defense responses by bacterial lipopolysaccharides. J. Endotoxin
577 Res. 13: 68–79
578 48. Vorholter FJ, Niehaus K and Puhler A. 2001. Lipopolysacccharide biosynthesis in
579 Xanthomonas campestris pv. campestris: a cluster of 15 genes is involved in the
580 biosynthesis of the LPS O-antigen and the LPS core. Mol Genet Genomics 266 (1):79-
581 95.
582 49. Jaenicke, S., Bunk, B., Wibberg, D., Spröer, C., Hersemann, L., Blom, J., … Vorhölter,
583 F.-J. 2016. Complete Genome Sequence of the Barley Pathogen *Xanthomonas*
584 *translucens* pv. translucens DSM 18974^T (ATCC 19319^T). Genome Announc., 4(6):
585 e01334–16.
586 50. Martin DP, Lemey P, Posada D 2011. Analysing recombination in nucleotide sequences.
587 Mol. Eco. Res. 11: 943-955.
588 51. Guy, L., Nystedt, B., Sun, Y., Näslund, K., Berglund, E. C., & Andersson, S. G. 2012. A
589 genome-wide study of recombination rate variation in *Bartonella henselae*. BMC
590 Evolutionary Biology 12: 65.
591 52. Didelot X, and Wilson DJ 2015. ClonalFrameML: Efficient Inference of Recombination
592 in Whole Bacterial Genomes. PLoS Comput Biol 11 (2): e1004041.

bioRxiv preprint doi: https://doi.org/10.1101/180901; this version posted August 25, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

593 53. Vandroemme, J, Cottyn, B., Baeyen, S., Vos P.D. and Maes M. 2013. Draft genome
594 sequence of *Xanthomonas fragariae* reveals reductive evolution and distinct virulence-
595 related gene content. *BMC Genomics* 14:829.
596 54. Costa TR, Felisberto-Rodrigues C, Meir A, Prevost MS, Redzej A, Trokter M, Waksman
597 G. 2015. Secretion systems in Gram-negative bacteria: structural and mechanistic
598 insights. Nat. Rev. Microb. 13(6): 343-59.
599 55. Green, E. R., & Mecsas, J. 2016. Bacterial Secretion Systems – An overview.
600 Microb. Spect. 4(1): VMBF–0012–2015.
601 56. Fraser C, Hanage WP, Spratt BG. 2007. Recombination and the nature of bacterial
602 speciation. *Science* **315**: 476–480.
603
604 57. Doroghazi JR, Buckley DH. 2011. A model for the effect of homologous recombination
605 on microbial diversification. *Genome Biol Evol* **3**: 1349–1356.
606
607 58. Tudor-Nelson, S., Minsavage, G., Stall, R., and Jones, J. 2003. Bacteriocin-like
608 substances from tomato race 3 strains of *Xanthomonas campestris* pv. *vesicatoria*.
609 *Phytopathology* 93:1415-1421.
610 59. Jibrin MO, Timilsina S, Potnis N, Minsavage GV, Shenge KC, Akpa AD, Alegbejo MD,
611 Vallad GE, Jones JB. 2014. First report of *Xanthomonas euvesicatoria* causing bacterial
612 spot disease in pepper in northwestern Nigeria. Plant Dis. 98:10
613 60. Barak, Jeri D., Vancheva Taca, Lefeuvre Pierre, Jones Jeffrey B., Timilsina Sujan,
614 Minsavage Gerald V., Vallad Gary E., Koebnik Ralf. 2016. Whole-Genome Sequences of

bioRxiv preprint doi: https://doi.org/10.1101/180901; this version posted August 25, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

615 Xanthomonas euvesicatoria Strains Clarify Taxonomy and Reveal a Stepwise Erosion of
616 Type 3 Effectors. Frontiers in Plant Science 7:1805
617 61. Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A., and
618 Struhl, K. 1994. Current protocols in molecular biology. New York: John Wiley and
619 Sons.
620 62. Markowitz VM, Chen IM, Chu K, Szeto E, Palaniappan K, Pillay M, Ratner A, Huang
621 J, Pagani I, Tringe S, Huntemann M, Billis K, Varghese N, Tennessen K, Mavromatis
622 K, Pati A, Ivanova NN, Kyrpides NC 2014. IMG/M 4 version of the integrated
623 metagenome comparative analysis system. Nucleic Acids Res 42: D568-73.
624 63. Thieme, F., Koebnik, R., Bekel, T., Berger, C., Boch, J., Büttner, D., Caldana C, Gaigalat
625 L, Goesmann A, Kay S, Kirchner O, Lanz C, Linke B, McHardy AC, Meyer
626 F, Mittenhuber G, Nies DH, Niesbach-Klösgen U, Patschkowski T, Rückert C, Rupp
627 O, Schneiker S, Schuster SC, Vorhölter FJ, Weber E, Pühler A, Bonas U, Bartels
628 D, Kaiser O.. 2005. Insights into genome plasticity and pathogenicity of the plant
629 pathogenic bacterium *Xanthomonas campestris* pv. *vesicatoria* revealed by the complete
630 genome sequence. *J. Bacteriol*. 187: 7254–7266.
631 64. Richter, M., & Rosselló-Móra, R. 2009. Shifting the genomic gold standard for the
632 prokaryotic species definition. *Proc Natll Acad Sci U S A*, *106*(45), 19126–19131.
633 65. Contreras-Moreira,B. and Vinuesa,P. 2013. GET HOMOLOGUES, a versatile software
634 package for scalable and robust microbial pangenome analysis. Appl.Environ.Microbiol.
635 79: 7696-7701.

bioRxiv preprint doi: https://doi.org/10.1101/180901; this version posted August 25, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

636 66. Altschul SF, Madden TL, Schaer AA, Zhang J, Zhang Z, Miller W and Lipman DJ
637 (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search
638 programs. Nucl. Acids Res. 25(17): 3389-3402.
639 67. Li L, Stoeckert CJ Jr, Roos DS. 2003. OrthoMCL: identification of ortholog groups for
640 eukaryotic genomes. Genome Res. 13(9): 2178-89.
641 68. Willenbrock H, Hallin PF, Wassenaar TM, Ussery DW. 2007. Characterization of
642 probiotic *Escherichia coli* isolates with a novel pan-genome microarray. Genome
643 Biology. 8(12): R267.
644 69. Edgar, Robert C. 2004. ***MUSCLE***: multiple sequence *alignment* with high accuracy and
645 high throughput. Nuc. Acid. Res. 32(5): 1792-97.
646 70. Nguyen, L.-T., Schmidt, H. A., von Haeseler, A., and Minh, B. Q. 2015. IQ-TREE: a fast
647 and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol.
648 Biol. Evol. 32: 268–274.
649 71. Guindon, S. and Gascuel, O. 2003. A simple, fast and accurate algorithm to estimate
650 large phylogenies by maximum likelihood. System Biology (52): 696-704.
651 72. Blom, J., Kreis, J., Spänig, S., Juhre, T., Bertelli, C., Ernst, C., & Goesmann, A. 2016.
652 EDGAR 2.0: an enhanced software platform for comparative gene content analyses. Nuc.
653 Acid. Res. 44: W22–W28.
654 73. Strayer, A.L., Jeyaprakash, A., Minsavage, G.V., Timilsina, S., Vallad, G.E., Jones, J.B.
655 & Paret, M.L. 2016. A multiplex real-time PCR assay differentiates four *Xanthomonas*
656 species associated with bacterial spot of tomato. Plant Dis. 100: 1660-1668.
657

bioRxiv preprint doi: https://doi.org/10.1101/180901; this version posted August 25, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

658 **Figure Legends**

659

660 **Figure 1**. Pangenome comparisons based on a combined subset of *Xp* and *Xe* genomes. Values
661 used for heatmap and dendrogram are distances between genomes based on Gower's coefficient.

662

663 **Figure 2.** Maximum likelihood phylogeny based on concatenated core genes from of 65
664 genomes of *X. euvesicatoria* and *X. perforans* with midpoint rooting. Branch labels are bootstrap
665 support values. Scale bar is in substitutions per site.

666

667 **Figure 3.** Homologous recombination in the core genomes of (A) *X. euvesicatoria* and (B) *X.*
668 *perforans* estimated by ClonalFrameML. For all branches of the genealogy and positions in the
669 genome, dark blue regions represent recombination and light blue represent invariant sites.
670 Polymorphic sites are shown in a color indicating their level of homoplasy: white indicates no
671 homoplasy and the range from yellow to red represent increasing degrees of homoplasy (52).
672 The corresponding analysis of all strains together is shown in Supplementary Figure S2.

673

674 **Figure 4**. Variation in LPS clusters. Genes are colored in blocked arrows, with similar colors
675 indicating genes belonging to the same COG (cluster of orthologous genes). a) Comparison of
676 NI1, NI38, and reference strains for *Xtt*, *Xe* and *Xp*. b. Comparison of representative LPS clusters
677 for *Xp* groups 1A (GEV936 and GEV1026), 1B (Xp5-6, Xp91-118), 2 (GEV1001) and the
678 Italian strain 4P1S2. Orange arrows indicate position of *etfA* genes across the compared sets.

679

bioRxiv preprint doi: https://doi.org/10.1101/180901; this version posted August 25, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

680 **Figure 5.** Reactions of a differential line carrying the *RXopJ4* gene after inoculation with a
681 Florida *Xp* strain and strain NI1. The *Xp* strain with *avrXv4* elicited a hypersensitive response
682 (left leaflet), while NI1 without *avrXv4* did not (right leaflet).
683

bioRxiv preprint doi: https://doi.org/10.1101/180901; this version posted August 25, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

684 **Table 1**. Average nucleotide identities (ANIb) of genomes of Nigerian, *X. perforans* and *X.*
685 *euvesicatoria* strains as compared to reference genomes. For all pairwise comparisons, see
686 Supplementary File 1.

Species/Groups	ANIb vs. Xp91-118	ANIb vs. Xe85-10
Nigerian strains		
NI1	98.92	98.58
NI38	98.48	99.79
X. perforans		
Xp91-118	100	
Group 1A strains	99.87-99.90	98.47-98.50
Group 1B strains	99.70-99.86	98.36-98.55
Group 2 strains	99.67-99.78	98.44-98.62
Xp17-12		
Xp4P1S2	99.59	98.44
X. euvesicatoria		
Xe85-10	98.48	100
Other *Xe* strains	98.32-98.44	99.66-99.95

687

bioRxiv preprint doi: https://doi.org/10.1101/180901; this version posted August 25, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

688 **Table 2.** Rates of recombination as calculated using ClonalFrameML[1].

Xanthomonas population	R/θ	δ	ν	R/m	Tr/Tv
All strains	0.49	292	0.03	3.82	5.36
All *Xp* (+NI1)	0.85	394	0.02	8.47	4.92
Florida *Xp*[#]	1.54	696	0.02	26.23	3.47
All *Xe*	0.43	409	0.028	5.06	5.35
Xe w/out NI38	0.36	731	0.015	4.02	4.02

689 [1]Values were rounded for presentation. R/θ is the rate of recombination to mutation; δ is the size
690 of the recombined fragment in bp; ν is the probability of recombination per site; R/m is the
691 relative effect of recombination to mutation (or ratio of per base pair substitution by
692 recombination to mutation, R/m= R/θ x ν x δ); Tr/Tv is the transition/transversion ratio used as
693 input in the analysis.

694 [#] NI1 (Nigeria) and 4P1S2 (Italy) were excluded.

bioRxiv preprint doi: https://doi.org/10.1101/180901; this version posted August 25, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

695 **Table 3**. Allelic assignments of NI1, NI38 and 4P1S2 Type 3 Secretion System T3SS genes
696 relative to *Xe* and *Xp* reference strains. Allele types are based on nucleotide sequence identity.
697 'Identical' indicates that *Xe* and *Xp* alleles were identical to each other, thus cannot be
698 distinguished. 'Unique' indicates alleles not found in the reference strains. Allelic assignments
699 based amino sequence identity are shown in Supplementary Table S3, because in some cases the
700 nucleotide variation shown does not lead to variation in the protein sequence.

Secretion system Genes	NI1 Allele Type	NI38 Allele Type	4P1S2 Allele Type
Hrc (Hypersensitivity, resistance and conserved) genes			
HrcA	**Unique**	Xe	Xp
HrcC	Xp	**Unique**	Xp
HrcT	Identical	Identical	Identical
HrcN	Xp	**Unique**	**Unique**
YscJ/HrcJ	Identical	Identical	**Unique**
HrcU	Identical	Identical	Identical
HrcV	Xp	**Unique**	**Unique**
YscQ/HrcQ	Xp	Xe	**Xe**
HrcR	**Unique**	Xe	**Xe**
HrcS	Identical	Xe	**Xe**
HrcD	**Unique**	Xe	Xp
Hrp (Hypersensitivity, resistance and pathogenicity) genes			
HrpA	Xp	Xe	**Unique**
HrpB1	Identical	Identical	**Unique**
HrpB2	Identical	Identical	Identical
HrpB4	Xp	**Xp**	Xp

bioRxiv preprint doi: https://doi.org/10.1101/180901; this version posted August 25, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

HrpB5	**Unique**	Identical	Identical
HrpB7	Identical	Identical	Xp
HrpD6	Xp	**Xp**	Xp
HrpE	**Unique**	**Unique**	**Xe**
*HrpF**	Xp*	Xe	Xp
*HrpG**	Xp*	Xe	**Unique**
HrpX	Xp	Xe	**Unique**
Hpa (Hrp Associated)			
Hpa1	Xp	**Xp**	Xp
Hpa2	Xp	**Xp**	Xp
HpaA	Xp	Xe	**Xe**
HpaB	**Unique**	Xe	**Unique**
HpaG	Xp	Xe	Xp
HpaP	Xp	**Unique**	Xp

701 **HrpG/F* are fusion proteins in *Xp*.

bioRxiv preprint doi: https://doi.org/10.1101/180901; this version posted August 25, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

702 **Table 4.** Effector profiles of NI1 and NI38 for species-specific effectors within the bacterial spot
703 xanthomonads.

Effectors*	Strains	
	NI1	**NI38**
Found uniquely in *Xe*		
AvrBs1	Absent*	Absent
XopC1	Absent	Ga0128168_10566
XopJ1	Absent	Ga0128168_10111
XopJ3 (*AvrRxv*)	Absent	Ga0128168_12162
XopO	Absent	Ga0128168_12403
XopAA	Absent	Ga0128168_11918
XopAJ	Ga0071335_10438	Ga0128168_10823
Found in *Xe*, *Xp* and *Xv*		
AvrBsT (YopJ)	Ga0071335_1442[U]	Absent
Found uniquely in *Xp*		
XopC2	Ga0071335_101102	Ga0128168_101017, Ga0128168_101018
XopJ4	Absent	Absent
XopAE	Ga0071335_113188	Ga0128168_11984, Ga0128168_11985
XopAF(*AvrXv3)*	Ga0071335_13475	Absent

bioRxiv preprint doi: https://doi.org/10.1101/180901; this version posted August 25, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Found uniquely in *Xg*		
AvrBs1	Absent	Absent
AvrHah1	Ga0071335_1551/Ga0071335_1571	Absent
XopAO	Absent	Absent
XopAQ	Ga0071335_1192^{U}	Absent
XopAS	Absent	Absent
Found in *Xe*/*Xp* but absent in *Xv*/*Xg*		
XopE1	Ga0071335_10358^{U}	Ga0128168_10456^{Xe}, Ga0128168_109715^{U}
XopF2	Absent	Ga0128168_103921^{U}
XopI	Ga0071335_12619^{Xe}	Ga0128168_10765^{Xe}
XopP	Ga0071335_101103^{Xp} Ga0071335_13041^{Xp} Ga0071335_1451^{U} Ga0071335_1461^{U}	Ga0128168_101020^{U} Ga0128168_101021^{Xe}
XopV	Ga0071335_13449^{U}	Ga0128168_100834^{NI1}
XopAK	Ga0071335_111115^{U}	Ga0128168_11919^{Xe}
XopAP	Ga0071335_107226^{U}	Ga0128168_100746^{Xe}
Found in *Xg*/*Xe* but		

bioRxiv preprint doi: https://doi.org/10.1101/180901; this version posted August 25, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

absent in *Xp/Xv*		
XopB	Absent	Ga0128168_114314

704 *Effectors unique to other species combinations, such as Xg/Xv, which were not found in NI1 or
705 NI38 are not shown.
706 [*Xe, Xp, U, NI1*] Superscripts indicate the allele types (U–unique). For effectors present only in one
707 species, the allele type is that of the species unless otherwise indicated.