TITLE

The Origin and Evolution of a Pandemic Lineage of the Kiwifruit Pathogen *Pseudomonas syringae* pv. *actinidiae*

AUTHORS

Honour C. McCann[a,c,*,#], Li Li[b,*], Yifei Liu[c], Dawei Li[b], Pan Hui[b], Canhong Zhong[b], Erik Rikkerink[d], Matthew Templeton[d,e], Christina Straub[a], Elena Colombi[a], Paul B. Rainey[a,f,g,†] & Hongwen Huang[b,c,†,#]

AFFILIATIONS

[a] New Zealand Institute for Advanced Study, Massey University, Private Bag 102904, Auckland 0745, New Zealand,.

[b] Key Laboratory of Plant Germplasm Enhancement and Specialty Agriculture, Wuhan Botanical Garden, Chinese Academy of Sciences, Wuhan 430074, China

[c] Key Laboratory of Plant Resources Conservation and Sustainable Utilization, South China Botanical Garden, Chinese Academy of Sciences, Guangzhou 510650, China

[d] New Zealand Institute for Plant and Food Research, 120 Mt Albert Road, Auckland 1025, New Zealand

[e] School of Biological Sciences, University of Auckland, Private Bag 92-019, Auckland 1142, New Zealand

[f] Max Planck Institute for Evolutionary Biology, August-Thienemann-Str. 2, Plön 24306, Germany

[g] Ecole Supérieure de Physique et de Chimie Industrielles de la Ville de Paris (ESPCI ParisTech), CNRS UMR 8231, PSL Research University, 75231 Paris Cedex 05, France

[*] Co-first authors

[†] Co-senior authors

Corresponding authors: h.mccann@massey.ac.nz, tel.: +64 94140800; huanghw@scbg.ac.cn., tel.: +86 020 37252778.

ABSTRACT

Recurring epidemics of kiwifruit (*Actinidia* sp.) bleeding canker disease are caused by *Pseudomonas syringae* pv. *actinidiae* (*Psa*), whose emergence has coincided with domestication of its host. The most recent pandemic has had a deleterious effect on kiwifruit production worldwide. In order to strengthen understanding of population structure, phylogeography and evolutionary dynamics of *Psa*, we isolated *Psa* from cultivated kiwifruit across six provinces in China. No *Psa* was isolated from wild *Actinidia* spp. Genome sequencing of fifty isolates and the inclusion of an additional thirty from previous studies show that China is the origin of the recently emerged pandemic lineage. However China harbours only a fraction of global *Psa* diversity, with greatest diversity found in Korea and Japan. Distinct transmission events were responsible for introduction of the pandemic lineage of *Psa* into New Zealand, Chile and Europe. Two independent transmission events occurred between China and Korea, and two Japanese isolates from 2014 cluster with New Zealand *Psa*. Despite high

similarity at the level of the core genome and negligible impact of within-lineage recombination, there has been substantial gene gain and loss even within the single clade from which the global pandemic arose.

SIGNIFICANCE STATEMENT

Bleeding canker disease of kiwifruit caused by *Pseudomonas syringae* pv. *actinidiae* (*Psa*) has come to prominence in the last three decades. Emergence has coincided with domestication of the host plant and provides a rare opportunity to understand ecological and genetic factors affecting the evolutionary origins of *Psa*. Here, based on genomic analysis of an extensive set of strains sampled from China and augmented by isolates from a global sample, we show, contrary to earlier predictions, that China is not the native home of the pathogen, but is nonetheless the source of the recent global pandemic. Our data identify specific transmission events, substantial genetic diversity and point to non-agricultural plants in either Japan or Korea as home to the source population.

66 **INTRODUCTION**

67 A pandemic of kiwifruit (*Actinidia* spp.) bleeding canker disease caused
68 by *Pseudomonas syringae* pv. *actinidiae* (*Psa*) emerged in 2008 with severe
69 consequences for production in Europe, Asia, New Zealand and Chile (1-7).
70 Earlier disease epidemics in China, South Korea and Japan had a regional impact,
71 however, as infections were often lethal and the pathogen rapidly disseminated,
72 it was predicted to pose a major threat to global kiwifruit production(8, 9).
73 Despite recognition of this threat – one subsequently realized in 2008 – little was
74 done to advance understanding of population structure, particularly across
75 regions of eastern Asia that mark the native home of the genus Actinidia.

76 The origins of agricultural diseases and their link with plant
77 domestication is shrouded by time, as most plant domestication events occurred
78 millennia ago. Kiwifruit (*Actinidia* spp.) is a rare exception because
79 domestication occurred during the last century (10, 11). Kiwifruit production
80 and trade in plant material for commercial and breeding purposes has recently
81 increased in Asia, Europe, New Zealand and Chile(12-16), preceding the
82 emergence of disease in some cases by less than a decade.

83 The first reports of a destructive bacterial canker disease in green-fleshed
84 kiwifruit (*A. chinensis var. deliciosa*) occurred in Shizuoka, Japan (17, 18). The
85 causal agent was described as *Pseudomonas syringae* pv. *actinidiae* (*Psa*) (18). An
86 outbreak of disease with symptoms similar to those produced by *Psa* was
87 reported to have occurred in 1983-1984 in Hunan, China, though no positive
88 identification was made or isolates stored at that time (17). *Psa* was also isolated
89 from infected green kiwifruit in Korea shortly thereafter (19). The cultivation of
90 more recently developed gold-fruiting cultivars derived from *A. chinensis var.*

91 *chinensis* (e.g. 'Hort16A') began only in the 2000s and an outbreak of global
92 proportions soon followed. The first published notices of the latest outbreak on
93 gold kiwifruit issued from Italy in 2008, with reports from neighbouring
94 European countries, New Zealand, Asia and Chile occurring soon after (1-6, 20).
95 Whole genome sequencing showed the most recent global outbreak of disease
96 was caused by a new lineage of *Psa* (previously referred to as *Psa*-V and now
97 referred to as *Psa*-3), while earlier disease incidents in Japan and Korea were
98 caused by strains forming separate clades referred to *Psa*-1 (previously *Psa*-J)
99 and *Psa*-2 (previously *Psa*-K), respectively) (21-24). These lineages are marked
100 by substantial variation in their complement of type III secreted effectors, which
101 are required for virulence in *P. syringae.* Despite the surprising level of within-
102 pathovar differences in virulence gene repertoires occurring subsequent to the
103 divergence of these three lineages, strains from each lineage are capable of
104 infecting and growing to high levels in both *A. chinensis var. deliciosa* and *A.*
105 *chinensis var. chinensis* (23).
106 The severity of the latest global outbreak is largely predicated on the
107 expansion in cultivation of clonally propagated highly susceptible *A. chinensis*
108 *var. chinensis* cultivars, with trade in plant material and pollen likely providing
109 opportunities for transmission between distant geographic regions. Identifying
110 the source from which *Psa* emerged to cause separate outbreaks remains an
111 important question. Intriguingly, despite the divergence in both the core and
112 flexible genome, these distinct lineages nevertheless exhibit evidence of
113 recombination with each other and unknown donors (23). This suggests each
114 lineage of *Psa* emerged from a recombining source population. Definitive
115 evidence for the location, extent of diversity and evolutionary processes

116 operating within this population remain elusive. Early reports suggested China
117 may be the source of the latest global outbreak (22, 23). Although the strains of
118 *Psa* available at that time did not provide unambiguous and well-supported
119 evidence of a Chinese origin, this speculation was based on the fact that kiwifruit
120 are native to China; it is the provenance of the plant material selected for
121 commercial and breeding purposes between China, New Zealand, Italy and other
122 kiwifruit growing regions; there is extensive trade in plant material between all
123 of these regions; and one Chinese isolate was found to carry an integrative and
124 conjugative element (ICE) that was also found in New Zealand *Psa*-3 isolates
125 (22).

126 In order to strengthen understanding of the population structure,
127 phylogeography and evolutionary dynamics of *Psa*, we isolated *Psa* from
128 cultivated kiwifruit across six provinces in China and obtained additional isolates
129 from South Korea and Japan. Genome sequencing of fifty isolates and the
130 inclusion of an additional thirty previously sequenced isolates show that while
131 China is the origin of the pandemic lineage of *Psa,* only a single clade is currently
132 present in China, while strains from multiple clades are present in both Korea
133 and Japan. Strains from the pandemic lineage are closely related and display
134 reduced pairwise nucleotide diversity relative to other lineages, indicating a
135 more recent origin. Distinct transmission events were responsible for the
136 introduction of the pandemic lineage of *Psa* into New Zealand, Chile and Europe.
137 Two independent transmission events occurred between China and Korea, and
138 two Japanese isolates from 2014 cluster with New Zealand *Psa*. Despite high
139 similarity at the level of the core genome and negligible impact of within-lineage

140 recombination, there has been substantial gene gain and loss even within the
141 single clade from which the global pandemic arose.

142

143 **RESULTS**

144 **The phylogeography of *Psa***

145 The genomes of 50 *P. syringae* pv. *actinidiae* (*Psa*) isolated from
146 symptomatic kiwifruit in China, Korea and New Zealand between 2010 and 2015
147 were sequenced (Table S1). Combined with 30 *Psa* genomes from earlier
148 outbreaks and different geographic regions (e.g. Italy and Chile), our samples
149 represent the main *Psa* genotypes from the countries producing 90% of kiwifruit
150 production worldwide. The completed reference genome of *Psa* NZ13 (ICMP
151 18884) comprises a 6,580,291bp chromosome and 74,423bp plasmid(23, 25).
152 Read mapping and variant calling with reference to *Psa* NZ13 chromosome
153 produced a 1,059,732bp non-recombinant core genome for all 80 genomes,
154 including 2,963 nonrecombinant SNPs. A maximum likelihood phylogenetic
155 analysis showed the four lineages of *Psa* known to cause bleeding canker disease
156 in kiwifruit were represented among the 80 strains (Figure S1). The first clade
157 (*Psa*-1) includes the pathotype strain of *Psa* isolated and described during the
158 first recorded epidemic of bleeding canker disease in Japan (1984-1988). The
159 second clade (*Psa*-2) includes isolates from an epidemic in South Korea (1997-
160 1998), and the third clade (*Psa*-3) includes isolates that define the global
161 pandemic lineage (2008-present). A fourth clade (*Psa*-5) is represented by a
162 single strain, as no additional sequences or isolates were available(26). The
163 average between and within-clade pairwise identity is 98.93% and 99.73%,

164 respectively (Table S2). All *Psa* isolated from kiwifruit across six different
165 provinces in China group are members of the same clade: *Psa*-3. A subset of
166 Chinese strains group with the *Psa* isolated during the global outbreak in Italy,
167 Portugal, New Zealand, and Chile. This subset is referred to as the pandemic
168 lineage of *Psa*-3.

169 In order to obtain greater resolution of the relationships between the new
170 Chinese and pandemic isolates, we identified the 4,853,421bp core genome of all
171 62 strains in *Psa*-3. The core genome includes both variant and invariant sites
172 and excludes regions either unique to or deleted from one or more strains. To
173 minimize the possibility of recombination affecting the reconstruction of
174 evolutionary relationships and genetic distance within *Psa*-3, ClonalFrameML
175 was employed to identify and remove SNPs with a high probability of being
176 introduced by recombination rather than mutation. The within-lineage ratio of
177 recombination to mutation (R/theta) is reduced in *Psa*-3 ($6.75 \times 10^{-2} \pm 3.24 \times 10^{-}$
178 ^{5}) relative to between lineage rates ($1.27 \pm 5.16 \times 10^{-4}$) ,and the mean divergence
179 of imported DNA within *Psa*-3 is $8.54 \times 10^{-3} \pm 5.18 \times 10^{-7}$ compared to 5.68×10^{-3}
180 $\pm 1.04 \times 10^{-8}$ between lineages. Although recombination has occurred within *Psa*
181 lineage 3, it is less frequent and has introduced fewer polymorphisms relative to
182 mutation: when accounting for polymorphisms present in recombinant regions
183 identified by ClonalFrameML and/or present on transposons, plasmids, and
184 other mobile elements, more than seven-fold more polymorphisms were
185 introduced by mutation relative to recombination (Table 1). Recombination has
186 a more pronounced impact between lineages, where substitutions are slightly
187 more likely to have been introduced by recombination than by mutation (Table
188 1).

189

190 **The source of pandemic *Psa***

191 Data show that there is greater diversity among the Chinese *Psa*-3
192 population than had been previously identified (Figure 1). Interestingly, clades
193 defining *Psa*-1 and *Psa*-3 exhibit similar levels of diversity (Table S2). These
194 clades share a common ancestor: assuming they are evolving at a similar rate,
195 they may have been present in Japan and China for a similar duration. The
196 strains isolated during the latest pandemic in Italy (I2, I3, I10, I11, I13), Portugal
197 (P1), New Zealand (NZ13, NZ31-35, NZ37-43, NZ45-49, NZ54), Chile (CL4), Japan
198 (J38, J39) and Korea (K5) during the latest kiwifruit canker pandemic cluster
199 with nine Chinese isolates (C1, C3, C29-31, C62, C67-69) (Figure 1). This
200 pandemic lineage exhibits little diversity at the level of the core genome, having
201 undergone clonal expansion only very recently. The NZ isolates form a
202 monophyletic group and share a common ancestor, indicating there was a single
203 transmission event of *Psa* into NZ. Strikingly, two recently isolated Japanese
204 pandemic *Psa*-3 isolated in 2014 group within the New Zealand isolates,
205 suggesting the pandemic lineage may have been introduced into Japan via New
206 Zealand (Figure 1). Italian and Portuguese pandemic strains also form a separate
207 group, indicative of a single transmission event from China to Italy. China is
208 undoubtedly the source of the strains responsible for the pandemic of kiwifruit
209 canker disease, yet the precise origins of the pandemic subclade remain unclear.
210 Isolates from four different provinces in Western China (Guizhou, Shaanxi,
211 Sichuan and Chongqing) are represented among the pandemic lineage, indicating
212 extensive regional transmission within China after emergence of the pandemic.
213 Yet each province harbouring pandemic isolates also harbors basally diverging

214 *Psa*-3 isolates (Figure 2). With the exception of a group of isolates from Sichuan,
215 there is no phylogeographic signal among the more divergent Chinese strains.
216 This suggests there was extensive regional transmission of *Psa* both prior and
217 subsequent to the emergence of the pandemic subclade in China. Korea harbors
218 both divergent and pandemic subclade *Psa*-3 strains. K5 groups with the Chilean
219 *Psa*-3 strain in the pandemic subclade, while K7 groups with the more divergent
220 Chinese isolates indicating that a transmission event from strains outside the
221 pandemic subclade may have occurred. This pool of diversity therefore
222 represents a reservoir from which novel strains are likely to emerge in the
223 future.

224 The reduced level of diversity within the core genome of pandemic *Psa*-3
225 demonstrates these strains have been circulating for a shorter period of time
226 relative to those responsible for earlier outbreaks in both Japan and Korea. In
227 order to estimate the divergence time of the pandemic lineages as well as the age
228 of the most recent common ancestor of all *Psa* clades displaying vascular
229 pathogenicity on kiwifruit, we performed linear regression of root-to-tip
230 distances against sampling dates using the RAxML phylogenies determined from
231 the non-recombinant core genome of all clades and of Psa-3 alone. No temporal
232 signal was identified in the data. There were poor correlations between
233 substitution accumulation and sampling dates, indicating the sampling period
234 may have been too short for sufficient substitutions to occur. There may also be
235 variation in the substitution rate within even a single lineage. Forty-four unique
236 non-recombinant SNPs were identified among the 21 pandemic *Psa*-3 genomes
237 sampled over five years in New Zealand (an average of 2.10 per genome) over
238 five years) producing an estimated rate of 8.7×10^{-8} substitutions per site per

239 year. The relatively slow substitution rate and the strong bottleneck effect
240 experienced during infections hinders efforts to reconstruct patterns of
241 transmission, as the global dissemination of a pandemic strain may occur
242 extremely rapidly {ref}. The estimated divergence time of *Psa* broadly
243 considered is likely older than the pandemic and epidemic events with which
244 they are associated: the earliest report of disease cause by lineage 1 occurred in
245 1984 and the first report of infection from the latest pandemic was issued in
246 2008 (see also (27)).
247

248 **Diversification and parallelism among *Psa*-3 isolates**

249 2,214 SNPs mapping to the core genome of *Psa*-3 were identified; 263 of
250 these mapped to recombinant regions identified by ClonalFrameML and/or
251 plasmid, prophage, integrative and conjugative elements, transposons and other
252 mobile genetic elements (Table 1, Figure 3). Of the 1,951 SNPs mapping to the
253 non-recombinant non-mobile core genome, 57.7% (1,125) are strain specific.
254 Most strain-specific SNPs are found in the two most divergent members of the
255 lineage: *Psa* C16 and C17, with 736 and 158 strain-specific SNPs, respectively.
256 The remaining isolates have an average of 3.9 strain-specific SNPs, ranging from
257 0 to 44 SNPs per strains (Figure 3). There are 826 SNPs shared between one or
258 more *Psa*-3 strains. The pandemic clade differs from the more divergent Chinese
259 strains by 72 shared SNPs. Within the pandemic lineage there are 125 strain-
260 specific SNPs, an average of 3.1 unique SNPs per strain (ranging from 0-27 SNPs)
261 and an additional 29 SNPs shared among pandemic strains. Protein-coding
262 sequence accounts for 88.4% of the non-recombinant, gap-free core genome of
263 this lineage. We observed that 78.9% (1,539/1,951) of mutations occurred in

264 protein coding sequence, significantly different from the expectation
265 (1,725/1,963) in the absence of selection (Pearson's χ^2 test: $P < 0.0001$, $\chi^2 =$
266 172.55). This suggests there is selection against mutations occurring in protein
267 coding sequences. Of the 953 non-synonymous substitutions introduced by
268 mutation in lineage 3, 927 resulted in amino acid substitutions, two resulted in
269 extensions and 24 resulted in premature truncations.

270 Multiple synonymous and non-synonymous mutations were identified in
271 271 genes. The accumulation of multiple independent mutations in the same
272 gene may be a function of gene length, mutational hotspots or directional
273 selection. We identified 29 genes exhibiting a greater than expected number of
274 independent mutations in the same gene, given the length of the gene and total
275 number of mutations affecting core genes (Table S3). A range of hypothetical
276 proteins, transcriptional regulators and transferases acquired between two and
277 four independent mutations. The fitness impact of these mutations – and the 38
278 amino-acid changing mutations in the ancestor of the pandemic subclade - is
279 unknown, yet it is possible these patterns are the outcome of selective pressures
280 imposed during bacterial residence within a similar host niche.

281 Two substitutions are shared exclusively by the European pandemic
282 strains (AKT28710.1 G1150Aand AKT33438.1 T651C) and one silent
283 substitution in a gene encoding an acyltransferase superfamily protein
284 (AKT31915.1 C273T) is shared among the European pandemic and six of nine
285 Chinese pandemic strains (C3, C29-31, C67, C69). As these six Chinese pandemic
286 strains were isolated from Shaanxi, Sichuan and Chongqing, they do not provide
287 any insight into the precise geographic origins of the European pandemic *Psa-3*,
288 though transmission from China to Italy is likely concomitant with dissemination

289 of the pandemic lineage across China. Six conserved and diagnostic
290 polymorphisms are present in the pandemic New Zealand and Japanese isolates
291 (Table S4). One of these is a silent substitution in an ion channel protein
292 (AKT31947.1 A213G), another is an intergenic (T->G) mutation at position
293 362,522 of the reference *Psa* NZ13 chromosome and the remaining four are
294 nonsynonymous substitutions in an adenylyltransferase (AKT32845.1, W977R);
295 chromosome segregation protein (AKT30494.1, H694Q); cytidylate kinase
296 (AKT29651.1, V173L) and peptidase protein (AKT32264.1, M418K).
297 The type III secretion system is known to be required for virulence in *P.*
298 *syringae*. A 44,620bp deletion event in *Psa* C17 resulted in the loss of 42 genes
299 encoding the structural apparatus and conserved type III secreted effectors in
300 *Psa* C17. This strain is highly compromised in its ability to grow in *A. chinensis*
301 var. *deliciosa* 'Hayward', attaining 1.2 x 10^7 cfu/g three days post inoculation
302 (dpi) and declining to 8.8 x 10^4 cfu/g at fourteen dpi (Figure S2). This is a
303 marked reduction compared to *Psa* NZ13, which attains 3.0 x 10^9 and 4.2 x 10^7
304 cfu/g three and fourteen dpi. *Psa* C17 nevertheless multiplies between day 0 and
305 day 3, indicating that even in the absence of type III-mediated host defense
306 disruption, *Psa* may still proliferate in host tissues. The loss of the TTSS does not
307 inhibit the growth of *Psa* C17 as strongly in the more susceptible *A. chinensis* var.
308 *chinensis* 'Hort16A' cultivar.
309 Two potentially significant deletion events occurred in the ancestor of the
310 pandemic subclade: a frameshift caused by a mutation and single base pair
311 deletion in a glucan succinyltransferase (*opgC*) and a 6,456bp deletion in the wss
312 operon (Figure S3). Osmoregulated periplasmic glucans (OPGs, in particular
313 *opgG* and *opgH*) are required for motility, biofilm formation and virulence in

314 various plant pathogenic bacteria and fungi (28-30). Homologs of *opgGH* remain
315 intact in the pandemic subcclade, yet the premature stop mutation in *opgC* likely
316 results in the loss of glucan succinylation. The soft-rot pathogen *Dickeya dadantii*
317 expresses OpgC in high osmolarity conditions, resulting in the substitution of
318 OPGs by O-succinyl residues(31). *D. dadantii opgC* deletion mutants did not
319 display any reduction in virulence(31). *Psa* is likely to encounter high osmolarity
320 during growth and transport in xylem conductive tissues, yet the impact of the
321 loss of *opgC* on *Psa* fitness has yet to be determined. The most striking difference
322 between the pandemic subcclade and more divergent Chinese *Psa*-3 strains is
323 the deletion of multiple genes involved in cellulose production and acetylation of
324 the polymer (Figure S3)(32). The loss of cellulose production and biofilm
325 production is not associated with a reduction in growth or symptom
326 development of *P. syringae* pv. *tomato* DC3000 on tomato, but may enhance
327 bacterial spread through xylem tissues during vascular infections (33). In *P.*
328 *fluorescens* SBW25 deletion of the Wss operon significantly compromises ability
329 to colonise plant surfaces and in particular the phyllosphere of sugar beet (*Beta*
330 *vulgaris*) seedlings (Gal et al 2003 Mol Ecol). It is possible that loss of this locus
331 aids movement through the vascular system and / or dissemination among
332 plants, by limiting capacity for surface colonization and biofilm formation.
333

334 **Dynamic genome evolution of *Psa*-3**

335 Despite the high similarity within the core genome, extensive variation is
336 evident in the pangenome of *Psa*-3 (Figure 3). The core genome (4,339 genes in
337 99-100% of strains, and 674 genes in 95%-99% of strains, or 58-62 genomes)
338 comprises 50.5% of the total pangenome (9,931 genes). 674 genes are present in

339 15-95% of strains (9-57 genomes), the so-called 'shell genes' (Figure 3). The
340 flexible genome is comprised of the 'shell' and 'cloud' genes; the latter describes
341 genes present in 0-15% of strains (one to six genomes in this case). Cloud genes
342 contribute most to the flexible genome: 3,950 genes are present in one to six
343 strains. This is a striking amount of variation in a pathogen described as clonal
344 and monomorphic. It should be noted that sequencing and assembly quality will
345 impact annotation and pangenome estimates: omitting the poor quality J39
346 assembly results in a core and soft-core genome differing by 18 genes and a
347 reduction of the cloud by 275. Despite a relatively slow rate of mutation and
348 limited within-clade homologous recombination, the amount of heterologous
349 recombination demonstrates that the genomes of these pathogens are highly
350 labile. Mobile genetic elements like bacteriophage, transposons and integrases
351 make a dramatic contribution to the flexible genome. Integrative and conjugative
352 elements (ICEs) are highly mobile elements and have recently been
353 demonstrated to be involved in the transfer of copper resistance in *Psa* (Figure
354 4)(34). Prodigious capacity for lateral gene transfer creates extreme
355 discordance between ICE type, host phylogeny and host geography making these
356 regions unsuitable markers of host evolution and origin.
357 Three divergent ICEs have been previously described from the global
358 pandemic lineage {McCann et al}. Within *Psa*-3 ICEs were found in 53 of 62
359 isolates (nine of the divergent Chinese isolates were devoid of any such element)
360 (Figure 1). No phylogeographic signal is evident. For example, strains from
361 Sichuan, Shaanxi, Korea, Italy and Portugal share an identical ICE. Even within a
362 single Chinese provence, mulitple ICEs exist (Shaanxi and Sichuan isolates
363 harbour four and three different ICEs, respectively). Moreover, ICE host range is

364 not limited to Psa alone: the ICE found in every NZ isolate (and also recorded in
365 Chinese isolate C1) exists in essentially identical form in a strain of *P. syringae*
366 pv. *avellenae* CRAPAV013 isolated from hazelnut in 1991 in Latina, Italy (it
367 exhibits 98% pairwise identity, differing from the New Zealand ICE by a
368 transposon, 66bp deletion and mere 6 SNPs).

369 **DISCUSSION**

370 We have uncovered an endemic population of *Psa* infecting cultivated
371 kiwifruit in China. All *Psa* isolated within China are members of the same lineage
372 as that responsible for the latest pandemic. The pandemic strains isolated in
373 Italy, Portugal, Chile and New Zealand form a subclade within this lineage along
374 with a subset of Chinese isolates, indicating that the pandemic ultimately
375 emerged from the Chinese population of lineage 3 strains. Italian pandemic
376 strains share a SNP with six of nine Chinese pandemic *Psa* strains, indicating
377 there was likely a direct transmission event from China to Italy prior to 2008.
378 The New Zealand isolates share six clade-defining mutations, indicating that a
379 separate and single transmission event was responsible for the outbreak of
380 disease there. Identification of the transmission pathway introducing *Psa* into
381 New Zealand is dependent on obtaining a sample of *Psa* sharing some or all of
382 the mutations characteristic of NZ *Psa* from either the overseas source
383 population or from infected plant material arriving into New Zealand from an
384 overseas location. The relatively low mutation rate in the core genome of *Psa*
385 places a lower boundary on the ability of genomic epidemiology to resolve
386 transmission events occurring either rapidly (as a consequence of human-
387 mediated long-distance dissemination) or at a local scale. The Japanese
388 pandemic strains cluster with the NZ strains, and share all six clade-defining
389 mutations. This suggests that pandemic *Psa*-3 was either introduced into Japan
390 via New Zealand, or from the same as-yet unknown region in China from which
391 transmission to New Zealand occurred. *Psa*-3 was first identified as causing
392 disease in four prefectures across Japan in April 2014(5). Japan imported pollen
393 and plant material from both China and New Zealand prior and subsequent to

394 *Psa*-3 detection in both those countries, though the amount of pollen imported
395 from New Zealand in 2012 (349kg) and 2013 (190kg) far outweighed the
396 amount imported from China (1 kg in both 2012 and 2013)(35).
397 Our phylogeographic study of a single lineage giving rise to a pandemic in
398 *P. syringae* has revealed far greater diversity than was previously appreciated.
399 Extensive diversity between *Psa* isolates collected from *Actinidia* spp. was
400 observed in the same province. The amount of diversity present within lineage 3
401 indicates this population was present and circulating in China before the
402 pandemic began. The emergence of the pandemic subclade moreover has not
403 resulted in the replacement of more ancestral strains: both pandemic and
404 divergent lineage 3 *Psa* were isolated from four out of six provinces.
405 Strains from three different lineages have been isolated in both Korea and
406 Japan, while China harbours strains from only a single lineage. The most basal
407 lineages of canker-causing *Psa* are comprised of Korean strains isolated between
408 1997 and 2014 (*Psa*-2) and a member of the recently identified lineage *Psa*-5.
409 One early isolate (*Psa* K3, 1997) groups with the Japanese isolates in lineage 1,
410 and a more recent Korean isolate *Psa* K7 (2014) groups with the more diverse
411 Chinese isolates in lineage 3. Korea therefore harbours a more diverse
412 population of *Psa* than China, with strains from three distinct lineages of *Psa* (1, 2
413 and non-pandemic subclade 3). A novel group of *Psa* emerged very recently in
414 Japan (*Psa*-5) and appears to share an ancestor with the Korean *Psa*-2 strains.
415 With the recent dissemination of pandemic *Psa*-3 and the historical presence of
416 *Psa*-1 in Japan, the Japanese population of *Psa* is comprised of three distinct
417 lineages of *Psa* (1, 5 and pandemic subclade 3). Though no strains from *Psa*-1
418 have been isolated in either Japan or Korea since 1997, at least two lineages

419 currently coexist in both Japan and Korea. This strongly suggests that the source
420 population of all *Psa* is not kiwifruit in China, but is more likely to reside in either
421 Korea or Japan. The potential transmission of a non-pandemic lineage 3 strain
422 from China to Korea and the emergence of a new lineage in Japan supports our
423 earlier assertion that variants will continue to emerge to cause local epidemics
424 and global pandemics in the future.

425 Considering that the divergence time of this monophyletic pathovar
426 predates the commercialisation of kiwifruit by hundreds if not thousands of
427 years,*Psa* is likely associated with a non-domesticated host(s) in the wild. Both
428 *A. chinensis var. deliciosa* or *A. chinensis var. chinensis* are found in natural
429 ecosystems and have overlapping habitat ranges with cultivated kiwifruit in
430 many areas. However, despite isolating 747 *Pseudomonas* strains from both wild
431 and cultivated kiwifruit during this sampling program, we did not identify *Psa*
432 among any of the 188 *Pseudomonas* spp. isolated from 98 wild *A. chinensis var.*
433 *deliciosa* or *A. chinensis var. chinensis* sampled across six provinces in China
434 (Table S5). Very few *Actinidia* spp. have ranges extending to South Korea and
435 Japan: *A. arguta*, *A. kolomikta*, *A. polygama* and *A. rufa*. *A. arguta* are broadly
436 distributed across both Korea and Japan. Early work by Ushiyama *et al.* (1992)
437 found that *Psa* could be isolated from symptomatic *A. arguta* plants in Japan. The
438 elucidation of whether this wild relative of kiwifruit harbours diverse strains of
439 *Psa* that may emerge to cause future outbreaks is currently under investigation.
440 Alternately, a host shift from another domesticated crop may have occurred after
441 expansion in kiwifruit cultivation.
442

443 Numerous epidemiological studies of human pathogens have
444 demonstrated environmental or zoonotic origins, but there are few such studies
445 of plant pathogens ((23, 36-46). Where ecological and genetic factors restrict
446 pathogens to a small number of plant hosts some progress has been made, but
447 for facultative pathogens such as *P. syringae* that colonise multiple hosts and are
448 widely distributed among both plant and non-plant habitats, the environmental
449 reservoirs of disease and factors affecting their evolutionary emergence are
450 difficult to unravel ((47, 48).
451 The emergence of *Psa* over the last three decades – concomitant with
452 domestication of kiwifruit– offers a rare opportunity to understand the
453 relationship between wild populations of both plants and microbes and the
454 ecological and evolutionary factors driving the origins of disease, including the
455 role of agriculture. It is now possible to exclude China as the native home to the
456 source population, but the precise location remains unclear. Nonetheless, it is
457 likely, given the extent of diversity among *Psa* isolates and the time-line to
458 domestication, that ancestral populations exist in non-agricultural plant
459 communities. Attention now turns to Korea and Japan and in particular the
460 interplay between genetic and ecological factors that have shaped Psa evolution.
461
462
463 .

464 **MATERIALS AND METHODS**

465 **Bacterial strains and sequencing**

466 Samples were procured by isolation from symptomatic plant tissue.
467 Bacterial strain isolations were performed from same-day sampled leaf and stem
468 tissue by homogenising leaf or stem tissue in 800uL 10mM $MgSO_4$ and plating
469 the homogenate on *Pseudomonas* selective media (King's B supplemented with
470 cetrimide, fucidin and cephalosporing, Oxoid). Plates were incubated 48 hours
471 between 25 and 30°C. Single colonies were restreaked and tested for oxidase
472 activity, and used to inoculate liquid overnight cultures in KB. Strains were then
473 stored at -80°C in 15% glycerol and the remainder of the liquid culture was
474 reserved for genomic DNA isolation by freezing the pelleted bacterial cells in a
475 96 well plate at -20°C. Genomic DNA extractions were performed using Promega
476 Wizard 96-well genomic DNA purification system.

477 Initial strain identification was performed by sequencing the citrate
478 synthase gene (*cts*, aka *gltA* (49)). Subsequent to strain identification, paired-end
479 sequencing was performed using the Illumina HiSeq 2500 platform (Novogene,
480 Guangzhou, China). Additional paired-end sequencing was performed at New
481 Zealand Genomics Limited (Auckland, New Zealand) using the MiSeq platform,
482 and raw sequence reads from some previously published isolates were shared by
483 Mazzaglia *et al.* (24).

484

485

486

487 **Variant Calling and Recombination Analyses**

488 The completely sequenced genome of *Pseudomonas syringae* pv. *actinidiae*
489 NZ13 was used as a reference for variant calling. A near complete version of this
490 genome was used as a reference in our previous publication and subsequently
491 finished by Templeton *et al.* (2015), where it is referred to as ICMP1884 (23, 25).
492 Variant calling was performed on all *P. syringae* pv. *actinidiae* isolates for which
493 read data was available.

494 Read data was corrected using the SPADEs correction module and
495 Illumina adapter sequences were removed with Trimmomatic allowing 2 seed
496 mismatches, with a palindrome and simple clip threshold of 30 and 10,
497 respectively (50, 51). Quality-based trimming was also performed using a sliding
498 window approach to clip the first 10 bases of each read as well as leading and
499 trailing bases with quality scores under 20, filtering out all reads with a length
500 under 50 (51). PhiX and other common sequence contaminants were filtered out
501 using the Univec Database and duplicate reads were removed (52).

502 Reads were mapped to the complete reference genome *Psa* NZ13 with
503 Bowtie2 and duplicates removed with SAM Tools (53, 54). Freebayes was used
504 to call variants with a minimum base quality 20 and minimum mapping quality
505 30 (55). Variants were retained if they had a minimum alternate allele count of
506 10 reads and fraction of 95% of reads supporting the alternate call. The average
507 coverage was calculated with SAM Tools and used as a guide to exclude
508 overrepresented SNPs (defined here as threefold higher coverage than the
509 average) which may be caused by mapping to repetitive regions. BCFtools
510 filtering and masking was used to generate final reference alignments including
511 SNPs falling within the quality and coverage thresholds described above and

512 excluding SNPs within 3bp of an insertion or deletion (indel) event or indels
513 separated by 2 or fewer base pairs. Invariant sites with a minimum coverage of
514 10 reads were also retained in the alignment, areas of low (less than 10 reads) or
515 no coverage are represented as gaps relative to the reference.
516 Freebayes variant calling includes indels and multiple nucleotide
517 insertions as well as single nucleotide insertions, however only SNPs were
518 retained for downstream phylogenetic analyses. An implementation of
519 ClonalFrame suitable for use with whole genomes was employed to identify
520 recombinant regions using a maximum likelihood starting tree generated by
521 RaxML (56, 57). All substitutions occurring within regions identified as being
522 introduced due to recombination by ClonalFrameML were removed from the
523 alignments. The reference alignments were manually curated to exclude
524 substitutions in positions mapping to mobile elements such as plasmids,
525 integrative and conjugative elements and transposons.
526

527 **Phylogenetic Analysis**

528 The maximum likelihood phylogenetic tree of 80 *Psa* strains comprising
529 new Chinese isolates and strains reflecting the diversity of all known lineages
530 was built with RAxML (version 7.2.8) using a 1,216,321bp core genome
531 alignment excluding all positions for which one or more genomes lacked
532 coverage of 10 reads or higher (57). The core genome alignmnt included 2,207
533 variant sites. Membership within each phylogenetic lineage corresponds to a
534 minimum average nucleotide identity of 99.70%. The average nucleotide identity
535 was determined using a BLAST-based approach in JspeciesWS (ANIb), using a
536 subset of 32 *Psa* genome assemblies spanning all lineages(58). In order to fully

537 resolve the relationships between more closely related recent outbreak strains, a
538 phylogeny was constructed using only the 62 *Psa*-3 strains. This was determined
539 using a 4,853,421bp core genome alignment including 1,951 non-recombinant
540 SNPs and invariant sites. Trees were built with the generalized time-reversible
541 model and gamma distribution of site-specific rate variation (GTR+Γ) and 100
542 bootstrap replicates. *Psa* C16 was used to root the tree as this was shown to be
543 the most divergent member of the phylogeny when including strains from
544 multiple lineages. Nodes shown have minimum bootstrap support values of 50.
545

546 **Identification of the core and mobile genome**

547 Genomes were assembled with SPAdes using the filtered, trimmed and
548 corrected reads (50). Assembly quality was improved with Pilon and annotated
549 with Prokka (59, 60). The pangenome of *Psa*-3 was calculated using the ROARY
550 pipeline(61). Orthologs present in 61 (out of a total of 62) genomes were
551 considered core; presence in 58-60, 9-57 and 1-8 were considered soft-core,
552 shell and cloud genomes, respectively. BLAST-based confirmation was used to
553 confirm the identity predicted virulence or pandemic-clade-restricted genes in
554 genome assemblies.
555

556 **Pathogenicity assays**

557 Growth assays were performed using both stab inoculation as in McCann
558 *et al*. (2013) an initial inoculum of 10^8cfu/mL and four replicate plants at day 0
559 and six at all subsequent sampling time points. Bacterial density in inoculated
560 tissue was assessed by homogenizing tissue and serial dilution plating. Statistical

561 significance between each treatment at each time point was assessed using two-
562 tailed t-tests with uneven variance.

563 ACKNOWLEDGEMENTS

564 We gratefully acknowledge the assistance of the following guides,
565 teachers, and graduate assistants who helped us identify sample locations in
566 China: Junjie Gong, Yancang Wang, Shengju Zhang, Zupeng Wang, Yangtao Guo,
567 Meiyan Chen, Kuntong Li, Moucai Wang, Jiaming He, Yonglin Zhao, Zhongshu Yu,
568 Yan Lv, Mingfei Yao, Shihua Pu, Tingwen Huang, Qiuling Hu, Caizhi He, and
569 Jiaqing Peng. Derk Wachsmuth at Max Planck Institute for computing server
570 support. James Connell for assistance with biosecurity regulations. Members of
571 the Rainey and Huang labs for discussion. Joel Vanneste for contributing strains.
572 This work was funded by grants from the New Zealand Ministry for Business,
573 Innovation and Employment (C11X1205), Canada Natural Sciences and
574 Engineering Research Council, Chinese Academy of Sciences Present's
575 International Fellowship Initiative (Grant NO. 2015PB063), China Scholarship
576 Council (Grant NO. 201504910013), National Natural Science Foundation of
577 China (Grant NO. 31572092), Science and Technology Service Network Initiative
578 Foundation of The Chinese Academy of Sciences (Grant NO. KFJ-EW-STS-076),
579 Protection and utilization of Crop Germplasm Resources Foundation of Ministry
580 of Agriculture (Grant NO. 2015NWB027).

581 **TABLES**

582 **Table 1. Origin of SNPs in core genomes**

		Intergenic	Synonymous	Non-synonymous	Total	
Lineage 3	Mutation	412	586	953	1,951	
	Recombination	35	137	91	263	7.42
All lineages	Mutation	457	1,494	1,008	2,959	
	Recombination	355	2,218	579	3,152	0.94

583 **Table S1. Strains**

584

ID	WGS origin	Host plant	Country	Year	Other collection/alias	Contigs	N50
C1[1]	Mazzaglia et al (2012)	*A. chinensis* 'Hongyang'	China, Shaanxi, Wei	2010	CH2010-6, M7	470	24,560
C3	This paper	*A. deliciosa* 'Hayward'	China, Shaanxi, Xi'an, Zhouzhi	2012	ZY2, CC770	318	46,356
C9[1]	Butler et al (2013)		China, Shaanxi, Wei County	2010	M228	346	38,960
C10	This paper	*A. chinensis* 'Hongyang'	China, Sichuan, Shifang	2012	850.1.1. CC822	325	45,191
C11	This paper	*A. chinensis* 'Hongyang'	China, Sichuan, Shifang	2012	850.2.2. CC823	335	50,545
C12	This paper	*A. chinensis* 'Hongyang'	China, Sichuan, Shifang	2012	850.4.1. CC826	339	48,835
C13	This paper	*A. chinensis* 'Hongyang'	China, Sichuan, Shifang	2012	850.5.1. CC827	326	44,835
C14	This paper	*A. chinensis* 'Hongyang'	China, Sichuan, Shifang	2012	850.6.1. CC828	326	44,683
C15	This paper	*A. chinensis* 'Hort16A'	China, Sichuan, Pengzhou, Cifeng	2012	913.1.1 CC835	327	54,039
C16	This paper	*A. chinensis* 'Hongyang'	China, Hubei, Enshi, Jianshi	2012	913.5.1, CC836	386	42,362
C17	This paper	*A. chinensis* 'Hongyang'	China, Sichuan, Qionglai, Huojing	2012	913.10.1, CC837	366	43,680
C18	This paper	*Actinidia* sp.	China, Shaanxi, Baoji, Meixian	2012	913.15.1, CC838	325	50,834
C24	This paper	*A. chinensis* 'Hongyang'	China, Chongqing, Wanzhou, Houshan	2014	120L3	338	48,742
C26	This paper	*A. chinensis* 'Hongyang'	China, Chongqing, Wanzhou, Houshan	2014	124L1	343	47,302
C27	This paper	*A. chinensis* 'Hongyang'	China, Chongqing, Wanzhou, Houshan	2014	124L7	364	48,757
C28	This paper	*A. chinensis* 'Hongyang'	China, Sichuan, Dujiangyan, Xujia	2014	139S2	343	46,114
C29	This paper	*A. deliciosa* 'Hayward'	China, Sichuan, Dujiangyan, Hongkou	2014	163S1	321	41,599
C30	This paper	*A. chinensis*	China, Sichuan, Dujiangyan, Hongkou	2014	165L4	334	43,004
C31	This paper	*A. chinensis*	China, Sichuan, Dujiangyan, Hongkou	2014	166L2	303	43,291
C48	This paper	*A. chinensis*	China, Hunan, Changde City, Shimen	2014	50L1	351	47,058
C54	This paper	*A. deliciosa*	China, Guizhou, Liupanshui, Panxian	2014	77L5	360	47,421
C62	This paper	*A. chinensis* 'Jinyan'	China, Guizhou, Liupanshui, Liuzhi	2014	GZ410	327	45,829
C66	This paper	*A. chinensis* 'Hongyang'	China, Hubei, Yichang, Yiling	2014	YC5	333	48,710

C67	This paper	*A. chinensis* 'Hongyang'	China, Chongqing, Qianjiang, Jinxi	2012	163W4	310	46,111
C68	This paper	*A. chinensis* 'Jinyan'	China, Guizhou, Liupanshui, Liuzhi	2014	GZ3-5	347	45,058
C69	This paper	*A. chinensis* 'Jinyan'	China, Shaanxi, Xi'an, Zhouzhi	2014	SH1-14	304	47,863
C70	This paper	*A. chinensis*	China, Sichuan, Dujiangyan, Xujia	2014	141S5	327	48,715
C73	This paper	*A. chinensis* 'Hongyang'	China, Sichuan, Dujiangyan, Xujia	2014	139L10	365	48,937
C74	This paper	*A. chinensis* 'Hongyang'	China, Sichuan, Dujiangyan, Xiange	2014	148L1	326	47,058
C75	This paper	*A. chinensis* 'Hongyang'	China, Sichuan, Dujiangyan, Xiange	2014	148L4	336	47,058
CL4[1]	Butler et al (2013	*A. deliciosa*	Chile, Maule	2010		370	37,470
I1[1]	Marceletti et al (2011)	*A. deliciosa* 'Hayward'	Italy, Roma	1992	NCPPB3871	405	27,730
I2[1]	Marceletti et al (2011)	*A. chinensis* 'Hort16A'	Italy, Latina	2008	CRAFRU8.43	523	22,372
I3[1]	Mazzaglia et al (2012)	*A. chinensis* 'Hort16A'	Italy, Lazio	2008	CFBP 7286	329	31,420
I10[1]	Butler et al (2013)	*A. deliciosa*	Italy, Roma	2010	ICMP18744, CRAFRU11.41	358	35,904
I11	Mazzaglia et al (2012)	*A. chinensis* 'Jin Tao'	Italy, Veneto	2008	CFBP 7285	359	33,568
I13	Mazzaglia et al (2012)	*A. deliciosa* 'Hayward'	Italy, Lazio	2008	CFBP 7287	357	36,668
J1[1]	Baltrus et al (2011)	*A. deliciosa*	Japan, Kanagawa	1984	MAFF 302091	248	65,551
J2[1]	Mazzaglia et al (2012)	*A. chinensis*	Japan	1988	PA459	634	17,643
J25[1]	Mazzaglia et al (2012)	*A. deliciosa* 'Hayward'	Japan, Shizuoka	1984	KW41	570	18,393
J29	McCann et al (2013)	A. arguta	Japan, Kanagawa	1987	MAFF302133, JpSar1	412	30,107
J30	McCann et al (2013)	A. arguta	Japan, Kanagawa	1987	MAFF302134, JpSar2	401	32,586
J31	McCann et al (2013)	*A. deliciosa* 'Hayward'	Japan, Kanagawa	1987	MAFF302143, JpKiw4	1723	4,505
J32	McCann et al (2013)	*A. deliciosa* 'Hayward'	Japan, Wakayama	1988	MAFF302145, JpWa1	465	33,908
J33	McCann et al (2013)	*A. deliciosa* 'Hayward'	Japan, Wakayama	1988	MAFF302146, JpWa2	410	31,305
J35	McCann et al (2013)	*A. deliciosa* 'Hayward'	Japan, Shizuoka	1984	NCPPB 3739, Kw11	368	39,207
J36[1]	Butler et al (2013)	*A. deliciosa* 'Hayward'	Japan, Shizuoka	1984	Kw1	417	27,018
J37[1]	Fujikawa & Sawada (2015)		Japan, Saga	2014	PRJDB2950	291	50,639
J38	This paper	*A. chinensis* 'Hort16A'	Japan	2014		313	48,009

J39[2]	This paper	*A. chinensis* 'Hort16A'	Japan, Saga	2014		1221	9,308
K3	Mazzaglia et al (2012)	*A. deliciosa*	Korea, Jeonnam	1997	KN.2	962	10,103
K4	This paper	*A. chinensis* 'Hort16A'	Korea	2014		258	46,118
K5	This paper	*A. chinensis* 'Hort16A'	Korea, Jeju	2014		313	46,355
K6	This paper	*A. chinensis* 'Hort16A'	Korea	2011		270	43,774
K7	This paper	*A. chinensis* 'Hort16A'	Korea	2014		330	46,153
K26	McCann et al (2013)	*A. chinensis*	Korea, Jeonnam	1997	KACC10584	290	36,930
K27	McCann et al (2013)	*A. chinensis*	Korea, Jeonnam	1998	KACC10594	413	25,076
K28	McCann et al (2013)	*A. chinensis*	Korea, Jeonnam	1997	KACC10574	297	37,347
NZ13	McCann et al (2013)	*A. deliciosa* 'Hayward'	New Zealand, Te Puke	2010		1	
NZ31[1]	Butler et al (2013)	*A. deliciosa*	New Zealand, Paengaroa	2010		382	33,149
NZ32[1]	Butler et al (2013)	*A. chinensis*	New Zealand, Te Puke	2010		367	31,162
NZ33[1]	Butler et al (2013)		New Zealand, Te Puke	2011	TP1	380	31,549
NZ34[1]	Butler et al (2013)		New Zealand, Te Puke	2011	6.1	386	46,155
NZ35	This paper	*Actinidia* sp. pollen	New Zealand, Te Puke	2010		330	47,026
NZ37	This paper	*Actinidia* sp.	New Zealand, Te Puke	2010	BF	317	46,211
NZ38	This paper	*A. deliciosa*	New Zealand, Te Puke	2014	627	322	46,155
NZ39	This paper	*A. deliciosa*	New Zealand, Te Puke	2014	670	315	46,155
NZ40	This paper	*A. deliciosa*	New Zealand, Te Puke	2014	793	322	45,201
NZ41	This paper	*A. deliciosa*	New Zealand, Te Puke	2014	854.2	322	46,356
NZ42	This paper	*A. deliciosa*	New Zealand, Te Puke	2014	632.1	322	48,669
NZ43	This paper	*A. deliciosa*	New Zealand, Te Puke	2014	694.1	322	44,416
NZ45	This paper	*A. deliciosa*	New Zealand, Te Puke	2014	1014	364	48,150
NZ46	This paper	*A. chinensis*	New Zealand, Matakana Island	2012		325	48,671
NZ47	This paper		New Zealand, Te Puke	2014	851	320	46,212
NZ48	This paper	*A. chinensis* 'Hort16A'	New Zealand, Te Puke	2013	821	320	48,955

NZ49	This paper		New Zealand, Te Puke	2011	691	325	46,155
NZ54	This paper	*A. chinensis*	New Zealand, Pukekohe	2014		323	48,754
NZ59	This paper		New Zealand	2015		315	44,129
NZ60	This paper		New Zealand	2015		332	38,090
P1	Mazzaglia et al (2012)	*A. deliciosa* 'Summer'	Portugal	2010	346	354	31,352

585 [1]Simulated reads were generated from contigs available for these previously sequenced genomes
586 [2]Different sequencing runs were employed for draft genome assembly and variant calling

587

588 **Table S2. Average percent identity within and between Psa lineages**

589

Lineage	*Psa*-1	*Psa*-2	*Psa*-3	*Psa*-5
Psa-1	99.70			
Psa-2	98.97	99.76		
Psa-3	99.06	98.91	99.73	
Psa-5	98.87	98.96	98.83	ND[1]

590

591 [1]Not determined for Psa-5 as only a single strain has been sequenced from this lineage. ANIb values
592 determined using representative strains for *Psa*-1 (J31, K3, J2, J1, J30, J29, J35, J32, J33, I1, J36,
593 J25), *Psa*-2 (K27, K6, K4, K26, K28), *Psa*-3 (C16, C17, C10, C11, C70, K7, C15, C54, C74, C69, K5,
594 C62, I13, NZ13) and *Psa*-5 (J37)

595

596 **Table S3. Genes with multiple mutations in *Psa*-3**

597 (See Table_S3.xlsx)

598

599 **Table S4. SNPs shared between all pandemic NZ and Japanese isolates**

600

Protein ID (NZ13)	Product	Codon
AKT31947.1	ion channel protein Tsx	71(silent)
AKT32845.1	bifunctional glutamine-synthetase adenylyltransferase	W977R
AKT30494.1	chromosome segregation protein SMC	H694Q
AKT29651.1	cytidylate kinase	V173L
AKT32264.1	peptidase PmbA	M418K
Intergenic		362,522(G->T)

601 **Table S5. Isolates identified by cultivation and disease status of host**

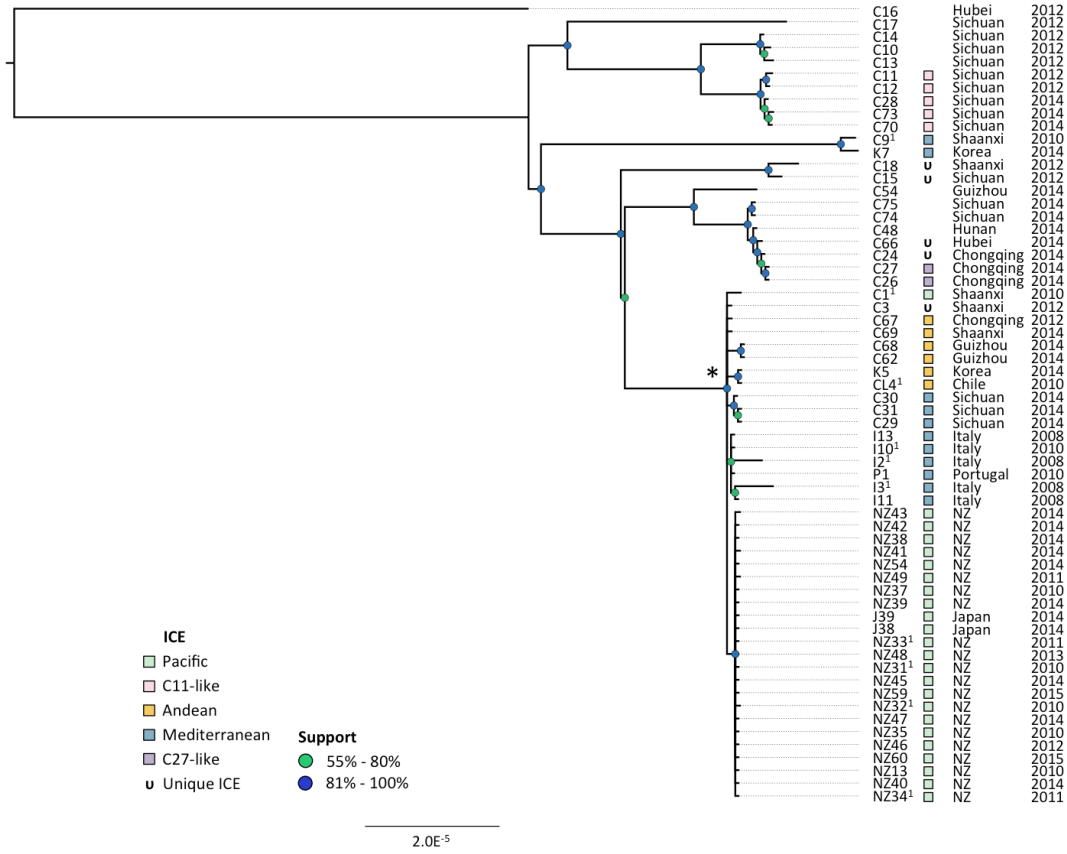
	A. arguta	*A. callosa*	*A. chinensis* var. *chinensis*	*A. chinensis* var. *deliciosa*	*Actinidia* sp.	*Camellia* sp.	*Prunus* sp.	Total
Cultivated	3		373	350	1	45	45	817
Disease	3		323	244		30	45	645
Not determined	2		121	81		20	10	234
Pseudomonas spp.	1		123	120		8	35	287
P. syringae			12	23		2		37
Psa			67	20				87
Healthy			27	44		1		72
Not determined			10	16		1		27
Pseudomonas spp.			17	28				45
P. syringae								0
Psa								0
Suspected			17	62		14		93
Not determined			9	27		1		37
Pseudomonas spp.			7	31		12		50
P. syringae			1	4		1		6
Psa								0
Wild	38	15	157	230		21		461
Disease	14		3					17
Not determined	2		1					3
Pseudomonas spp.	11		2					13
P. syringae	1							1
Psa								0
Healthy	10	15	62	214		21		322
Not determined	3	3	33	121		16		176
Pseudomonas spp.		12	15	81		4		112
P. syringae	7		14	12		1		34

Psa								0
Suspected	14		92	16				122
Not determined	4		33	11				48
Pseudomonas spp.	10		42	5				57
P. syringae			17					17
Psa								0
Total	41	15	530	580	1	66	45	1278

602
603
604

605 **FIGURES**

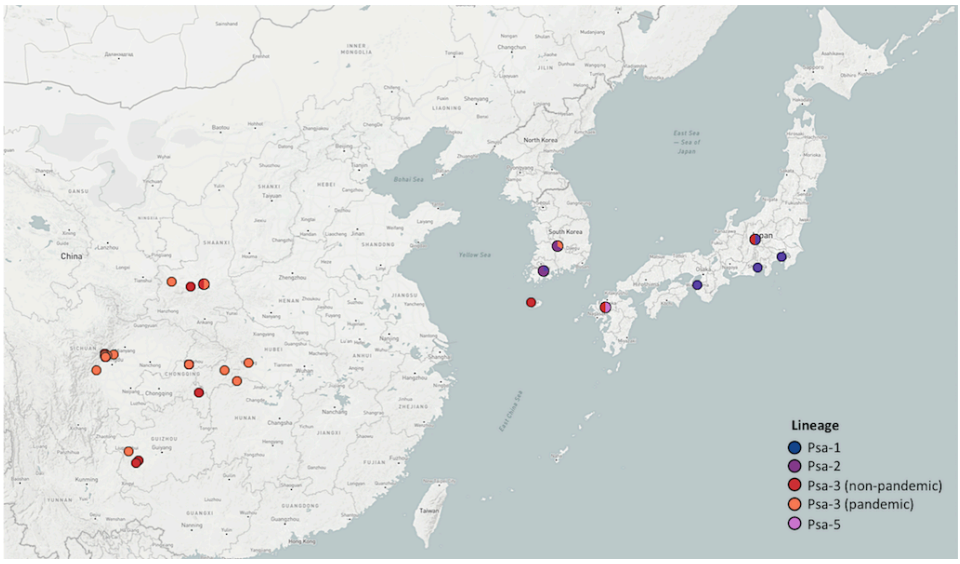
606 **Figure 1. Phylogeny of *Psa* lineage 3**



607

608 **Figure 2. Psa-3 isolation locations in East Asia**

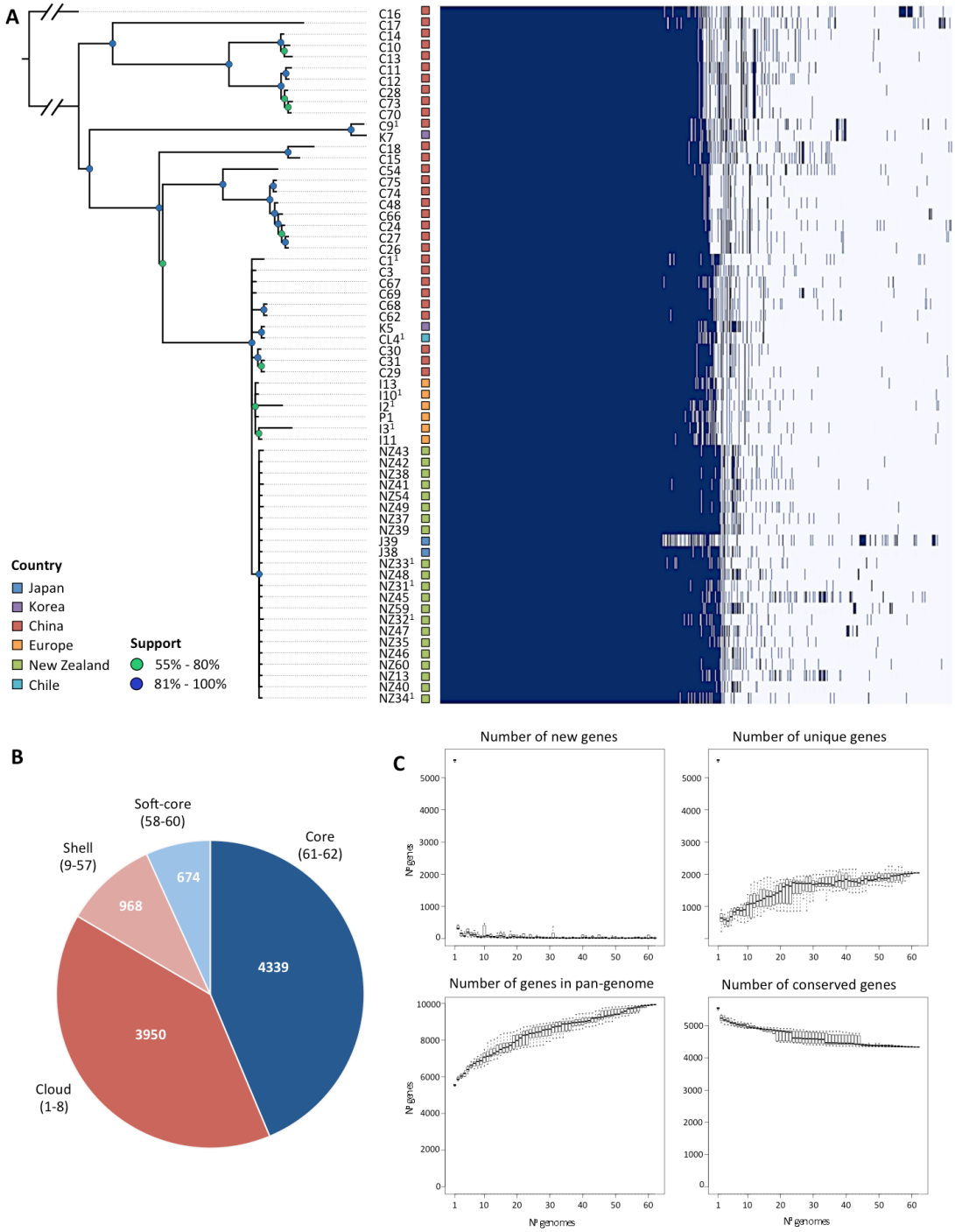
609



610

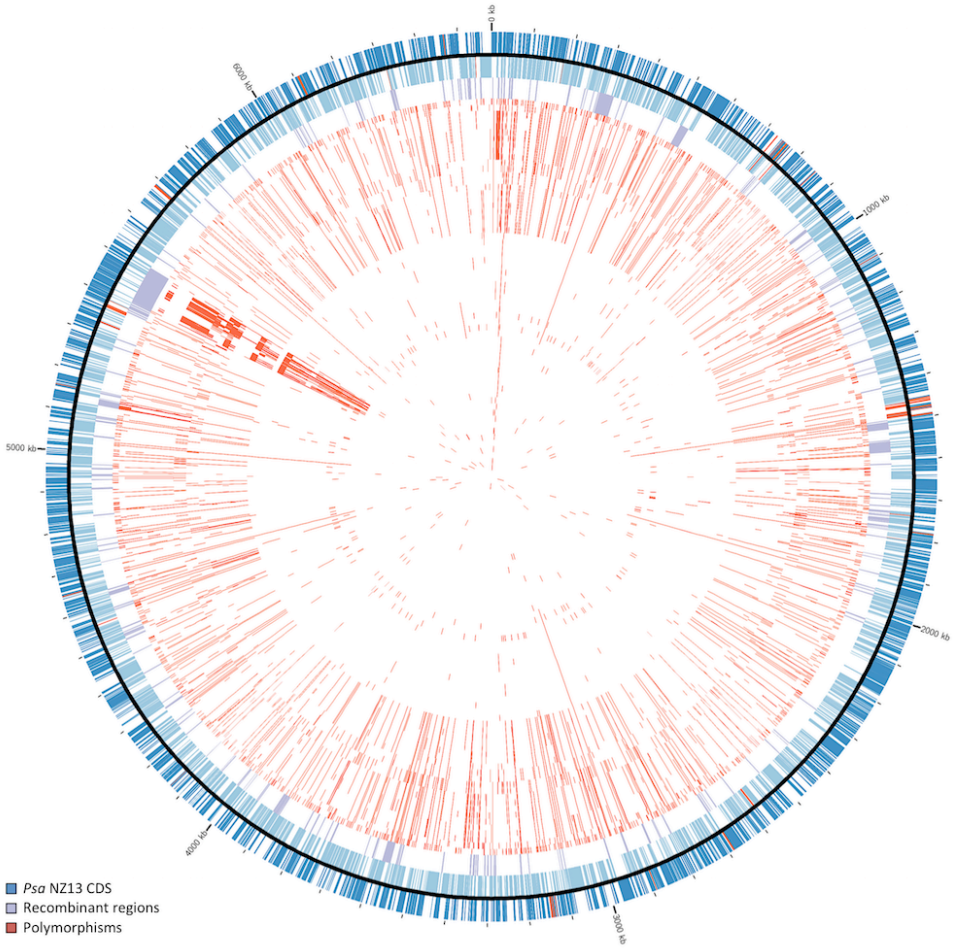
611

Figure 3. Pangenome of *Psa*-3



617 **Figure 4. Genomic context of polymorphisms in *Psa*-3**

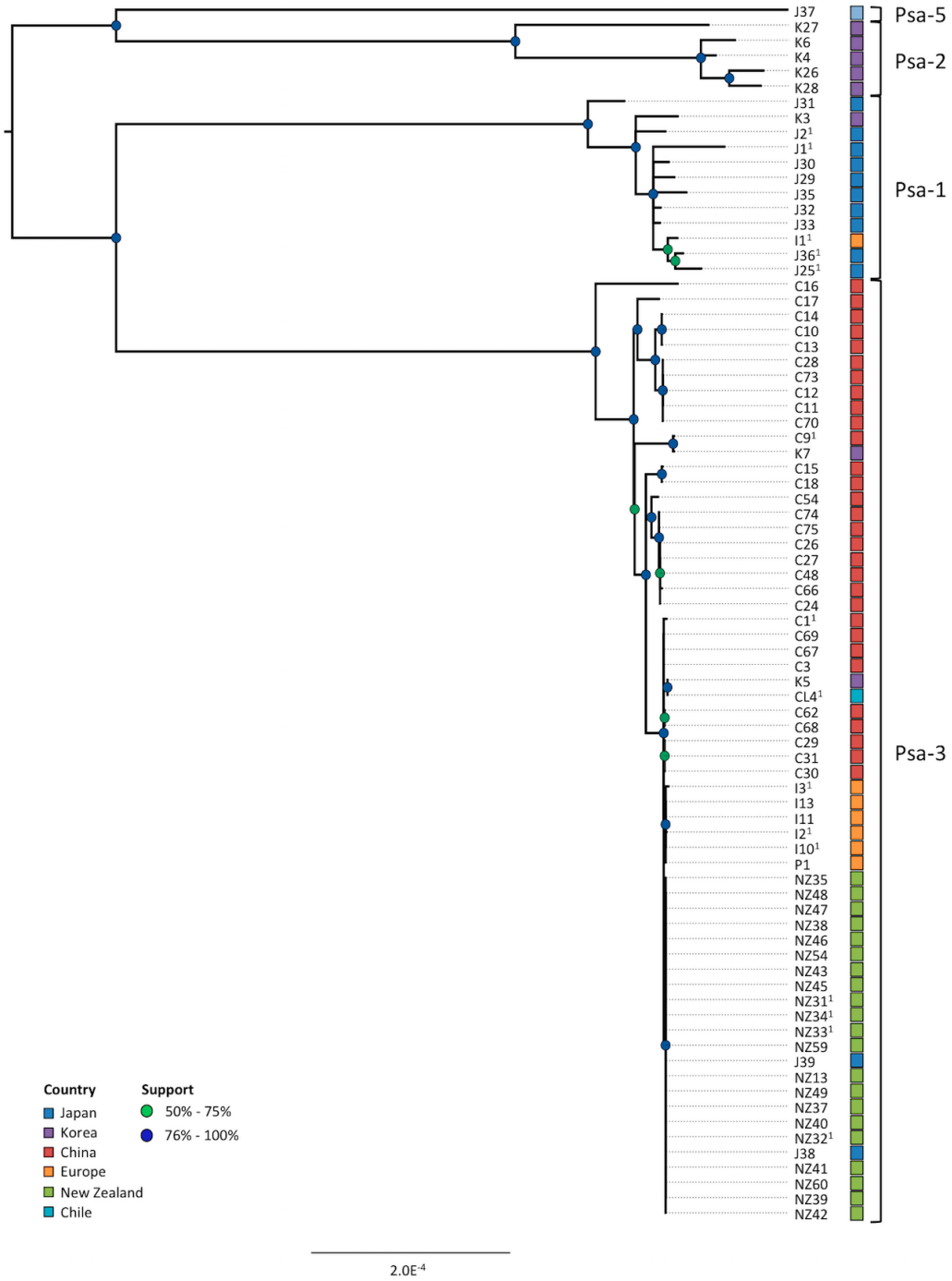
618



619

620

621 **Figure S1. Phylogeny of all *Psa***

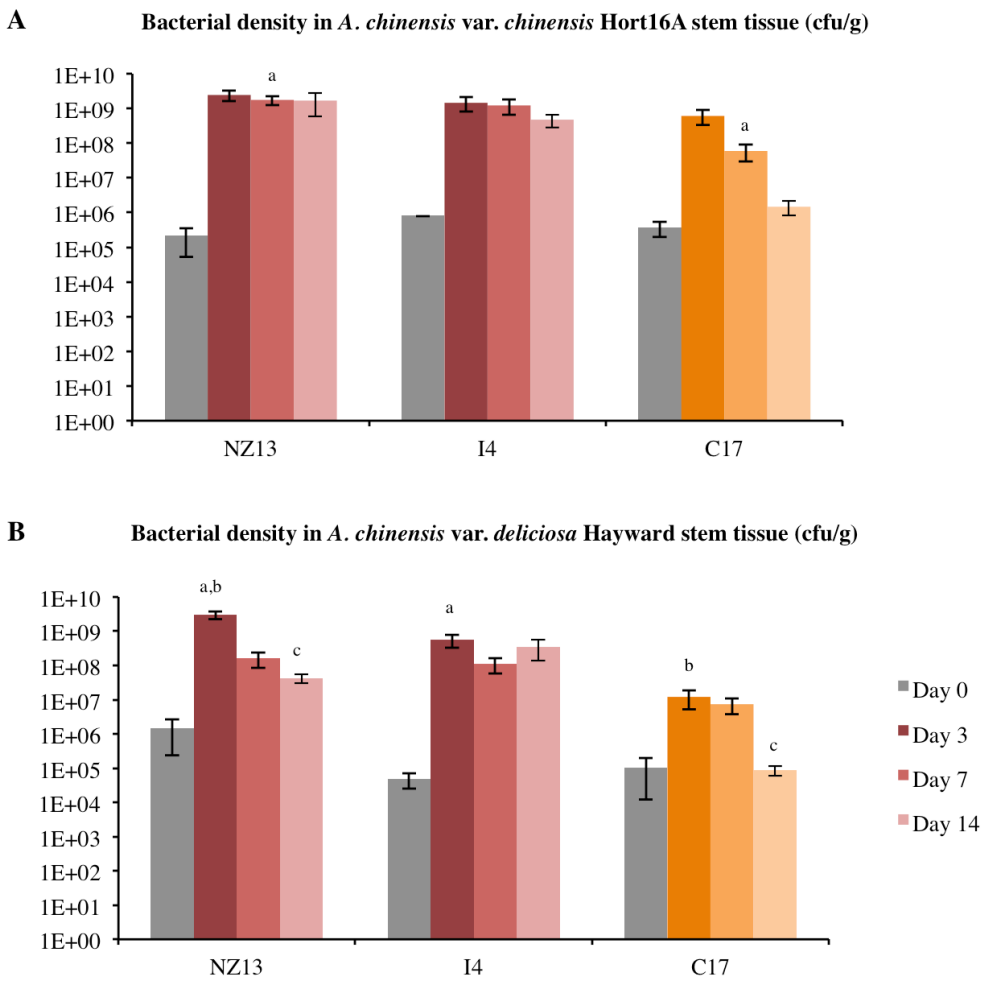


622

623

624 **Figure S2. Bacterial growth assay of *Psa* on *A. chinensis***

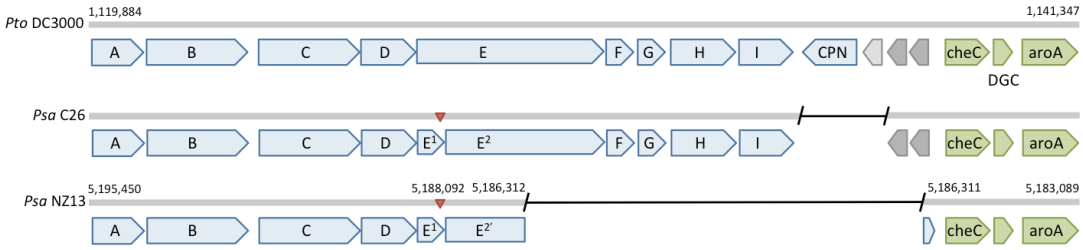
625



626

627 **Figure S3. Wss operon disruption in *Psa*-3**

628



629

630 **FIGURE LEGENDS**

631 **Figure 1. Phylogeny of *Psa* lineage 3**

632 Maximum likelihood tree based on 4,853,421bp non-recombinant core genome
633 alignment including 1,951 variant sites. All nodes displayed have bootstrap
634 support values above 55% (55-80% in green, 81-100% in blue). Year and
635 province (China) or country of isolation is displayed. Integrative and conjugative
636 elements (ICEs) present in each host genome are indicated.

637

638 **Figure 2. Psa-3 isolation locations in East Asia**

639 Filled circles' positions correspond to location of isolation (select Japanese and
640 Korean isolates do not have reliable isolation location information). Colour
641 corresponds to the phylogenetic position of the isolates as shown in Figure S1.
642 Map generated in Microreact.

643

644 **Figure 3. Pangenome of *Psa*-3**

645 A. Presence/absence matrix of all core and accessory genes in *Psa*-3, ordered
646 according to strains' phylogenetic relationships. Country of isolation is indicated
647 at left. B. The core and flexible genome of *Psa*-3. The core, soft-core, shell and
648 cloud genomes are defined according to the numbers in parentheses. C. Number
649 of new, unique and conserved genes with addition of each genome.

650

651 **Figure 4. Genomic context of polymorphisms in *Psa*-3**

652 Polymorphisms and recombinant regions mapped onto *Psa* NZ13 reference
653 genome using CIRCOS(62). *Psa* NZ13 CDS are displayed in the first and second
654 ring (blue), with annotated Type 3 secretion system and effectors highlighted
655 (red). Inner rings display polymorphisms in *Psa*-3 genomes ordered from most
656 to least divergent relative to *Psa* NZ13. The integrative and conjugative element
657 (ICE) is the most polymorphic region.

658

659 **Figure S1. Phylogeny of *Psa***

660 RaxML Maximum likelihood tree based on 1,216,321bp non-recombinant core
661 genome alignment including 2,207 variant sites. All nodes displayed have
662 bootstrap support values above 50% (50-75% in green, 76-100% in blue).
663 Province of isolation is displayed for Chinese isolates.

664

665 **Figure S2. Bacterial growth assay on *A. deliciosa***

666 Bacterial growth of pandemic *Psa* NZ13, I4 (red) and divergent C17 (orange)
667 strains on *A. chinensis* var. *chinensis* 'Hort16A' and *A. chinensis* var. *deliciosa*
668 'Hayward'. Mean *in planta* bacterial density in stem tissue (cfu/g) at 0, 3 and 7
669 days post-inoculation is shown (mean ± SEM) with superscript denoting
670 significant difference between strains at each sampling time ($P<0.05$, two-tailed
671 t-test, unequal variance). Four replicate plants were assayed at day 0, and six
672 replicates at each subsequent time point.

673

674 **Figure S3. Wss operon disruption in *Psa*-3**

675 Genes encoding components of the *wss* operon (blue), hypothetical and
676 conserved hypothetical (light and dark grey), chemotaxis, diguanylate cyclase
677 and *aroA* (green). Deletions (black line) and position of single base pair insertion
678 (red triangle) displayed with reference to *Pto* DC3000. Insertion results in
679 frameshift mutation in *wssE*, two predicted derivatives annotated as *wssE1* and
680 *wssE2*. The subsequent 6.5kb deletion in the ancestor of the pandemic subclade
681 results in the truncation of *wssE2*, annotated as *wssE2'*.

682 **REFERENCES**

683 1. Everett KR, et al. (2011) First report of *Pseudomonas syringae* pv.
684 *actinidiae* causing kiwifruit bacterial canker in New Zealand. *Australasian*
685 *Plant Dis Notes* 6(1):67–71.

686 2. Vanneste JL, et al. (2011) First report of *Pseudomonas syringae* pv.
687 *actinidiae*, the causal agent of bacterial canker of kiwifruit in France. *Plant*
688 *disease* 95(10):1311–1311.

689 3. Balestra GM, Renzi M, Mazzaglia A (2010) First report of bacterial canker
690 of *Actinidia deliciosa* caused by *Pseudomonas syringae* pv. *actinidiae* in
691 Portugal. *New Disease Reports*.

692 4. Abelleira A, et al. (2011) First report of bacterial canker of kiwifruit caused
693 by *Pseudomonas syringae* pv. *actinidiae* in Spain. *Plant disease*
694 95(12):1583–1583.

695 5. Sawada H, et al. (2015) Characterization of biovar 3 strains of
696 *Pseudomonas syringae* pv. *actinidiae* isolated in Japan. *Annals of the*
697 *Phytopathological Society of Japan* 81(2):111–126.

698 6. Koh YJ, et al. (2012) Occurrence of a new type of *Pseudomonas syringae* pv.
699 *actinidiae* strain of bacterial canker on kiwifruit in Korea. *The Plant*
700 *Pathology Journal* 28(4):423–427.

701 7. Zhao ZB, Gao XN, Huang QL, Huang LL, Qin HQ (2013) Identification and
702 characterization of the causal agent of bacterial canker of kiwifruit in the
703 Shaanxi province of China. *Journal of Plant ….*

704 8. Serizawa S, Ichikawa T, Takikawa Y, Tsuyumu S, Goto M (1989)
705 Occurrence of bacterial canker of kiwifruit in Japan: Description of
706 symptoms, isolation of the pathogen and screening of bactericides. *Annals*
707 *of the Phytopathological Society of Japan* 55:427–436.

708 9. Koh YJ, Jung JS, Hur JS (2002) current status of occurrence of major
709 diseases on kiwifruits and their control in Korea. *Acta Horticulturae: V*
710 *International Symposium on Kiwifruit*.

711 10. Ferguson AR, Huang H (2007) Genetic resources of kiwifruit:
712 domestication and breeding. *Horticultural reviews*:1–121.

713 11. Ferguson AR (2011) Kiwifruit: Evolution of a crop. *ISHS Acta*
714 *Horticulturae: VII International Symposium on Kiwifruit 913* 913:31–42.

715 12. Huang H, Wang Y, Zhang Z, Jiang Z, Wang S (2004) *Actinidia* germplasm
716 resources and kiwifruit industry in China. *HortScience* 39(6):1165–1172.

717 13. Shim KK, Ha YM (1999) Kiwifruit production and research in Korea. *Acta*
718 *Hortic* (498):127–132.

719 14. Testolin R, Ferguson AR (2009) Kiwifruit (*Actinidia* spp.) production and
720 marketing in Italy. *New Zealand Journal of Crop and Horticultural Science*
721 37(1):1–32.

722 15. Ferguson AR (2015) Kiwifruit in the world. *Acta Hortic* (1096):33–46.

723 16. Cruzat C (2014) The kiwifruit in Chile and in the world. *Revista Brasileira*
724 *de Fruticultura*.

725 17. Fang Y, Xiaoxiang Z, Tao WY (1990) Preliminary studies on kiwifruit
726 disease in Hunan province. *Sichuan Fruit Science and Technology* 18:28–
727 29.

728 18. Takikawa Y, Serizawa S, Ichikawa T, Tsuyumu S, Goto M (1989)
729 *Pseudomonas syringae* pv. *actinidiae* pv. nov.: The causal bacterium of
730 canker of kiwifruit in Japan. *Jpn J Phytopathol* 55(4):437–444.

731 19. Koh Y, Cha JB, Chung JH, Lee HD (1994) Outbreak and spread of bacterial
732 canker in kiwifruit. *Korean Journal of Plant Pathology* 10:68–72.

733 20. European and Mediterranean Plant Protection Organization (2011) EPPO
734 Reporting Service - Pests & Diseases. 1–21.

735 21. Marcelletti S, Ferrante P, Petriccione M, Firrao G, Scortichini M (2011)
736 *Pseudomonas syringae* pv. *actinidiae* draft genomes comparison reveal
737 strain-specific features involved in adaptation and virulence to *Actinidia*
738 species. *PLoS ONE* 6(11):e27297.

739 22. Butler MI, et al. (2013) *Pseudomonas syringae* pv. *actinidiae* from recent
740 outbreaks of kiwifruit bacterial canker belong to different clones that
741 originated in China. *PLoS ONE* 8(2):e57464.

742 23. McCann HC, et al. (2013) Genomic analysis of the kiwifruit pathogen
743 *Pseudomonas syringae* pv. *Actinidiae* provides insight into the origins of an
744 emergent plant disease. *PLoS Pathog* 9(7):e1003503.

745 24. Mazzaglia A, et al. (2012) *Pseudomonas syringae* pv. actinidiae (PSA)
746 isolates from recent bacterial canker of kiwifruit outbreaks belong to the
747 same genetic lineage. *PLoS ONE* 7(5):e36518.

748 25. Templeton MD, Warren BA, Andersen MT, Rikkerink EHA, Fineran PC
749 (2015) Complete DNA sequence of *Pseudomonas syringae* pv. *actinidiae*,
750 the causal agent of kiwifruit canker disease. *Genome Announcements*
751 3(5):e01054–15.

752 26. Fujikawa T, Sawada H (2016) Genome analysis of the kiwifruit canker
753 pathogen *Pseudomonas syringae* pv. *actinidiae* biovar 5. *Nature Publishing*
754 *Group* 6:21399–11.

755 27. O'Brien HE, et al. (2012) Extensive remodeling of the *Pseudomonas*
756 *syringae* pv. *avellanae* type III secretome associated with two independent

757 host shifts onto hazelnut. *BMC Microbiol* 12:141.

758 28. Wu X, et al. (2014) Deciphering the components that coordinately regulate
759 virulence factors of the soft rot pathogen *Dickeya dadantii*. *MPMI*
760 27(10):1119–1131.

761 29. Page F, et al. (2001) Osmoregulated periplasmic glucan synthesis is
762 required for *Erwinia chrysanthemi* Pathogenicity. *Journal of Bacteriology*
763 183(10):3134–3141.

764 30. Klosterman SJ, et al. (2011) Comparative genomics yields insights into
765 niche adaptation of plant vascular wilt pathogens. *PLoS Pathog*
766 7(7):e1002137–19.

767 31. Bontemps-Gallo S, et al. (2016) The *opgC* gene is required for OPGs
768 succinylation and is osmoregulated through RcsCDB and EnvZ/OmpR in
769 the phytopathogen *Dickeya dadantii*. *Nature Publishing Group*:1–14.

770 32. Spiers AJ, Kahn SG, Bohannon J, Travisano M, Rainey PB (2002) Adaptive
771 divergence in experimental populations of *Pseudomonas fluorescens*. I.
772 Genetic and phenotypic bases of wrinkly spreader fitness. *Genetics*
773 161(1):33–46.

774 33. Prada-Ramírez HA, et al. (2015) AmrZ regulates cellulose production in
775 *Pseudomonas syringae* pv. *tomato* DC3000. *Molecular Microbiology*
776 99(5):960–977.

777 34. Colombi E, et al. (2016) Evolution of copper resistance in the kiwifruit
778 pathogen *Pseudomonas syringae* pv. *actinidiae* through acquisition of
779 integrative conjugative elements and plasmids. *In review.*

780 35. Japanese Ministry of Agriculture, Forestry and Fisheries, Yokohama Plant
781 Protection Station, Research Division (2016) Pest risk analysis report on
782 *Pseudomonas syringae* pv. *actinidiae*. 1–24.

783 36. Mather AE, Reid S, Maskell DJ, Parkhill J (2013) Distinguishable epidemics
784 of multidrug-resistant *Salmonella* Typhimurium DT104 in different hosts.
785 *Science* 341. doi:10.1126/science.1241628.

786 37. Wagner DM, et al. (2014) *Yersinia pestis* and the plague of Justinian 541–
787 543 AD: A genomic analysis. *The Lancet Infectious Diseases* 14(4):319–326.

788 38. Andam CP, Worby CJ, Chang Q, Campana MG (2016) Microbial genomics of
789 ancient plagues and outbreaks. *TRENDS in Microbiology*:1–13.

790 39. Cauchemez S, et al. (2016) Unraveling the drivers of MERS-CoV
791 transmission. *Proc Natl Acad Sci USA* 113(32):9081–9086.

792 40. Almeida RPP, Nunney L (2015) How do plant diseases caused by *Xylella*
793 *fastidiosa* emerge? *Plant Disease* 99(11):1457–1467.

794 41. Schwartz AR, et al. (2015) Phylogenomics of *Xanthomonas* field strains
795 infecting pepper and tomato reveals diversity in effector repertoires and
796 identifies determinants of host specificity. *Front Microbiol* 6:208–17.

797 42. Clarke CR, et al. (2015) Genome-enabled phylogeographic investigation of
798 the quarantine pathogen *Ralstonia solanacearum* race 3 biovar 2 and
799 screening for sources of resistance against its core effectors.
800 *Phytopathology* 105(5):597–607.

801 43. Vinatzer BA, Monteil CL, Clarke CR (2014) Harnessing population
802 genomics to understand how bacterial pathogens emerge, adapt to crop
803 hosts, and disseminate. *Annu Rev Phytopathol* 52(1):19–43.

804 44. Stukenbrock EH, Bataillon T (2012) A population genomics perspective on
805 the emergence and adaptation of new plant pathogens in agro-ecosystems.
806 *PLoS Pathog* 8(9):e1002893.

807 45. Shapiro LR, et al. (2016) Horizontal gene acquisitions, mobile element
808 proliferation, and genome decay in the host-restricted plant pathogen
809 *Erwinia tracheiphila*. *Genome Biology and Evolution* 8(3):649–664.

810 46. Quibod IL, et al. (2016) Effector diversification contributes to
811 *Xanthomonas oryzae* pv. *oryzae* phenotypic adaptation in a semi-isolated
812 environment. *Scientific Reports* **6**:34137.

813 47. Singh RP, et al. (2011) The emergence of Ug99 races of the stem rust
814 fungus is a threat to world wheat production. *Annu Rev Phytopathol*
815 49:465–481.

816 48. Monteil CL, Yahara K, Studholme DJ, Mageiros L (2016) Population-
817 genomic insights into emergence, crop-adaptation, and dissemination of
818 *Pseudomonas syringae* pathogens. *Microbial Genomics* 2(10).
819 doi:10.1099/mgen.0.000089.

820 49. Sarkar SF, Guttman DS (2004) Evolution of the core genome of
821 *Pseudomonas syringae*, a highly clonal, endemic plant pathogen. *Applied*
822 *and Environmental Microbiology* 70(4):1999–2012.

823 50. Bankevich A, et al. (2012) SPAdes: A new genome assembly algorithm and
824 its applications to single-cell sequencing. *Journal of Computational Biology*
825 19(5):455–477.

826 51. Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for
827 Illumina sequence data. *Bioinformatics* 30(15):btu170–2120.

828 52. Mukherjee S, Huntemann M, Ivanova N, Kyrpides NC, Pati A (2015) Large-
829 scale contamination of microbial isolate genomes by Illumina PhiX control.
830 *Standards in Genomic Sciences* 10(1):18.

831 53. Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie
832 2. *Nat Meth* 9(4):357–359.

833 54. Li H (2011) A statistical framework for SNP calling, mutation discovery,
834 association mapping and population genetical parameter estimation from
835 sequencing data. *Bioinformatics* 27(21):2987–2993.

836 55. Garrison E, Marth G (2012) Haplotype-based variant detection from short-
837 read sequencing. *arXiv*.

838 56. Didelot X, Wilson DJ (2015) ClonalFrameML: Efficient inference of
839 recombination in whole bacterial genomes. *PLoS Computational Biology*
840 11(2):e1004041–18.

841 57. Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and
842 post-analysis of large phylogenies. *Bioinformatics*.

843 58. Richter M, Rosselo-Mora R, Glockner FO, Peplies J (2016) JSpeciesWS: a
844 web server for prokaryotic species circumscription based on pairwise
845 genome comparison. *Bioinformatics*:929–931.

846 59. Walker BJ, et al. (2014) Pilon: An integrated tool for comprehensive
847 microbial variant detection and genome assembly improvement. *PLoS ONE*
848 9(11):e112963–14.

849 60. Seemann T (2014) Prokka: rapid prokaryotic genome annotation.
850 *Bioinformatics* 30(14):2068–2069.

851 61. Page AJ, et al. (2015) Roary: rapid large-scale prokaryote pan genome
852 analysis. *Bioinformatics* 31(22):3691–3693.

853 62. Krzywinski M, Schein J, Birol I, Connors J (2009) Circos: an information
854 aesthetic for comparative genomics. *Genome Research* **19**(9):1639-45.

855