TITLE

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

AUTHORS

Honour C. McCann[a,c,1,3], Li Li[b,1], 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,2] & Hongwen Huang[b,c,2,3]

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

[1] Co-first authors

[2] Co-senior authors

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

ABSTRACT

Recurring epidemics of kiwifruit (*Actinidia* spp.) 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 sampled 746 *Pseudomonas* isolates from cultivated and wild kiwifruit across six provinces in China, of which 87 were *Psa*. Of 234 *Pseudomonas* isolated from wild *Actinidia* spp. none were identified as *Psa*. 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.

INTRODUCTION

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

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

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

chinensis (e.g. 'Hort16A') began only in the 2000s and an outbreak of global proportions soon followed. The first published notices of the latest outbreak on gold kiwifruit issued from Italy in 2008, with reports from neighbouring European countries, New Zealand, Asia and Chile occurring soon after (1-6, 20). Whole genome sequencing showed the most recent global outbreak of disease was caused by a new lineage of *Psa* (previously referred to as *Psa*-V and now referred to as *Psa*-3), while earlier disease incidents in Japan and Korea were caused by strains forming separate clades referred to *Psa*-1 (previously *Psa*-J) and *Psa*-2 (previously *Psa*-K), respectively) (21-24). These lineages are marked by substantial variation in their complement of type III secreted effectors, which are required for virulence in *P. syringae.* Despite the surprising level of within-pathovar differences in virulence gene repertoires occurring subsequent to the divergence of these three lineages, strains from each lineage are capable of infecting and growing to high levels in both *A. chinensis* var*. deliciosa* and *A. chinensis* var*. chinensis* (23).

The severity of the latest global outbreak is largely predicated on the expansion in cultivation of clonally propagated highly susceptible *A. chinensis var. chinensis* cultivars, with trade in plant material and pollen likely providing opportunities for transmission between distant geographic regions. Identifying the source from which *Psa* emerged to cause separate outbreaks remains an important question. Intriguingly, despite the divergence in both the core and flexible genome, these distinct clades nevertheless exhibit evidence of recombination with each other and unknown donors (23). This suggests each lineage of *Psa* emerged from a recombining source population. Definitive evidence for the location, extent of diversity and evolutionary processes

118 operating within this population remain elusive. Early reports suggested China
119 may be the source of the latest global outbreak (22, 23). Although the strains of
120 *Psa* available at that time did not provide unambiguous and well-supported
121 evidence of a Chinese origin, this speculation was based on the fact that kiwifruit
122 are native to China; it is the provenance of the plant material selected for
123 commercial and breeding purposes in China, New Zealand, Italy and other
124 kiwifruit growing regions; there is extensive trade in plant material between all
125 of these regions; and one Chinese isolate was found to carry an integrative and
126 conjugative element (ICE) that was also found in New Zealand *Psa*-3 isolates
127 (22).

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

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

144

145 **RESULTS**

146 **The phylogeography of *Psa***

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

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

171 In order to obtain greater resolution of the relationships between the new
172 Chinese and pandemic isolates, we identified the 4,853,421bp core genome of all
173 62 strains in *Psa*-3. The core genome includes both variant and invariant sites
174 and excludes regions either unique to or deleted from one or more strains. To
175 minimize the possibility of recombination affecting the reconstruction of
176 evolutionary relationships and genetic distance within *Psa*-3, ClonalFrameML
177 was employed to identify and remove SNPs with a high probability of being
178 introduced by recombination rather than mutation. The within-lineage ratio of
179 recombination to mutation (R/theta) is reduced in *Psa*-3 ($6.75 \times 10^{-2} \pm 3.24 \times 10^{-5}$) relative to between lineage rates ($1.27 \pm 5.16 \times 10^{-4}$) ,and the mean divergence
181 of imported DNA within *Psa*-3 is $8.54 \times 10^{-3} \pm 5.18 \times 10^{-7}$ compared to $5.68 \times 10^{-3} \pm 1.04 \times 10^{-8}$ between lineages. Although recombination has occurred within *Psa*
183 lineage 3, it is less frequent and has introduced fewer polymorphisms relative to
184 mutation: when accounting for polymorphisms present in recombinant regions
185 identified by ClonalFrameML and/or present on transposons, plasmids, and
186 other mobile elements, more than seven-fold more polymorphisms were
187 introduced by mutation relative to recombination (Table 1). Recombination has
188 a more pronounced impact between clades, where substitutions are slightly
189 more likely to have been introduced by recombination than by mutation (Table
190 1).

The source of pandemic *Psa*

Data show that there is greater diversity among the Chinese *Psa*-3 population than had been previously identified (Figure 2). Interestingly, clades defining *Psa*-1 and *Psa*-3 exhibit similar levels of diversity (Table S2). These clades share a common ancestor: assuming they are evolving at a similar rate, they may have been present in Japan and China for a similar duration. The strains isolated during the latest pandemic in Italy (I2, I3, I10, I11, I13), Portugal (P1), New Zealand (NZ13, NZ31-35, NZ37-43, NZ45-49, NZ54), Chile (CL4), Japan (J38, J39) and Korea (K5) during the latest kiwifruit canker pandemic cluster with nine Chinese isolates (C1, C3, C29-31, C62, C67-69) (Figure 2). This pandemic lineage exhibits little diversity at the level of the core genome, having undergone clonal expansion only very recently. The NZ isolates form a monophyletic group and share a common ancestor, indicating there was a single transmission event of *Psa* into NZ. Two recently isolated Japanese pandemic *Psa*-3 isolated in 2014 group within the New Zealand isolates, suggesting the pandemic lineage may have been introduced into Japan via New Zealand (Figure 2). Italian and Portuguese pandemic strains also form a separate group, indicative of a single transmission event from China to Italy. China is undoubtedly the source of the strains responsible for the pandemic of kiwifruit canker disease, yet the precise origins of the pandemic subclade remain unclear. Isolates from four different provinces in Western China (Guizhou, Shaanxi, Sichuan and Chongqing) are represented among the pandemic lineage, indicating extensive regional transmission within China after emergence of the pandemic. Yet each province harbouring pandemic isolates also harbors basally diverging

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

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

241 year. The relatively slow substitution rate and the strong bottleneck effect
242 experienced during infections hinders efforts to reconstruct patterns of
243 transmission, as the global dissemination of a pandemic strain may occur
244 extremely rapidly (27, 28). The estimated divergence time of *Psa* broadly
245 considered is likely older than the pandemic and epidemic events with which
246 they are associated: the earliest report of disease cause by lineage 1 occurred in
247 1984 and the first report of infection from the latest pandemic was issued in
248 2008.
249

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

251 2,214 SNPs mapping to the core genome of *Psa*-3 were identified; 263 of
252 these mapped to recombinant regions identified by ClonalFrameML and/or
253 plasmid, prophage, integrative and conjugative elements, transposons and other
254 mobile genetic elements (Table 1, Figure 4). The highest density of
255 polymorphism occurs in and around the integrative and conjugative element
256 (ICE) in *Psa* NZ13 (Figure 4). Of the 1,951 SNPs mapping to the non-recombinant
257 non-mobile core genome, 57.7% (1,125) are strain specific. Most strain-specific
258 SNPs are found in the two most divergent members of the lineage: *Psa* C16 and
259 C17, with 736 and 158 strain-specific SNPs, respectively. The remaining isolates
260 have an average of 3.9 strain-specific SNPs, ranging from 0 to 44 SNPs per
261 strains. There are 826 SNPs shared between one or more *Psa*-3 strains. The
262 pandemic clade differs from the more divergent Chinese strains by 72 shared
263 SNPs. Within the pandemic lineage there are 125 strain-specific SNPs, an average
264 of 3.1 unique SNPs per strain (ranging from 0-27 SNPs) and an additional 29
265 SNPs shared among pandemic strains. Protein-coding sequence accounts for

266 88.4% of the non-recombinant, gap-free core genome of this lineage. We
267 observed that 78.9% (1,539/1,951) of mutations occurred in protein coding
268 sequence, significantly different from the expectation (1,725/1,963) in the
269 absence of selection (Pearson's χ^2 test: $P < 0.0001$, $\chi^2 = 172.55$). This suggests
270 there is selection against mutations occurring in protein coding sequences. Of
271 the 953 non-synonymous substitutions introduced by mutation in *Psa*-3, 927
272 resulted in amino acid substitutions, two resulted in extensions and 24 resulted
273 in premature truncations.

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

285 Two substitutions are shared exclusively by the European pandemic
286 strains (AKT28710.1 G1150Aand AKT33438.1 T651C) and one silent
287 substitution in a gene encoding an acyltransferase superfamily protein
288 (AKT31915.1 C273T) is shared among the European pandemic and six of nine
289 Chinese pandemic strains (C3, C29-31, C67, C69). As these six Chinese pandemic
290 strains were isolated from Shaanxi, Sichuan and Chongqing, they do not provide

any insight into the precise geographic origins of the European pandemic *Psa-3*, though transmission from China to Italy is likely concomitant with dissemination of the pandemic lineage across China. Six conserved and diagnostic polymorphisms are present in the pandemic New Zealand and Japanese isolates (Table S4). One of these is a silent substitution in an ion channel protein (AKT31947.1 A213G), another is an intergenic (T->G) mutation at position 362,522 of the reference *Psa* NZ13 chromosome and the remaining four are nonsynonymous substitutions in an adenylyltransferase (AKT32845.1, W977R); chromosome segregation protein (AKT30494.1, H694Q); cytidylate kinase (AKT29651.1, V173L) and peptidase protein (AKT32264.1, M418K).

The type III secretion system is known to be required for virulence in *P. syringae*. A 44,620bp deletion event in *Psa* C17 resulted in the loss of 42 genes encoding the structural apparatus and conserved type III secreted effectors in *Psa* C17. This strain is highly compromised in its ability to grow in *A. chinensis* var. *deliciosa* 'Hayward', attaining 1.2 x 10^7 cfu/g three days post inoculation (dpi) and declining to 8.8 x 10^4 cfu/g at fourteen dpi (Figure S1). This is a marked reduction compared to *Psa* NZ13, which attains 3.0 x 10^9 and 4.2 x 10^7 cfu/g three and fourteen dpi. *Psa* C17 nevertheless multiplies between day 0 and day 3, indicating that even in the absence of type III-mediated host defense disruption, *Psa* may still proliferate in host tissues. The loss of the TTSS does not inhibit the growth of *Psa* C17 as strongly in the more susceptible *A. chinensis* var. *chinensis* 'Hort16A' cultivar.

Two potentially significant deletion events occurred in the ancestor of the pandemic subclade: a frameshift caused by a mutation and single base pair deletion in a glucan succinyltransferase (*opgC*) and a 6,456bp deletion in the *wss*

316 operon (Figure S2). Osmoregulated periplasmic glucans (OPGs, in particular
317 *opgG* and *opgH*) are required for motility, biofilm formation and virulence in
318 various plant pathogenic bacteria and fungi (29-31). Homologs of *opgGH* remain
319 intact in the pandemic subcclade, yet the premature stop mutation in *opgC* likely
320 results in the loss of glucan succinylation. The soft-rot pathogen *Dickeya dadantii*
321 expresses OpgC in high osmolarity conditions, resulting in the substitution of
322 OPGs by O-succinyl residues(32). *D. dadantii opgC* deletion mutants did not
323 display any reduction in virulence(32). *Psa* is likely to encounter high osmolarity
324 during growth and transport in xylem conductive tissues, yet the impact of the
325 loss of *opgC* on *Psa* fitness has yet to be determined. The most striking difference
326 between the pandemic subclade and more divergent Chinese *Psa*-3 strains is the
327 deletion of multiple genes involved in cellulose production and acetylation of the
328 polymer (Figure S2)(33). The loss of cellulose production and biofilm production
329 is not associated with a reduction in growth or symptom development of *P.*
330 *syringae* pv. *tomato* DC3000 on tomato, but may enhance bacterial spread
331 through xylem tissues during vascular infections (34). In *P. fluorescens* SBW25
332 deletion of the Wss operon significantly compromises ability to colonise plant
333 surfaces and in particular the phyllosphere of sugar beet (*Beta vulgaris*)
334 seedlings (35). It is possible that loss of this locus aids movement through the
335 vascular system and / or dissemination among plants, by limiting capacity for
336 surface colonization and biofilm formation.
337

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

339 Despite the high similarity within the core genome, extensive variation is
340 evident in the pangenome of *Psa*-3 (Figure 5). The core genome (4,339 genes in

341 99-100% of strains, and 674 genes in 95%-99% of strains, or 58-62 genomes)
342 comprises 50.5% of the total pangenome (9,931 genes). 674 genes are present in
343 15-95% of strains (9-57 genomes), the so-called 'shell genes' (Figure 5). The
344 flexible genome is comprised of the 'shell' and 'cloud' genes; the latter describes
345 genes present in 0-15% of strains (one to six genomes in this case). Cloud genes
346 contribute most to the flexible genome: 3,950 genes are present in one to six
347 strains. This is a striking amount of variation in a pathogen described as clonal
348 and monomorphic. It should be noted that sequencing and assembly quality will
349 impact annotation and pangenome estimates: omitting the low quality J39
350 assembly results in a core and soft-core genome differing by 18 genes and a
351 reduction of the cloud by 275. Despite a relatively slow rate of mutation and
352 limited within-clade homologous recombination, the amount of heterologous
353 recombination demonstrates that the genomes of these pathogens are highly
354 labile. Mobile genetic elements like bacteriophage, transposons and integrases
355 make a dramatic contribution to the flexible genome. Integrative and conjugative
356 elements (ICEs) are highly mobile elements and have recently been
357 demonstrated to be involved in the transfer of copper resistance in *Psa* (36).
358 Prodigious capacity for lateral gene transfer creates extreme discordance
359 between ICE type, host phylogeny and host geography making these regions
360 unsuitable markers of host evolution and origin.

361 Three divergent ICEs have been previously described from the global
362 pandemic lineage (23). Within *Psa*-3 ICEs were found in 53 of 62 isolates (nine
363 of the divergent Chinese isolates were devoid of any such element) (Figure 2). No
364 phylogeographic signal is evident. For example, strains from Sichuan, Shaanxi,
365 Korea, Italy and Portugal share an identical ICE. Even within a single Chinese

366 provence, mulitple ICEs exist (Shaanxi and Sichuan isolates harbour four and
367 three different ICEs, respectively). Moreover, ICE host range is not limited to Psa
368 alone: the ICE found in every NZ isolate (and also recorded in Chinese isolate C1)
369 exists in essentially identical form in a strain of *P. syringae* pv. *avellenae*
370 CRAPAV013 isolated from hazelnut in 1991 in Latina, Italy (it exhibits 98%
371 pairwise identity, differing from the New Zealand ICE by a transposon, 66bp
372 deletion and a mere 6 SNPs).

DISCUSSION

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

398 *Psa*-3 detection in both those countries, though the amount of pollen imported
399 from New Zealand in 2012 (349kg) and 2013 (190kg) far outweighed the
400 amount imported from China (1 kg in both 2012 and 2013) (37).

401 Our phylogeographic study of a single lineage giving rise to a pandemic in
402 *P. syringae* has revealed far greater diversity than was previously appreciated.
403 Extensive diversity between *Psa* isolates collected from *Actinidia* spp. was
404 observed in the same province. The amount of diversity present within lineage 3
405 indicates this population was present and circulating in China before the
406 pandemic began. The emergence of the pandemic subclade moreover has not
407 resulted in the replacement of more ancestral strains: both pandemic and
408 divergent lineage 3 *Psa* were isolated from four out of six provinces.

409 Strains from three different lineages have been isolated in both Korea and
410 Japan, while China harbours strains from only a single lineage. The most basal
411 lineages of canker-causing *Psa* are comprised of Korean strains isolated between
412 1997 and 2014 (*Psa*-2) and a member of the recently identified lineage *Psa*-5.
413 One early isolate (*Psa* K3, 1997) groups with the Japanese isolates in lineage 1,
414 and a more recent Korean isolate *Psa* K7 (2014) groups with the more diverse
415 Chinese isolates in lineage 3. Korea therefore harbours a more diverse
416 population of *Psa* than China, with strains from three distinct lineages of *Psa* (1, 2
417 and non-pandemic subclade 3). A novel group of *Psa* was recently identified in
418 Japan (*Psa*-5) and appears to share an ancestor with the Korean *Psa*-2 strains.
419 With the recent dissemination of pandemic *Psa*-3 and the historical presence of
420 *Psa*-1 in Japan, the Japanese population of *Psa* is comprised of three distinct
421 lineages of *Psa* (1, 5 and pandemic subclade 3). Though no strains from *Psa*-1
422 have been isolated in either Japan or Korea since 1997, at least two lineages

423 currently coexist in both Japan and Korea. This strongly suggests that the source
424 population of all *Psa* is not China, but likely resides in either Korea or Japan. The
425 potential transmission of a non-pandemic lineage 3 strain from China to Korea
426 and the emergence of a new lineage in Japan supports our earlier assertion that
427 variants will continue to emerge to cause local epidemics and global pandemics
428 in the future (23).

429 Considering that the divergence time of this monophyletic pathovar
430 predates the commercialisation of kiwifruit by hundreds if not thousands of
431 years, *Psa* is likely associated with a non-domesticated host(s) in the wild. Both
432 *A. chinensis var. deliciosa* or *A. chinensis var. chinensis* are found in natural
433 ecosystems and have overlapping habitat ranges with cultivated kiwifruit in
434 many areas. However, despite isolating 746 *Pseudomonas* strains from both wild
435 and cultivated kiwifruit during this sampling program, we did not identify *Psa*
436 among any of the 188 *Pseudomonas* spp. isolated from 98 wild *A. chinensis var.*
437 *deliciosa* or *A. chinensis var. chinensis* sampled across six provinces in China
438 (Table S5). Very few *Actinidia* spp. have ranges extending to South Korea and
439 Japan: *A. arguta*, *A. kolomikta*, *A. polygama* and *A. rufa*. *A. arguta* are broadly
440 distributed across both Korea and Japan. Early work by Ushiyama *et al.* (1992)
441 found that *Psa* could be isolated from symptomatic *A. arguta* plants in Japan. The
442 possibility that this wild relative of kiwifruit harbours diverse strains of *Psa* that
443 may emerge to cause future outbreaks is currently under investigation.
444 Alternately, a host shift from another domesticated crop may have occurred after
445 expansion in kiwifruit cultivation.
446

447 Numerous epidemiological studies of human pathogens have
448 demonstrated environmental or zoonotic origins, but there are few such studies
449 of plant pathogens (23, 38-48). Where ecological and genetic factors restrict
450 pathogens to a small number of plant hosts some progress has been made, but
451 for facultative pathogens such as *P. syringae* that colonise multiple hosts and are
452 widely distributed among both plant and non-plant habitats, the environmental
453 reservoirs of disease and factors affecting their evolutionary emergence are
454 difficult to unravel (49, 50).

455 The emergence of *Psa* over the last three decades – concomitant with
456 domestication of kiwifruit– offers a rare opportunity to understand the
457 relationship between wild populations of both plants and microbes and the
458 ecological and evolutionary factors driving the origins of disease, including the
459 role of agriculture. It is now possible to exclude China as the native home to the
460 source population, but the precise location remains unclear. Nonetheless, it is
461 likely, given the extent of diversity among *Psa* isolates and the time-line to
462 domestication, that ancestral populations exist in non-agricultural plant
463 communities. Attention now turns to Korea and Japan and in particular the
464 interplay between genetic and ecological factors that have shaped Psa evolution.
465
466
467

MATERIALS AND METHODS

Bacterial strains and sequencing

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

Initial strain identification was performed by sequencing the citrate synthase gene (*cts*, aka *gltA* (51)). Subsequent to strain identification, paired-end sequencing was performed using the Illumina HiSeq 2500 platform (Novogene, Guangzhou, China). Additional paired-end sequencing was performed at New Zealand Genomics Limited (Auckland, New Zealand) using the MiSeq platform, and raw sequence reads from some previously published isolates were shared by Mazzaglia *et al.* (24).

Variant Calling and Recombination Analyses

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

Read data was corrected using the SPADEs correction module and Illumina adapter sequences were removed with Trimmomatic allowing 2 seed mismatches, with a palindrome and simple clip threshold of 30 and 10, respectively (52, 53). Quality-based trimming was also performed using a sliding window approach to clip the first 10 bases of each read as well as leading and trailing bases with quality scores under 20, filtering out all reads with a length under 50 (53). PhiX and other common sequence contaminants were filtered out using the Univec Database and duplicate reads were removed (54).

Reads were mapped to the complete reference genome *Psa* NZ13 with Bowtie2 and duplicates removed with SAM Tools (55, 56). Freebayes was used to call variants with a minimum base quality 20 and minimum mapping quality 30 (57). Variants were retained if they had a minimum alternate allele count of 10 reads and fraction of 95% of reads supporting the alternate call. The average coverage was calculated with SAM Tools and used as a guide to exclude overrepresented SNPs (defined here as threefold higher coverage than the average) which may be caused by mapping to repetitive regions. BCFtools filtering and masking was used to generate final reference alignments including SNPs falling within the quality and coverage thresholds described above and

516 excluding SNPs within 3bp of an insertion or deletion (indel) event or indels
517 separated by 2 or fewer base pairs. Invariant sites with a minimum coverage of
518 10 reads were also retained in the alignment, areas of low (less than 10 reads) or
519 no coverage are represented as gaps relative to the reference.
520 Freebayes variant calling includes indels and multiple nucleotide
521 insertions as well as single nucleotide insertions, however only SNPs were
522 retained for downstream phylogenetic analyses. An implementation of
523 ClonalFrame suitable for use with whole genomes was employed to identify
524 recombinant regions using a maximum likelihood starting tree generated by
525 RaxML (58, 59). All substitutions occurring within regions identified as being
526 introduced due to recombination by ClonalFrameML were removed from the
527 alignments. The reference alignments were manually curated to exclude
528 substitutions in positions mapping to mobile elements such as plasmids,
529 integrative and conjugative elements and transposons.
530

531 **Phylogenetic Analysis**

532 The maximum likelihood phylogenetic tree of 80 *Psa* strains comprising
533 new Chinese isolates and strains reflecting the diversity of all known lineages
534 was built with RAxML (version 7.2.8) using a 1,216,321bp core genome
535 alignment excluding all positions for which one or more genomes lacked
536 coverage of 10 reads or higher (59). The core genome alignmnt included 2,207
537 variant sites. Membership within each phylogenetic lineage corresponds to a
538 minimum average nucleotide identity of 99.70%. The average nucleotide identity
539 was determined using a BLAST-based approach in JspeciesWS (ANIb), using a
540 subset of 32 *Psa* genome assemblies spanning all lineages(60). In order to fully

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

Identification of the core and mobile genome

Genomes were assembled with SPAdes using the filtered, trimmed and corrected reads (52). Assembly quality was improved with Pilon and annotated with Prokka (61, 62). The pangenome of *Psa*-3 was calculated using the ROARY pipeline(63). Orthologs present in 61 (out of a total of 62) genomes were considered core; presence in 58-60, 9-57 and 1-8 were considered soft-core, shell and cloud genomes, respectively. BLAST-based confirmation was used to confirm the identity predicted virulence or pandemic-clade-restricted genes in genome assemblies.

Pathogenicity assays

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

565 significance between each treatment at each time point was assessed using two-
566 tailed t-tests with uneven variance.
567

568 ACKNOWLEDGEMENTS

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

TABLES

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

588 **Table S1. Strains**
589

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

[1]Simulated reads were generated from contigs available for these previously sequenced genomes

[2]Different sequencing runs were employed for draft genome assembly and variant calling

592

593 **Table S2. Average percent identity within and between Psa lineages**
594

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]

595
596 [1]Not determined for Psa-5 as only a single strain has been sequenced from this lineage. ANIb values
597 determined using representative strains for *Psa*-1 (J31, K3, J2, J1, J30, J29, J35, J32, J33, I1, J36,
598 J25), *Psa*-2 (K27, K6, K4, K26, K28), *Psa*-3 (C16, C17, C10, C11, C70, K7, C15, C54, C74, C69, K5,
599 C62, I13, NZ13) and *Psa*-5 (J37)
600

601 **Table S3. Genes with multiple mutations in *Psa*-3**
602 (See Table_S3.xlsx)

603

604 **Table S4. SNPs shared between all pandemic NZ and Japanese isolates**
605

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)

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

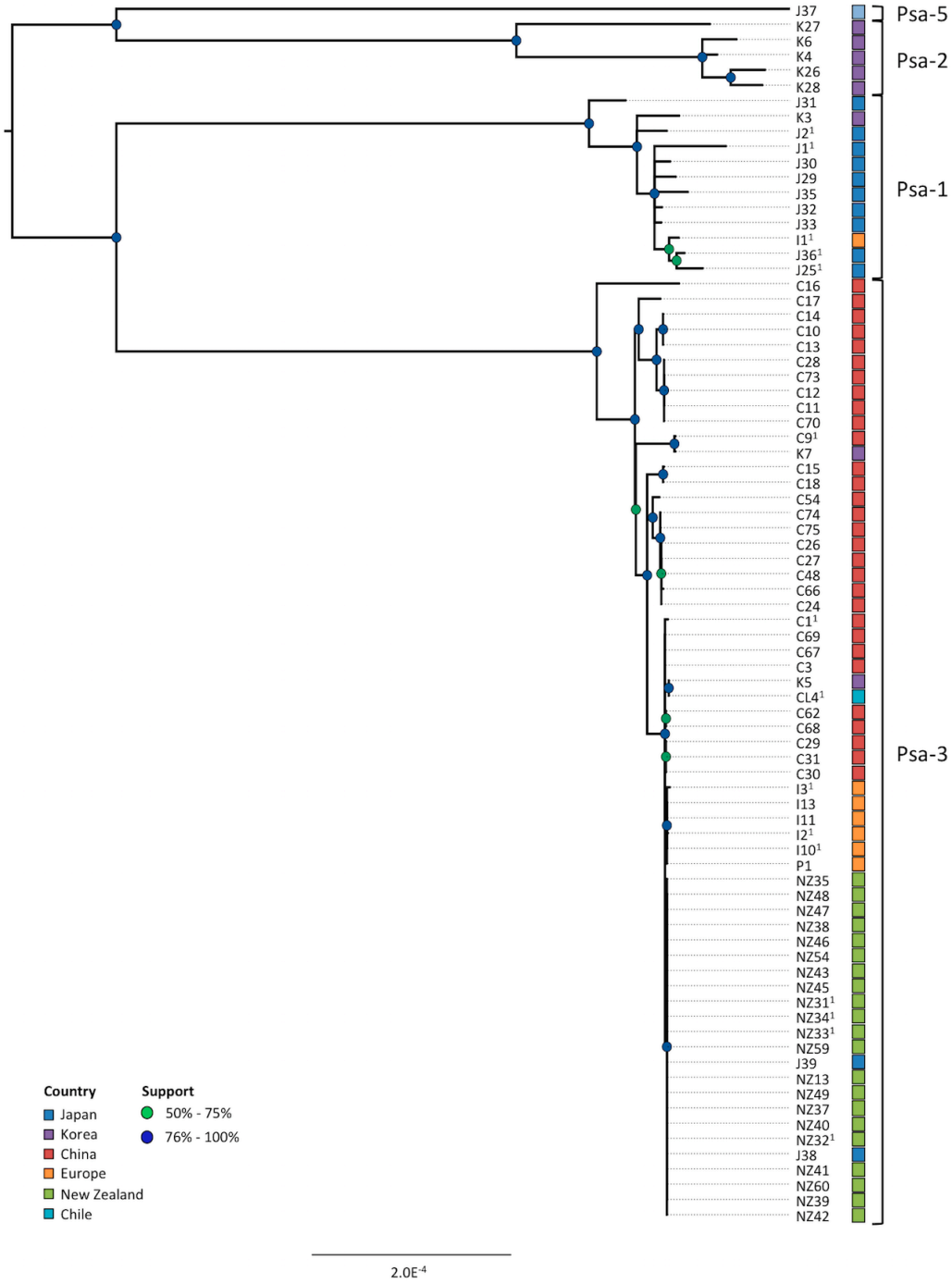
	A. arguta	*A. callosa*	*A. chinensis* var. *chinensis*	*A. chinensis* var. *deliciosa*	*Actinidia* sp.	*Camellia* sp.	*Prunus* sp.	Total
Cultivated	3		373	350	1	45	45	817
Disease	3		323	244		30	45	645
Not determined	2		121	81		20	10	234
Pseudomonas spp.	1		123	120		8	35	287
P. syringae			12	23		2		37
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

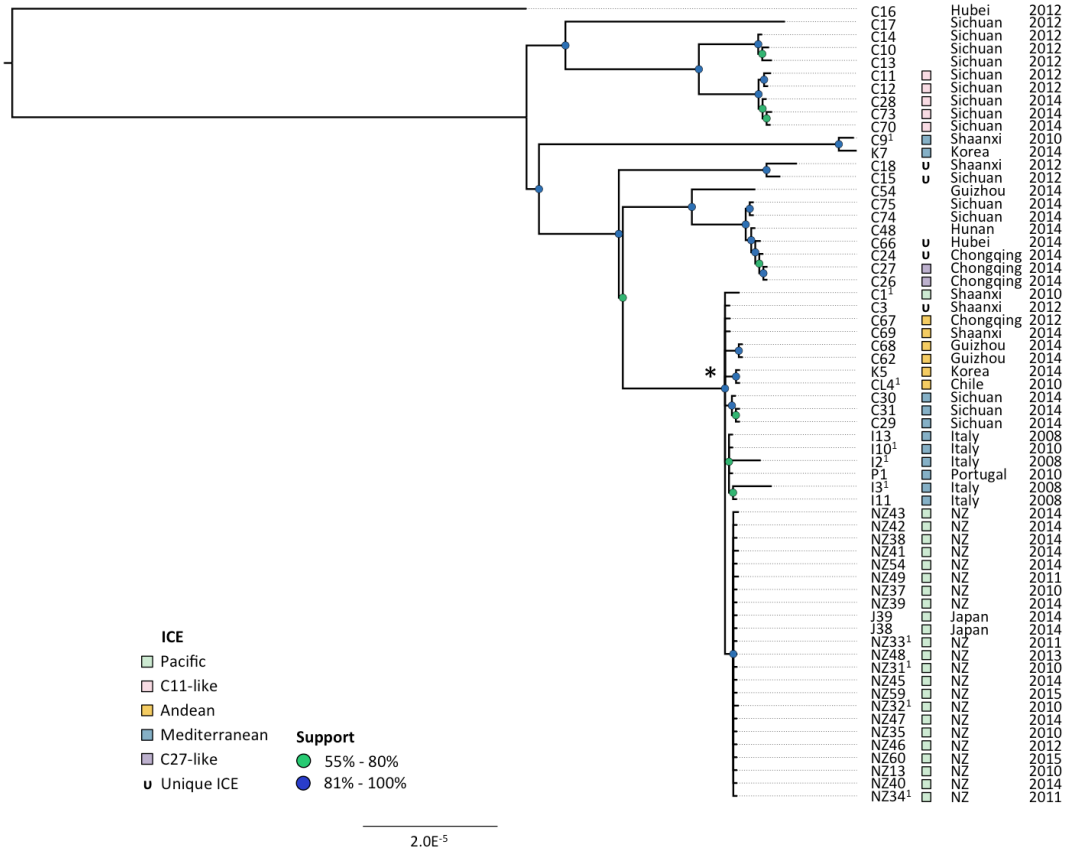
607
608
609

FIGURES

Figure 1. Phylogeny of all *Psa*



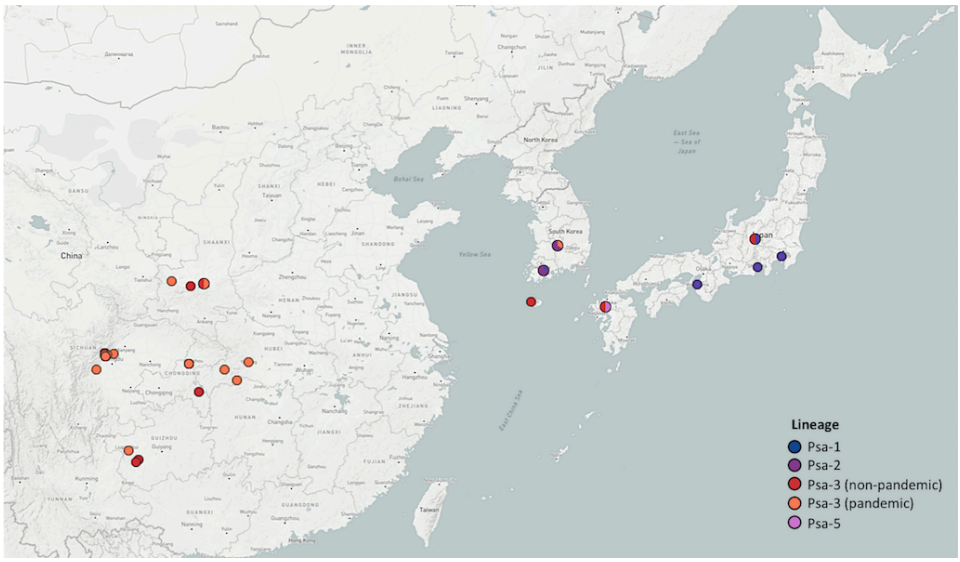
614 **Figure 2. Phylogeny of *Psa* lineage 3**



615

616 **Figure 3. Psa isolation locations in East Asia**

617

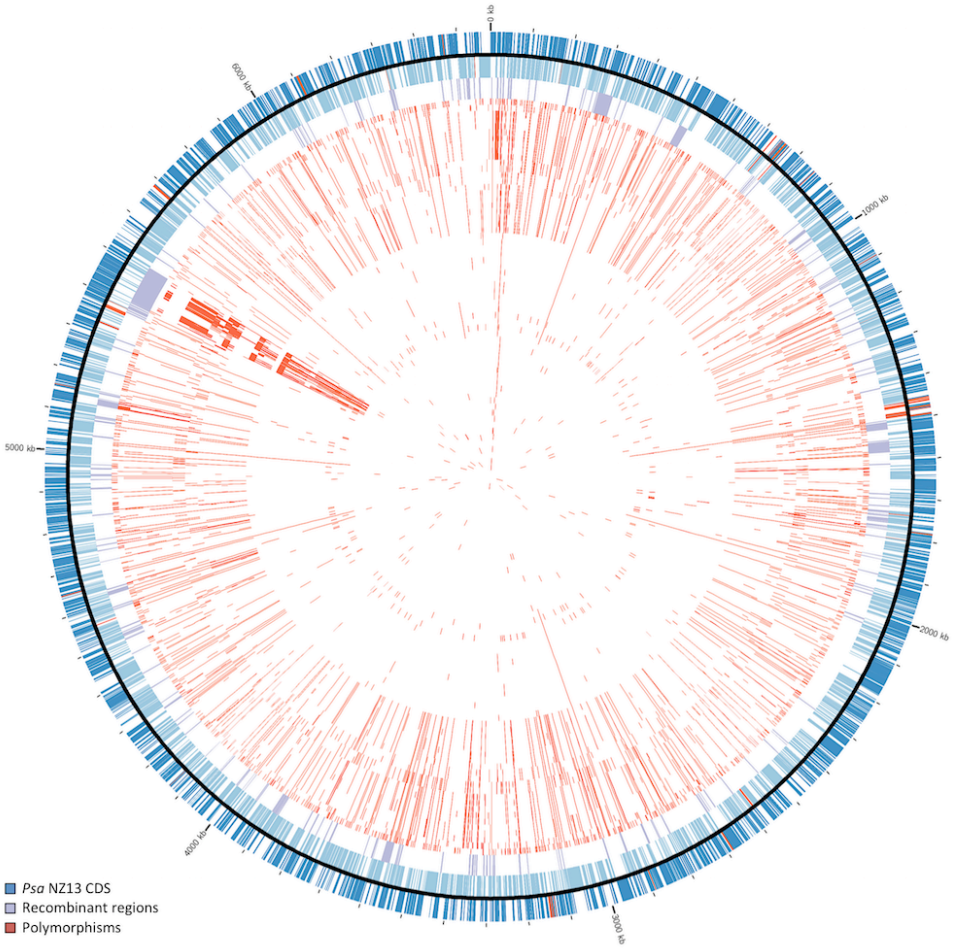


618

619

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

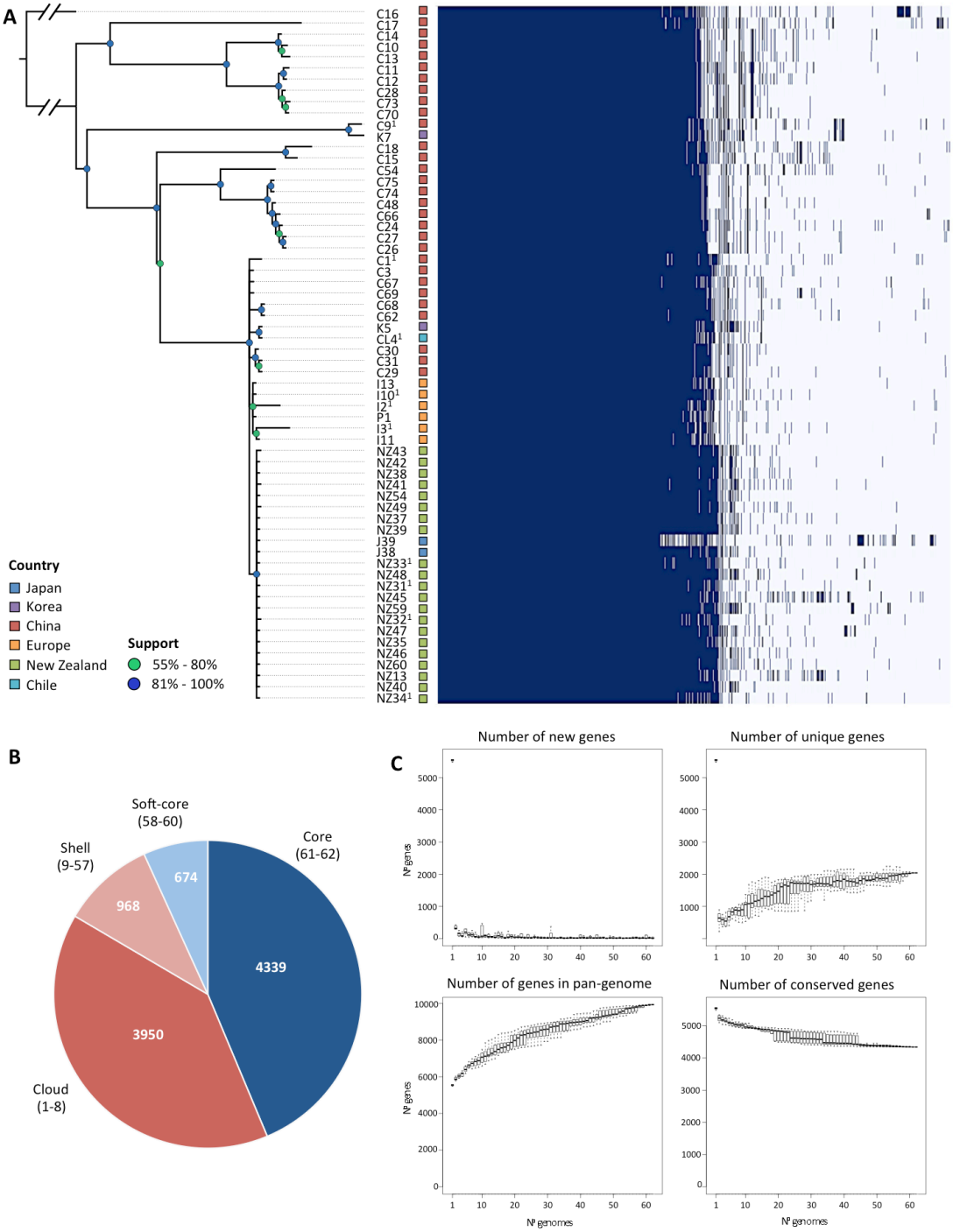
621



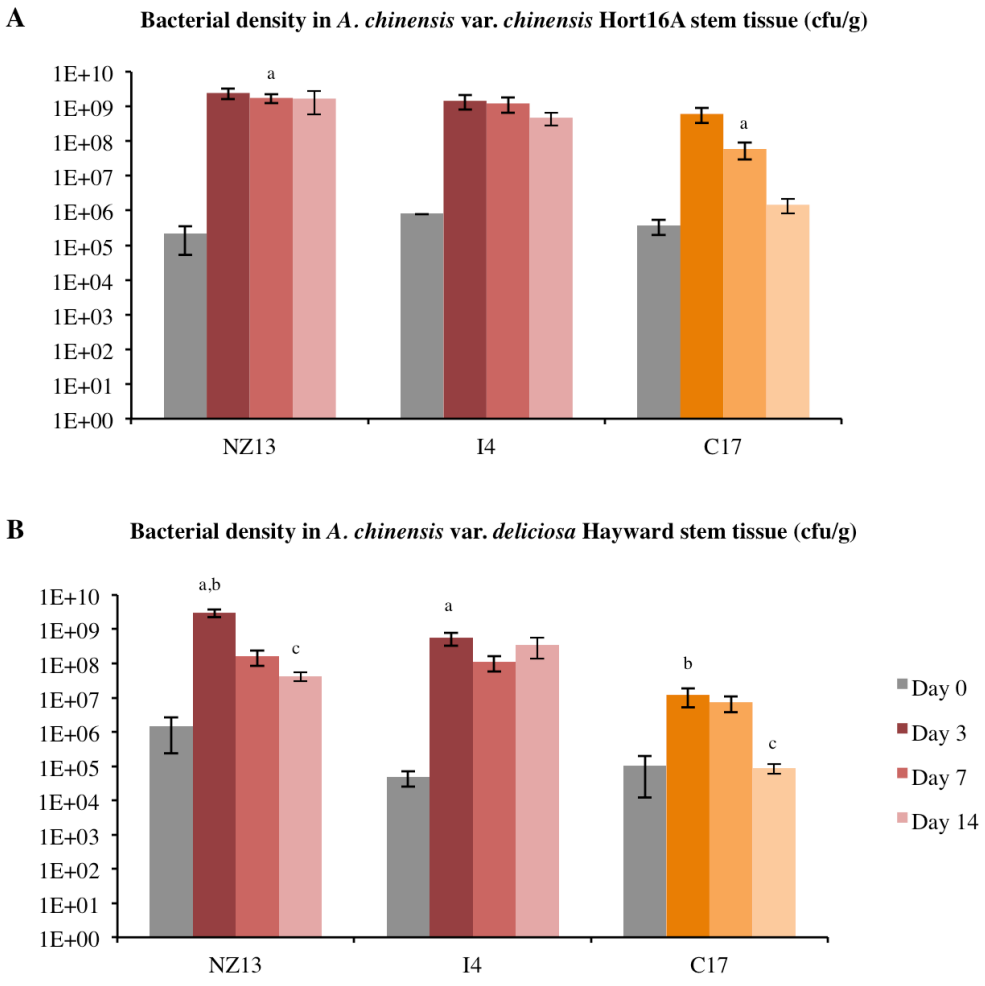
622

623

Figure 5. Pangenome of *Psa*-3

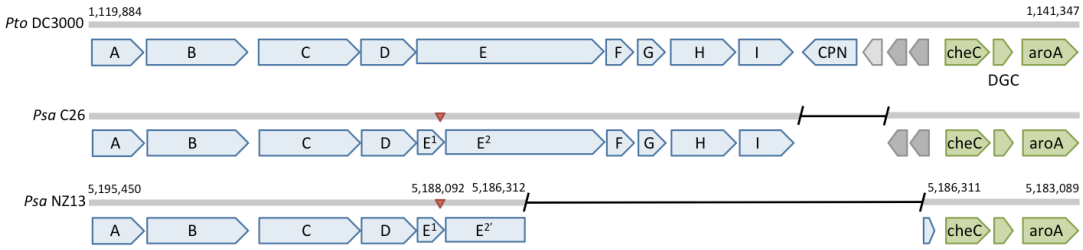


629 **Figure S1. Bacterial growth assay of *Psa* on *A. chinensis***
630



631

632 **Figure S2. *Wss* operon disruption in *Psa*-3**
633



634

635 # FIGURE LEGENDS

636 **Figure 1. Phylogeny of *Psa***

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

641

642 **Figure 2. Phylogeny of *Psa* lineage 3**

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

648

649 **Figure 3. Psa-3 isolation locations in East Asia**

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

654

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

656 Polymorphisms and recombinant regions mapped onto *Psa* NZ13 reference
657 genome using CIRCOS(64). *Psa* NZ13 CDS are displayed in the first and second

658 ring (blue), with annotated Type 3 secretion system and effectors highlighted
659 (red). Inner rings display polymorphisms in *Psa*-3 genomes ordered from most
660 to least divergent relative to *Psa* NZ13. The integrative and conjugative element
661 (ICE) is the most polymorphic region.
662

663 **Figure 5. Pangenome of *Psa*-3**
664 A. Presence/absence matrix of all core and accessory genes in *Psa*-3, ordered
665 according to strains' phylogenetic relationships. Country of isolation is indicated
666 at left. B. The core and flexible genome of *Psa*-3. The core, soft-core, shell and
667 cloud genomes are defined according to the numbers in parentheses. C. Number
668 of new, unique and conserved genes with addition of each genome.
669

670 **Figure S1. Bacterial growth assay on *A. deliciosa***
671 Bacterial growth of pandemic *Psa* NZ13, I4 (red) and divergent C17 (orange)
672 strains on *A. chinensis* var. *chinensis* 'Hort16A' and *A. chinensis* var. *deliciosa*
673 'Hayward'. Mean *in planta* bacterial density in stem tissue (cfu/g) at 0, 3 and 7
674 days post-inoculation is shown (mean ± SEM) with superscript denoting
675 significant difference between strains at each sampling time ($P<0.05$, two-tailed
676 t-test, unequal variance). Four replicate plants were assayed at day 0, and six
677 replicates at each subsequent time point.
678

679 **Figure S2. Wss operon disruption in *Psa*-3**
680 Genes encoding components of the *wss* operon (blue), hypothetical and
681 conserved hypothetical (light and dark grey), chemotaxis, diguanylate cyclase

682 and *aroA* (green). Deletions (black line) and position of single base pair insertion
683 (red triangle) displayed with reference to *Pto* DC3000. Insertion results in
684 frameshift mutation in *wssE*, two predicted derivatives annotated as *wssE1* and
685 *wssE2*. The subsequent 6.5kb deletion in the ancestor of the pandemic subclade
686 results in the truncation of *wssE2*, annotated as *wssE2'*.

687 **REFERENCES**

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

691 2. Vanneste JL, et al. (2011) First Report of *Pseudomonas syringae* pv.
692 *actinidiae*, the Causal Agent of Bacterial Canker of Kiwifruit in France.
693 *Plant Disease* 95(10):1311–1311.

694 3. Balestra GM, Renzi M, Mazzaglia A (2010) First report of bacterial canker
695 of *Actinidia deliciosa* caused by *Pseudomonas syringae* pv. *actinidiae* in
696 Portugal. *New Disease Reports* 22: 10.

697 4. Abelleira A, et al. (2011) First Report of Bacterial Canker of Kiwifruit
698 Caused by *Pseudomonas syringae* pv. *actinidiae* in Spain. *Plant Disease*
699 95(12):1583–1583.

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

703 6. Koh YJ, et al. (2012) Occurrence of a New Type of *Pseudomonas syringae*
704 pv. *actinidiae* Strain of Bacterial Canker on Kiwifruit in Korea. *The Plant*
705 *Pathology Journal* 28(4):423–427.

706 7. Zhao ZB, Gao XN, Huang QL, Huang LL, Qin HQ (2013) Identification and
707 characterization of the causal agent of bacterial canker of kiwifruit in the
708 Shaanxi province of China. *Journal of Plant Pathology* 95(1): 155-162.

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

713 9. Koh YJ, Jung JS, Hur JS (2002) Current Status of Occurrence of Major
714 Diseases on Kiwifruits and Their Control in Korea. *Acta Horticulturae* 610:
715 437-443.

716 10. Ferguson AR, Huang H (2007) Genetic resources of kiwifruit:
717 domestication and breeding. *Horticultural Reviews*, ed Janick J. (John Wiley
718 & Sons, Hoboken), pp 1-121.

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

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

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

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

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

16. Cruzat C (2014) The kiwifruit in Chile and in the world. *Revista Brasileira de Fruticultura*. 36(1): 112-123.

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

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

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

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

21. Marcelletti S, Ferrante P, Petriccione M, Firrao G, Scortichini M (2011) *Pseudomonas syringae* pv. *actinidiae* Draft Genomes Comparison Reveal Strain-Specific Features Involved in Adaptation and Virulence to *Actinidia* Species. *PLoS ONE* 6(11):e27297.

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

23. McCann HC, et al. (2013) Genomic Analysis of the Kiwifruit Pathogen *Pseudomonas syringae* pv. *actinidiae* Provides Insight into the Origins of an Emergent Plant Disease. *PLoS Pathog* 9(7):e1003503.

24. Mazzaglia A, et al. (2012) *Pseudomonas syringae* pv. *actinidiae* (PSA) Isolates from Recent Bacterial Canker of Kiwifruit Outbreaks Belong to the Same Genetic Lineage. *PLoS ONE* 7(5):e36518.

25. Templeton MD, Warren BA, Andersen MT, Rikkerink EHA, Fineran PC (2015) Complete DNA Sequence of *Pseudomonas syringae* pv. *actinidiae*, the Causal Agent of Kiwifruit Canker Disease. *Genome Announcements* 3(5):e01054–15.

26. Fujikawa T, Sawada H (2016) Genome analysis of the kiwifruit canker pathogen *Pseudomonas syringae* pv. *actinidiae* biovar 5. *Scientific Reports* 6:21399–11.

27. Biek R, Pybus OG, Lloyd-Smith JO, Didelot X (2015) Measurably evolving pathogens in the genomic era. *Trends in Ecology & Evolution* 30(6):306–

313.

28. Grad YH, Lipsitch M (2014) Epidemiologic data and pathogen genome sequences: a powerful synergy for public health. *Genome Biol* 15(11):538.

29. Wu X, et al. (2014) Deciphering the Components That Coordinately Regulate Virulence Factors of the Soft Rot Pathogen *Dickeya dadantii*. *MPMI* 27(10):1119–1131.

30. Page F, et al. (2001) Osmoregulated Periplasmic Glucan Synthesis Is Required for *Erwinia chrysanthemi* Pathogenicity. *Journal of Bacteriology* 183(10):3134–3141.

31. Klosterman SJ, et al. (2011) Comparative Genomics Yields Insights into Niche Adaptation of Plant Vascular Wilt Pathogens. *PLoS Pathog* 7(7):e1002137–19.

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

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

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

35. Gal M, Preston GM, Massey RC, Spiers AJ, Rainey PB (2003) Genes encoding a cellulosic polymer contribute toward the ecological success of *Pseudomonas fluorescens* SBW25 on plant surfaces. *Mol Ecol* 12(11):3109–3121.

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

37. Japanese Ministry of Agriculture, Forestry and Fisheries, Yokohama Plant Protection Station, Research Division (2016) Pest Risk Analysis Report on *Pseudomonas syringae* pv. *actinidiae*. 1–24.

38. Mather AE, Reid S, Maskell DJ, Parkhill J (2013) Distinguishable epidemics of multidrug-resistant *Salmonella* Typhimurium DT104 in different hosts. *Science* 341(6153):1514-7.

39. Wagner DM, et al. (2014) *Yersinia pestis* and the Plague of Justinian 541–543 AD: a genomic analysis. *The Lancet Infectious Diseases* 14(4):319–326.

40. Andam CP, Worby CJ, Chang Q, Campana MG (2016) Microbial Genomics of Ancient Plagues and Outbreaks. *TRENDS in Microbiology*:1–13.

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

42. Almeida RPP, Nunney L (2015) How Do Plant Diseases Caused by *Xylella fastidiosa* Emerge? *Plant Disease* 99(11):1457–1467.

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

44. Clarke CR, et al. (2015) Genome-Enabled Phylogeographic Investigation of the Quarantine Pathogen *Ralstonia solanacearum* Race 3 Biovar 2 and Screening for Sources of Resistance Against Its Core Effectors. *Phytopathology* 105(5):597–607.

45. Vinatzer BA, Monteil CL, Clarke CR (2014) Harnessing Population Genomics to Understand How Bacterial Pathogens Emerge, Adapt to Crop Hosts, and Disseminate. *Annu Rev Phytopathol* 52(1):19–43.

46. Stukenbrock EH, Bataillon T (2012) A Population Genomics Perspective on the Emergence and Adaptation of New Plant Pathogens in Agro-Ecosystems. *PLoS Pathog* 8(9):e1002893.

47. Shapiro LR, et al. (2016) Horizontal Gene Acquisitions, Mobile Element Proliferation, and Genome Decay in the Host-Restricted Plant Pathogen *Erwinia Tracheiphila*. *Genome Biology and Evolution* 8(3):649–664.

48. Quibod IL, et al. (2016) Effector Diversification Contributes to *Xanthomonas oryzae* pv. *oryzae* Phenotypic Adaptation in a Semi-Isolated Environment. *Scientific Reports* 6:34137.

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

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

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

52. Bankevich A, et al. (2012) SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *Journal of Computational Biology* 19(5):455–477.

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

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

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

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

847 57. Garrison E, Marth G (2012) Haplotype-based variant detection from short-
848 read sequencing. *arXiv*:1207.3907.

849 58. Didelot X, Wilson DJ (2015) ClonalFrameML: Efficient Inference of
850 Recombination in Whole Bacterial Genomes. *PLoS Computational Biology*
851 11(2):e1004041–18.

852 59. Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and
853 post-analysis of large phylogenies. *Bioinformatics*.

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

857 61. Walker BJ, et al. (2014) Pilon: An Integrated Tool for Comprehensive
858 Microbial Variant Detection and Genome Assembly Improvement. *PLoS*
859 *ONE* 9(11):e112963–14.

860 62. Seemann T (2014) Prokka: rapid prokaryotic genome annotation.
861 *Bioinformatics* 30(14):2068–2069.

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

864 64. Krzywinski M, Schein J, Birol I, Connors J (2009) Circos: an information
865 aesthetic for comparative genomics. Genome Research 19:1639-1635.

866