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 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*. 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

48 within-lineage recombination, there has been substantial gene gain and loss even
49 within the single clade from which the global pandemic arose.
50
51

52 SIGNIFICANCE STATEMENT

53

54 Bleeding canker disease of kiwifruit caused by *Pseudomonas syringae* pv.
55 *actinidiae* (*Psa*) has come to prominence in the last three decades. Emergence has
56 coincided with domestication of the host plant and provides a rare opportunity to
57 understand ecological and genetic factors affecting the evolutionary origins of *Psa*.
58 Here, based on genomic analysis of an extensive set of strains sampled from China
59 and augmented by isolates from a global sample, we show, contrary to earlier
60 predictions, that China is not the native home of the pathogen, but is nonetheless the
61 source of the recent global pandemic. Our data identify specific transmission events,
62 substantial genetic diversity and point to non-agricultural plants in either Japan or
63 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

89 in the 2000s and an outbreak of global proportions soon followed. The first published
90 notices of the latest outbreak on gold kiwifruit issued from Italy in 2008, with reports
91 from neighbouring European countries, New Zealand, Asia and Chile occurring soon
92 after (1-6, 20). Whole genome sequencing showed the most recent global outbreak of
93 disease was caused by a new lineage of *Psa* (previously referred to as *Psa*-V and now
94 referred to as *Psa*-3), while earlier disease incidents in Japan and Korea were caused
95 by strains forming separate clades referred to *Psa*-1 (previously *Psa*-J) and *Psa*-2
96 (previously *Psa*-K), respectively (21-24). These clades are marked by substantial
97 variation in their complement of type III secreted effectors, which are required for
98 virulence in *P. syringae*. Despite the surprising level of within-pathovar differences in
99 virulence gene repertoires occurring subsequent to the divergence of these three
100 clades, strains from each clade are capable of infecting and growing to high levels in
101 both *A. chinensis* var. *deliciosa* and *A. chinensis* var. *chinensis* (23).
102 The severity of the latest global outbreak is largely predicated on the
103 expansion in cultivation of clonally propagated highly susceptible *A. chinensis var.*
104 *chinensis* cultivars, with trade in plant material and pollen likely providing
105 opportunities for transmission between distant geographic regions. Identifying the
106 source from which *Psa* emerged to cause separate outbreaks remains an important
107 question. Intriguingly, despite the divergence in both the core and flexible genome,
108 these distinct clades nevertheless exhibit evidence of recombination with each other
109 and unknown donors (23). This suggests each lineage of *Psa* emerged from a
110 recombining source population. Definitive evidence for the location, extent of
111 diversity and evolutionary processes operating within this population remain elusive.
112 Early reports suggested China may be the source of the latest global outbreak (22,
113 23). Although the strains of *Psa* available at that time did not provide unambiguous

and well-supported evidence of a Chinese origin, this speculation was based on the fact that kiwifruit are native to China; it is the provenance of the plant material selected for commercial and breeding purposes in China, New Zealand, Italy and other kiwifruit growing regions; there is extensive trade in plant material between all of these regions; and one Chinese isolate was found to carry an integrative and conjugative element (ICE) that was also found in New Zealand *Psa*-3 isolates (22).

In order to strengthen understanding of the population structure, phylogeography and evolutionary dynamics of *Psa*, we isolated *Psa* from cultivated kiwifruit across six provinces in China and obtained additional isolates from South Korea and Japan. Genome sequencing of fifty isolates and the inclusion of an additional thirty previously sequenced isolates show that while China is the origin of the pandemic lineage of *Psa,* only a single clade is currently present in China, while strains from multiple clades are present in both Korea and Japan. Strains from the pandemic lineage are closely related and display reduced pairwise nucleotide diversity relative to other lineages, indicating a more recent origin. Distinct transmission events were responsible for the 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.

RESULTS

The phylogeography of *Psa*

The genomes of 50 *P. syringae* pv. *actinidiae* (*Psa*) isolated from symptomatic kiwifruit in China, Korea and New Zealand between 2010 and 2015 were sequenced (Table S1). Combined with 30 *Psa* genomes from earlier outbreaks and different geographic regions (e.g. Italy and Chile), our samples represent the main *Psa* genotypes from the countries producing 90% of kiwifruit production worldwide. The completed reference genome of *Psa* NZ13 (ICMP 18884) comprises a 6,580,291bp chromosome and 74,423bp plasmid (23, 25). Read mapping and variant calling with reference to *Psa* NZ13 chromosome produced a 1,059,722bp non-recombinant core genome for all 80 genomes, including 2,953 nonrecombinant SNPs. A maximum likelihood phylogenetic analysis showed the four clades of *Psa* known to cause bleeding canker disease in kiwifruit were represented among the 80 strains (Figure 1). The first clade (*Psa*-1) includes the pathotype strain of *Psa* isolated and described during the first recorded epidemic of bleeding canker disease in Japan (1984-1988). The second clade (*Psa*-2) includes isolates from an epidemic in South Korea (1997-1998), and the third clade (*Psa*-3) includes isolates that define the global pandemic lineage (2008-present). A fourth clade (*Psa*-5) is represented by a single strain, as no additional sequences or isolates were available (26). The average between and within-clade pairwise identity is 98.93% and 99.73%, respectively (Table S2). All *Psa* isolated from kiwifruit across six different provinces in China group are members of the same clade: *Psa*-3. A subset of Chinese strains group with the *Psa* isolated during the global outbreak in Italy, Portugal, New Zealand, and Chile. This subset is referred to as the pandemic lineage of *Psa*-3.

In order to obtain greater resolution of the relationships between the new Chinese and pandemic isolates, we identified the 4,853,413bp core genome of all 62 strains in *Psa*-3. The core genome includes both variant and invariant sites and excludes regions either unique to or deleted from one or more strains. To minimize the possibility of recombination affecting the reconstruction of evolutionary relationships and genetic distance within *Psa*-3, ClonalFrameML was employed to identify and remove 258 SNPs with a high probability of being introduced by recombination rather than mutation, retaining 1,948 nonrecombinant SNPs. The within-lineage ratio of 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 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*-3, it is less frequent and has introduced fewer polymorphisms relative to mutation: when accounting for polymorphisms present in recombinant regions identified by ClonalFrameML and/or present on transposons, plasmids, and other mobile elements, more than seven-fold more polymorphisms were introduced by mutation relative to recombination (Table 1). Recombination has a more pronounced impact between clades, where substitutions are slightly more likely to have been introduced by recombination than by mutation (Table 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

210 therefore represents a reservoir from which novel strains are likely to emerge in the
211 future.

212 The reduced level of diversity within the core genome of pandemic *Psa*-3
213 demonstrates these strains have been circulating for a shorter period of time relative to
214 those responsible for earlier outbreaks in both Japan and Korea. In order to estimate
215 the divergence time of the pandemic lineages as well as the age of the most recent
216 common ancestor of all *Psa* clades displaying vascular pathogenicity on kiwifruit, we
217 performed linear regression of root-to-tip distances against sampling dates using the
218 RAxML phylogenies determined from the non-recombinant core genome of all clades
219 and of *Psa*-3 alone. No temporal signal was identified in the data. There were poor
220 correlations between substitution accumulation and sampling dates, indicating the
221 sampling period may have been too short for sufficient substitutions to occur. There
222 may also be variation in the substitution rate within even a single lineage. Forty-four
223 unique non-recombinant SNPs were identified among the 21 pandemic *Psa*-3
224 genomes sampled over five years in New Zealand (an average of 2.10 per genome)
225 over five years) producing an estimated rate of 8.7×10^{-8} substitutions per site per
226 year. The relatively slow substitution rate and the strong bottleneck effect experienced
227 during infections hinders efforts to reconstruct patterns of transmission, as the global
228 dissemination of a pandemic strain may occur extremely rapidly (27, 28). The
229 estimated divergence time of *Psa* broadly considered is likely older than the pandemic
230 and epidemic events with which they are associated: the earliest report of disease
231 cause by lineage 1 occurred in 1984 and the first report of infection from the latest
232 pandemic was issued in 2008.
233

Diversification and parallelism among *Psa*-3 isolates

2,206 SNPs mapping to the core genome of *Psa*-3 were identified; 258 of these mapped to recombinant regions identified by ClonalFrameML and/or plasmid, prophage, integrative and conjugative elements, transposons and other mobile genetic elements (Table 1, Figure 4). The highest density of polymorphism occurs in and around the integrative and conjugative element (ICE) in *Psa* NZ13 (Figure 4). Of the 1,948 SNPs mapping to the non-recombinant non-mobile core genome, 58.1% (1,132) are strain specific. Most strain-specific SNPs are found in the two most divergent members of the lineage: *Psa* C16 and C17, with 736 and 157 strain-specific SNPs, respectively. The remaining isolates have an average of 4.0 strain-specific SNPs, ranging from 0 to 44 SNPs per strains. There are 816 SNPs shared between two or more *Psa*-3 strains. The pandemic clade differs from the more divergent Chinese strains by 72 shared SNPs. Within the pandemic lineage there are 125 strain-specific SNPs, an average of 3.1 unique SNPs per strain (ranging from 0-27 SNPs) and an additional 29 SNPs shared among pandemic strains. Protein-coding sequence accounts for 88.4% of the non-recombinant, gap-free core genome of this clade. We observed that 78.9% (1,536/1,948) of mutations occurred in protein coding sequence, significantly different from the expectation (1,722/1,948) in the absence of selection (Pearson's χ^2 test: $P < 0.0001$, $\chi^2 = 173.46$). This suggests there is selection against mutations occurring in protein coding sequences. Of the 953 substitutions introduced by mutation in *Psa*-3, 927 resulted in amino acid substitutions, two resulted in extensions and 24 resulted in premature truncations.

Multiple synonymous and non-synonymous mutations were identified in 269 genes. The accumulation of multiple independent mutations in the same gene may be a function of gene length, mutational hotspots or directional selection. A range of

hypothetical proteins, membrane proteins, transporters, porins and type III and IV secretion system proteins acquired between two and seven mutations. The fitness impact of these mutations – and the 38 amino-acid changing mutations in the ancestor of the pandemic subclade – is unknown, yet it is possible these patterns are the outcome of selective pressures imposed during bacterial residence within a similar host niche.

Two substitutions are shared exclusively by the European pandemic strains (AKT28710.1 G1150A and AKT33438.1 T651C) and one silent substitution in a gene encoding an acyltransferase superfamily protein (AKT31915.1 C273T) is shared among the European pandemic and six of nine Chinese pandemic strains (C3, C29-31, C67, C69). As these six Chinese pandemic 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 S3). 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.

284 *deliciosa* 'Hayward', attaining 1.2 x 10^7 cfu/g three days post inoculation (dpi) and
285 declining to 8.8 x 10^4 cfu/g at fourteen dpi (Figure S1). This is a marked reduction
286 compared to *Psa* NZ13, which attains 3.0 x 10^9 and 4.2 x 10^7 cfu/g three and fourteen
287 dpi, respectively. *Psa* C17 nevertheless multiplies between day 0 and day 3,
288 indicating that even in the absence of type III-mediated host defense disruption, *Psa*
289 may still proliferate in host tissues. The loss of the TTSS does not inhibit the growth
290 of *Psa* C17 as strongly in the more susceptible *A. chinensis* var. *chinensis* 'Hort16A'
291 cultivar.

292 Two potentially significant deletion events occurred in the ancestor of the
293 pandemic subclade: a frameshift caused by a mutation and single base pair deletion in
294 a glucan succinyltransferase (*opgC*) and a 6,456bp deletion in the *wss* operon (Figure
295 S2). Osmoregulated periplasmic glucans (OPGs, in particular *opgG* and *opgH*) are
296 required for motility, biofilm formation and virulence in various plant pathogenic
297 bacteria and fungi (29-31). Homologs of *opgGH* remain intact in the pandemic
298 subcclade, yet the premature stop mutation in *opgC* likely results in the loss of glucan
299 succinylation. The soft-rot pathogen *Dickeya dadantii* expresses OpgC in high
300 osmolarity conditions, resulting in the substitution of OPGs by O-succinyl residues
301 (32). *D. dadantii opgC* deletion mutants did not display any reduction in virulence
302 (32). *Psa* is likely to encounter high osmolarity during growth and transport in xylem
303 conductive tissues, yet the impact of the loss of *opgC* on *Psa* fitness has yet to be
304 determined. The most striking difference between the pandemic subclade and more
305 divergent Chinese *Psa*-3 strains is the deletion of multiple genes involved in cellulose
306 production and acetylation of the polymer (Figure S2) (33). The loss of cellulose
307 production and biofilm production is not associated with a reduction in growth or
308 symptom development of *P. syringae* pv. *tomato* DC3000 on tomato, but may

309 enhance bacterial spread through xylem tissues during vascular infections (34). In *P.*
310 *fluorescens* SBW25 deletion of the Wss operon significantly compromises ability to
311 colonise plant surfaces and in particular the phyllosphere of sugar beet (*Beta vulgaris*)
312 seedlings (35). It is possible that loss of this locus aids movement through the
313 vascular system and / or dissemination between plants, by limiting capacity for
314 surface colonization and biofilm formation.
315
316 **Dynamic genome evolution of *Psa*-3**
317 Despite the high similarity within the core genome, extensive variation is
318 evident in the pangenome of *Psa*-3 (Figure 5). The core genome (4,339 genes in 99-
319 100% of strains, and 674 genes in 95%-99% of strains, or 58-62 genomes) comprises
320 50.5% of the total pangenome (9,931 genes). 968 genes are present in 15-95% of
321 strains (9-57 genomes), the so-called 'shell genes' (Figure 5). The flexible genome is
322 comprised of the 'shell' and 'cloud' genes; the latter describes genes present in 0-15%
323 of strains (one to six genomes in this case). Cloud genes contribute most to the
324 flexible genome: 3,950 genes are present in one to six strains. This is a striking
325 amount of variation in a pathogen described as clonal and monomorphic. It should be
326 noted that sequencing and assembly quality will impact annotation and pangenome
327 estimates: omitting the low quality J39 assembly results in a core and soft-core
328 genome differing by 18 genes and a reduction of the cloud by 275. Despite a
329 relatively slow rate of mutation and limited within-clade homologous recombination,
330 the amount of heterologous recombination demonstrates that the genomes of these
331 pathogens are highly labile. Mobile genetic elements like bacteriophage, transposons
332 and integrases make a dramatic contribution to the flexible genome. Integrative and
333 conjugative elements (ICEs) are highly mobile elements and have recently been

334 demonstrated to be involved in the transfer of copper resistance in *Psa* (36).
335 Prodigious capacity for lateral gene transfer creates extreme discordance between ICE
336 type, host phylogeny and host geography making these regions unsuitable markers of
337 host evolution and origin.

338 Three divergent ICEs have been previously described from the global
339 pandemic lineage (23). Within *Psa*-3 ICEs were found in 53 of 62 isolates (nine of the
340 divergent Chinese isolates were devoid of any such element) (Figure 2). No
341 phylogeographic signal is evident. For example, strains from Sichuan, Shaanxi,
342 Korea, Italy and Portugal share an identical ICE. Even within a single Chinese
343 province, mulitple ICEs exist (Shaanxi and Sichuan isolates harbour four and three
344 different ICEs, respectively). Moreover, ICE host range is not limited to *Psa* alone:
345 the ICE found in every NZ isolate (and also recorded in Chinese isolate C1) exists in
346 essentially identical form in a strain of *P. syringae* pv. *avellenae* CRAPAV013
347 isolated from hazelnut in 1991 in Latina, Italy (it exhibits 98% pairwise identity,
348 differing from the New Zealand ICE by a transposon, 66bp deletion and a mere 6
349 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 *Psa*-3 detection in both those countries, though the amount of pollen imported from

New Zealand in 2012 (349 kg) and 2013 (190 kg) far outweighed the amount imported from China (1 kg in both 2012 and 2013) (37).

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

Strains from three different clades have been isolated in both Korea and Japan, while China harbours strains from only a single clade. The most basal clades of canker-causing *Psa* are comprised of Korean strains isolated between 1997 and 2014 (*Psa*-2) and a member of the recently identified lineage *Psa*-5. One early isolate (*Psa* K3, 1997) groups with the Japanese isolates in lineage 1, and a more recent Korean isolate *Psa* K7 (2014) groups with the more diverse Chinese isolates in *Psa*-3. Korea therefore harbours a more diverse population of *Psa* than China, with strains from three distinct lineages of *Psa* (1, 2 and non-pandemic subclade 3). The novel group of *Psa* recently identified in Japan (*Psa*-5) appears to share an ancestor with the Korean *Psa*-2 strains. With the recent dissemination of pandemic *Psa*-3 and the historical presence of *Psa*-1 in Japan, the Japanese population of *Psa* is comprised of three distinct clades of *Psa* (1, 5 and pandemic *Psa*-3). Though no strains from *Psa*-1 have been isolated in either Japan or Korea since 1997, at least two lineages currently coexist in both Japan and Korea. This strongly suggests that the source population of all *Psa* is not China, but likely resides in either Korea or Japan. The potential

399 transmission of a non-pandemic *Psa*-3 strain from China to Korea and the
400 identification of a new clade in Japan supports our earlier assertion that variants will
401 continue to emerge to cause local epidemics and global pandemics in the future (23).
402 Considering that the divergence time of this monophyletic pathovar predates
403 the commercialisation of kiwifruit by hundreds if not thousands of years, *Psa* is likely
404 associated with a non-domesticated host(s) in the wild. Both *A. chinensis var.*
405 *deliciosa* or *A. chinensis var. chinensis* are found in natural ecosystems and have
406 overlapping habitat ranges with cultivated kiwifruit in many areas. However, despite
407 isolating 746 *Pseudomonas* strains from both wild and cultivated kiwifruit during this
408 sampling program, we did not identify *Psa* among any of the 188 *Pseudomonas* spp.
409 isolated from 98 wild *A. chinensis var. deliciosa* or *A. chinensis var. chinensis*
410 sampled across China (Table S4). Very few *Actinidia* spp. have ranges extending to
411 South Korea and Japan: *A. arguta*, *A. kolomikta*, *A. polygama* and *A. rufa*. *A. arguta*
412 are broadly distributed across both Korea and Japan. Early work by Ushiyama *et al.*
413 (1992) found that *Psa* could be isolated from symptomatic *A. arguta* plants in Japan.
414 The possibility that this wild relative of kiwifruit harbours diverse strains of *Psa* that
415 may emerge to cause future outbreaks is currently under investigation. Alternately, a
416 host shift from another domesticated crop may have occurred after expansion in
417 kiwifruit cultivation.
418 Numerous epidemiological studies of human pathogens have demonstrated
419 environmental or zoonotic origins, but there are few such studies of plant pathogens
420 (23, 38-48). Where ecological and genetic factors restrict pathogens to a small
421 number of plant hosts some progress has been made, but for facultative pathogens
422 such as *P. syringae* that colonise multiple hosts and are widely distributed among

423 both plant and non-plant habitats, the environmental reservoirs of disease and factors
424 affecting their evolutionary emergence are difficult to unravel (49, 50).

425 The emergence of *Psa* over the last three decades – concomitant with
426 domestication of kiwifruit– offers a rare opportunity to understand the relationship
427 between wild populations of both plants and microbes and the ecological and
428 evolutionary factors driving the origins of disease, including the role of agriculture. It
429 is now possible to exclude China as the native home to the source population, but the
430 precise location remains unclear. Nonetheless, it is likely, given the extent of diversity
431 among *Psa* isolates and the time-line to domestication, that ancestral populations exist
432 in non-agricultural plant communities. Attention now turns to Korea and Japan and in
433 particular the interplay between genetic and ecological factors that have shaped *Psa*
434 evolution.
435

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 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

461 by Templeton *et al*. (2015), where it is referred to as ICMP18884 (23, 25). Variant
462 calling was performed on all *P*. *syringae* pv. *actinidiae* isolates for which read data
463 was available.

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

472 Reads were mapped to the complete reference genome *Psa* NZ13 with
473 Bowtie2 and duplicates removed with SAM Tools (55, 56). Freebayes was used to
474 call variants with a minimum base quality 20 and minimum mapping quality 30 (57).
475 Variants were retained if they had a minimum alternate allele count of 10 reads and
476 fraction of 95% of reads supporting the alternate call. The average coverage was
477 calculated with SAM Tools and used as a guide to exclude overrepresented SNPs
478 (defined here as threefold higher coverage than the average) which may be caused by
479 mapping to repetitive regions. BCFtools filtering and masking was used to generate
480 final reference alignments including SNPs falling within the quality and coverage
481 thresholds described above and excluding SNPs within 3bp of an insertion or deletion
482 (indel) event or indels separated by 2 or fewer base pairs. Invariant sites with a
483 minimum coverage of 10 reads were also retained in the alignment, areas of low (less
484 than 10 reads) or no coverage are represented as gaps relative to the reference.

Freebayes variant calling includes indels and multiple nucleotide insertions as well as single nucleotide insertions, however only SNPs were retained for downstream phylogenetic analyses. An implementation of ClonalFrame suitable for use with whole genomes was employed to identify recombinant regions using a maximum likelihood starting tree generated by RaxML (58, 59). All substitutions occurring within regions identified as being introduced due to recombination by ClonalFrameML were removed from the alignments. The reference alignments were manually curated to exclude substitutions in positions mapping to mobile elements such as plasmids, integrative and conjugative elements and transposons.

Phylogenetic Analysis

The maximum likelihood phylogenetic tree of 80 *Psa* strains comprising new Chinese isolates and strains reflecting the diversity of all known lineages was built with RAxML (version 7.2.8) using a 1,062,844bp core genome alignment excluding all positions for which one or more genomes lacked coverage of 10 reads or higher (59). Removal of 3,122 recombinant positions produced a 1,059,722bp core genome alignment including 2,953 variant sites. Membership within each phylogenetic clade corresponds to a minimum average nucleotide identity of 99.70%. The average nucleotide identity was determined using a BLAST-based approach in JspeciesWS (ANIb), using a subset of 32 *Psa* genome assemblies spanning all clades (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 4,853,155bp core genome alignment (excluding 258 recombinant SNPs), comprising invariant sites and 1,948 non-recombinant SNPs and invariant sites. Trees were built with the generalized time-reversible model and gamma

510 distribution of site-specific rate variation (GTR+Γ) and 100 bootstrap replicates. *Psa*
511 C16 was used to root the tree as this was shown to be the most divergent member of
512 the phylogeny when including strains from multiple lineages. Nodes shown have
513 minimum bootstrap support values of 50.
514
515 **Identification of the core and mobile genome**
516 Genomes were assembled with SPAdes using the filtered, trimmed and
517 corrected reads (52). Assembly quality was improved with Pilon and annotated with
518 Prokka (61, 62). The pangenome of *Psa*-3 was calculated using the ROARY pipeline
519 (63). Orthologs present in 61 (out of a total of 62) genomes were considered core;
520 presence in 58-60, 9-57 and 1-8 were considered soft-core, shell and cloud genomes,
521 respectively. BLAST-based confirmation was used to confirm the identity predicted
522 virulence or pandemic-clade-restricted genes in genome assemblies.
523
524 **Pathogenicity assays**
525 Growth assays were performed using both stab inoculation as in McCann *et al*.
526 (2013) an initial inoculum of 10^8cfu/mL and four replicate plants at day 0 and six at
527 all subsequent sampling time points. Bacterial density in inoculated tissue was
528 assessed by serial dilution plating of homogenized tissue. Statistical significance
529 between each treatment at each time point was assessed using two-tailed t-tests with
530 uneven variance.
531

ACKNOWLEDGEMENTS

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

TABLES

Table 1. Origin of SNPs in core genomes

		Intergenic	Synon	Nonsyn	Extension	Trunc	Total	
***Psa*-3**	Mutation	412	583	927	2	24	1,948	
	Recombination	35	136	85	1	1	258	7.55
All *Psa*	Mutation	457	1,493	982	3	18	2,953	
	Recombination	355	2,212	558	2	5	3,132[1]	0.94

[1]For ten positions in the alignment, two recombination events were predicted to occur

554 **Table S1. Strains**
555

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.	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

558

559 **Table S2. Average percent identity within and between Psa lineages**
560

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]

561
562 [1]Not determined for *Psa*-5 as only a single strain has been sequenced from this lineage. ANIb values
563 determined using representative strains for *Psa*-1 (J31, K3, J2, J1, J30, J29, J35, J32, J33, I1, J36, J25),
564 *Psa*-2 (K27, K6, K4, K26, K28), *Psa*-3 (C16, C17, C10, C11, C70, K7, C15, C54, C74, C69, K5, C62,
565 I13, NZ13) and *Psa*-5 (J37)
566
567

568 **Table S3. SNPs shared between all pandemic NZ and Japanese isolates**
569

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)

570 **Table S4. Isolates identified by cultivation and disease status of host**
571

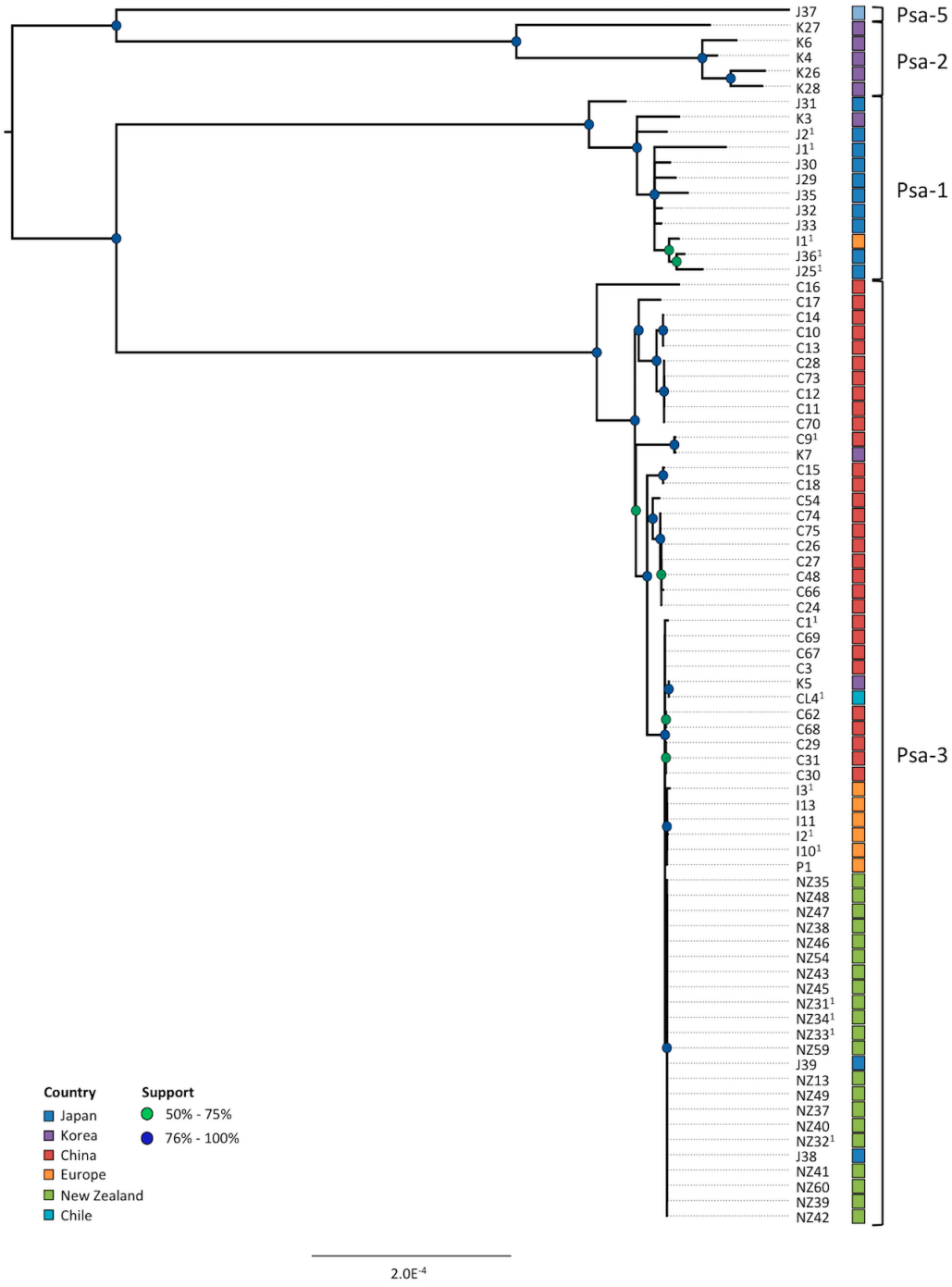
	A. arguta	*A. callosa*	*A. chinensis* var. *chinensis*	*A. chinensis* var. *deliciosa*	*Actinidia* sp.	*Camellia* sp.	*Prunus* sp.	Total
Cultivated	3		373	350	1	45	45	817
Disease	3		323	244		30	45	645
Not determined	2		121	81		20	10	234
Pseudomonas spp.	1		123	120		8	35	287
P. syringae			12	23		2		37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

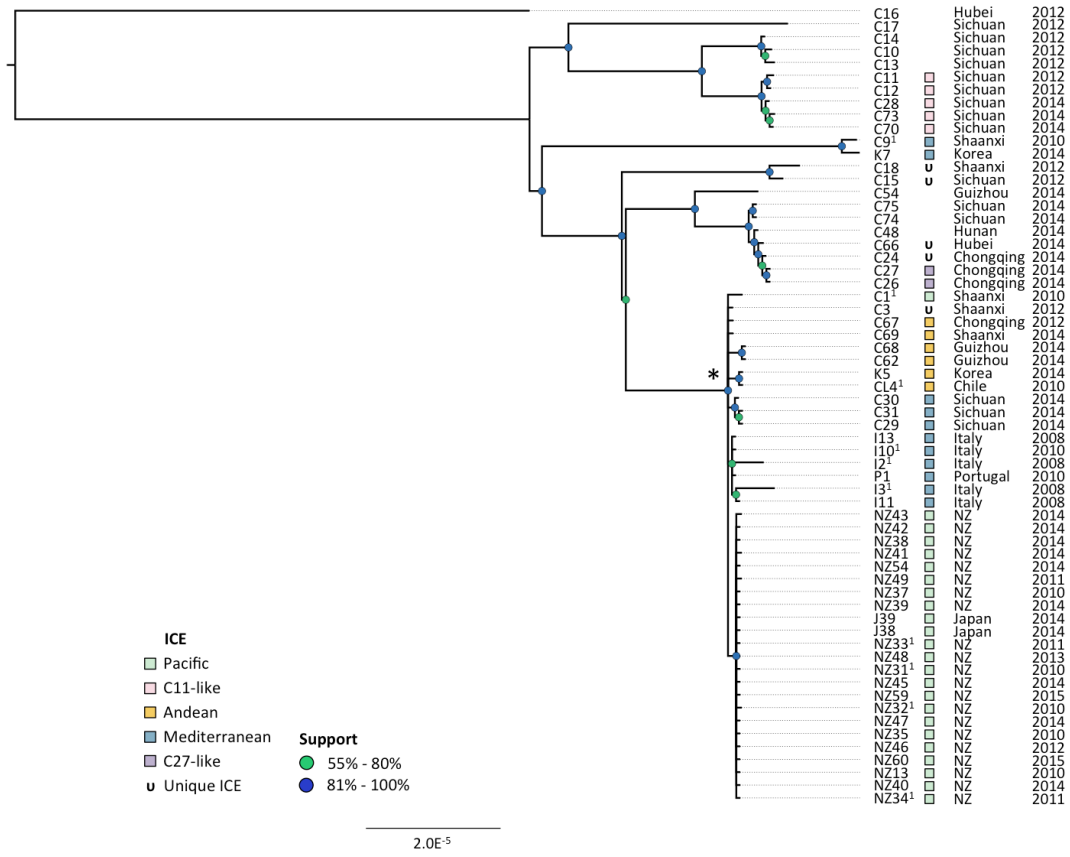
572
573
574

FIGURES

Figure 1. Phylogeny of *Psa*



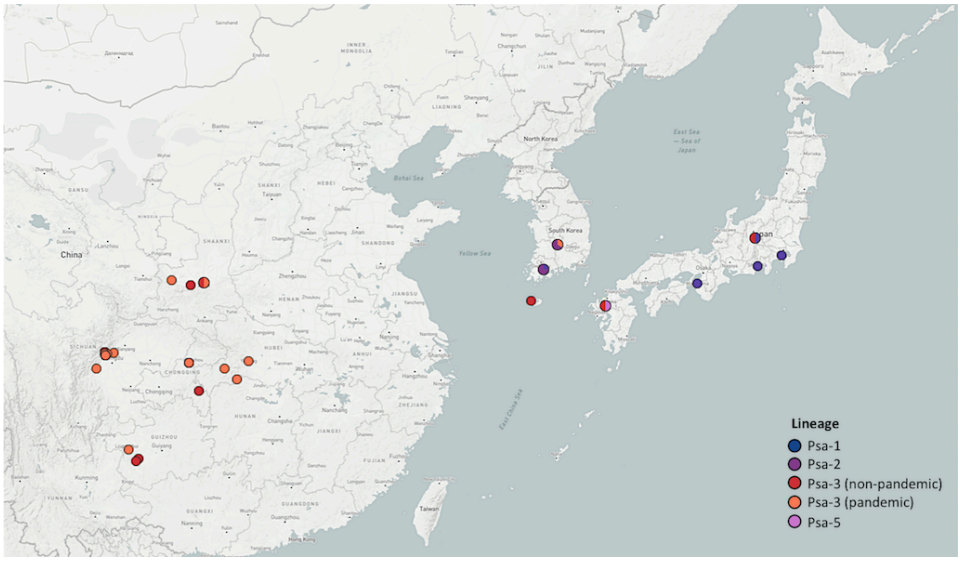
579 **Figure 2. Phylogeny of *Psa*-3**



580

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

582



583

584

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

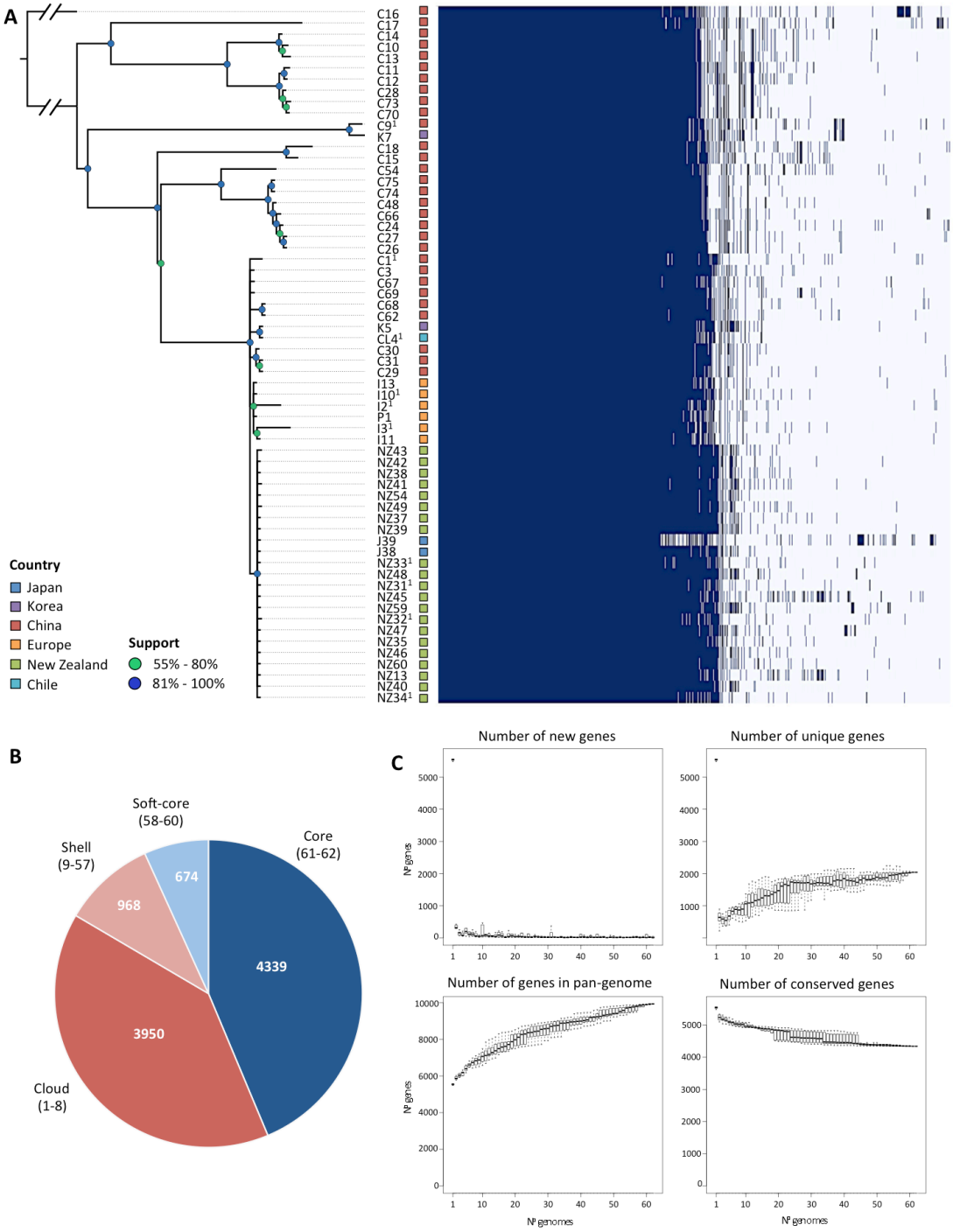
586

587

588

589 **Figure 5. Pangenome of *Psa*-3**

590



591

592

593

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

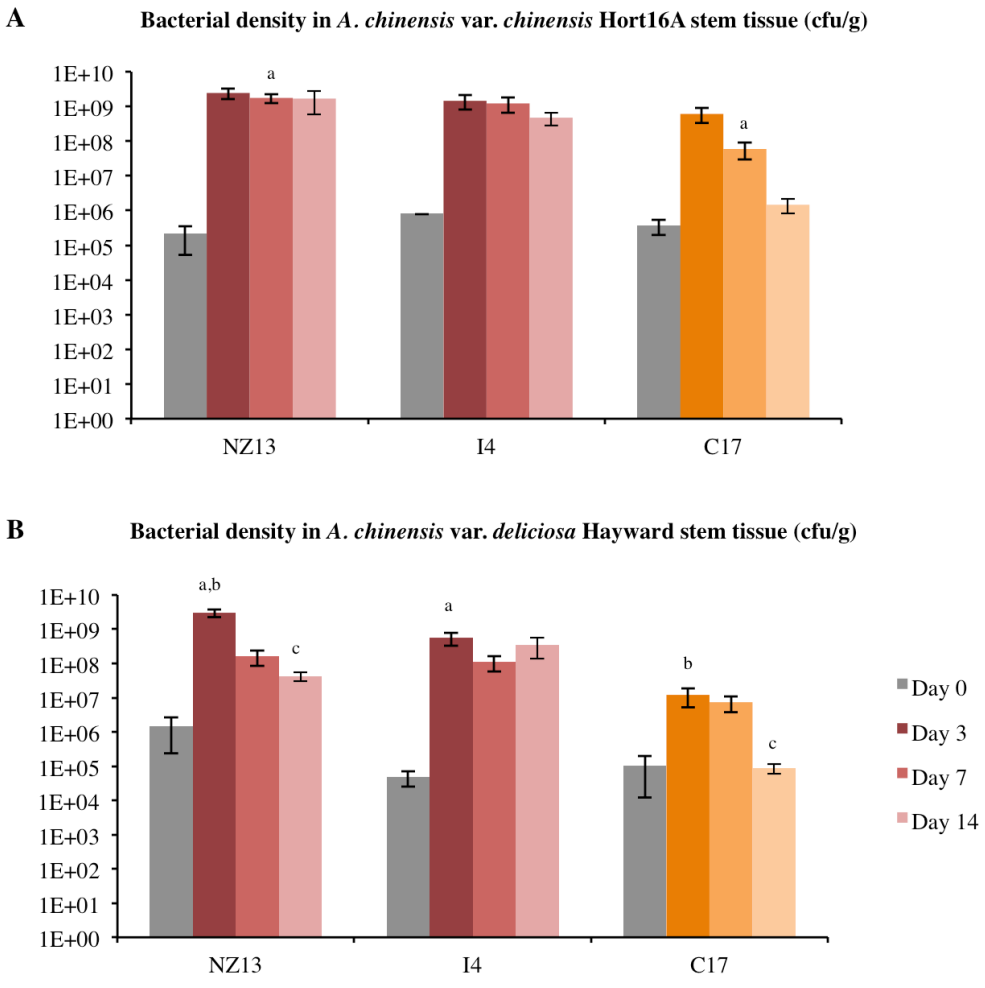


Figure S2. *Wss* operon disruption in *Psa*-3

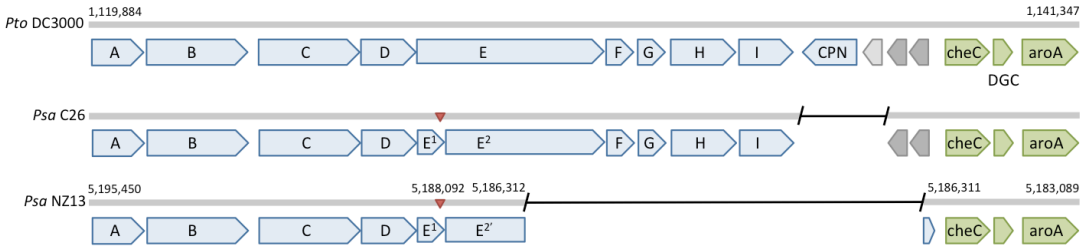


FIGURE LEGENDS

Figure 1. Phylogeny of *Psa*

RaxML Maximum likelihood tree based on 1,059,722bp non-recombinant core genome alignment including 2,953 variant sites. All nodes displayed have bootstrap support values above 50% (50-75% in green, 76-100% in blue). Province of isolation is displayed for Chinese isolates.

Figure 2. Phylogeny of *Psa*-3

Maximum likelihood tree based on 4,853,155bp non-recombinant core genome alignment including 1,948 variant sites. All nodes displayed have bootstrap support values above 55% (55-80% in green, 81-100% in blue). Year and province (China) or country of isolation is displayed. Integrative and conjugative elements (ICEs) present in each host genome are indicated.

Figure 3. Psa isolation locations in East Asia

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

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

Polymorphisms and recombinant regions mapped onto *Psa* NZ13 reference genome using CIRCOS (64). *Psa* NZ13 CDS are displayed in the first and second ring (blue), with annotated Type 3 secretion system and effectors highlighted (red). Inner rings

display polymorphisms in *Psa*-3 genomes ordered from most to least divergent relative to *Psa* NZ13 (see Figure 2). The most polymorphic region corresponds to the location of the integrative and conjugative element (ICE) in *Psa* NZ13.

Figure 5. Pangenome of *Psa*-3

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

Figure S1. Bacterial growth assay on *A. deliciosa*

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

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

Genes encoding components of the *wss* operon (blue), hypothetical and conserved hypothetical (light and dark grey), chemotaxis, diguanylate cyclase and *aroA* (green). Deletions (black line) and position of single base pair insertion (red triangle)

displayed with reference to *Pto* DC3000. Insertion results in frameshift mutation in *wssE*, two predicted derivatives annotated as *wssE1* and *wssE2*. The subsequent 6.5kb deletion in the ancestor of the pandemic subclade results in the truncation of *wssE2*, annotated as *wssE2'*.

REFERENCES

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

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

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

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

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

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

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

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

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

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

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

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

13. Shim KK, Ha YM (1999) Kiwifruit production and research in Korea. *Acta 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.

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

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

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

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

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

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

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

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

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

793 52. Bankevich A, *et al*. (2012) SPAdes: A New Genome Assembly Algorithm and
794 Its Applications to Single-Cell Sequencing. *Journal of Computational Biology*
795 19(5):455–477.

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

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

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

803 56. Li H (2011) A statistical framework for SNP calling, mutation discovery,
804 association mapping and population genetical parameter estimation from

sequencing data. *Bioinformatics* 27(21):2987–2993.

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

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

59. Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30(9):1312-1313.

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

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

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

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

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