Molecular Plant Pathology

MOLECULAR PLANT PATHOLOGY (2014) 15(5), 461–465 DOI: 10.1111/mpp.12103

Incongruence between multi-locus sequence analysis (MLSA) and whole-genome-based phylogenies: *Pseudomonas syringae* pathovar *pisi* as a cautionary tale

DAVID A. BALTRUS[1,*], KEVIN DOUGHERTY[1], STEPHEN M. BECKSTROM-STERNBERG[2], JAMES S. BECKSTROM-STERNBERG[2] AND JEFFREY T. FOSTER[2]

[1]*School of Plant Sciences, University of Arizona, Tucson, AZ 85721-0036, USA*
[2]*Center for Microbial Genetics and Genomics, Northern Arizona University, Flagstaff, AZ 86011-4073, USA*

SUMMARY

Previous phylogenies, built using a subset of genomic loci, split *Pseudomonas syringae* pv. *pisi* into two well-supported clades and implied convergence in host range for these lineages. The analysis of phenotypic and genotypic data within the context of this phylogenetic relationship implied further convergence at the level of virulence gene loss and acquisition. We generate draft genome assemblies for two additional *P. syringae* strains, isolated from diseased pea plants, and demonstrate incongruence between phylogenies created from a subset of the data compared with the whole genomes. Our whole-genome analysis demonstrates that strains classified as pv. *pisi* actually form a coherent monophyletic clade, so that apparent convergence is actually the product of shared ancestry. We use this example to urge caution when making evolutionary inferences across closely related strains of *P. syringae*.

INTRODUCTION

Pseudomonas syringae pathovars *syringae* (*Psy*) and *pisi* (*Ppi*) can both cause bacterial blight in pea (*Pisum sativum*) (Mazarei and Kerr, 1990). In a previous report, Hwang *et al*. (2005) used multi-locus sequence analysis (MLSA) to identify two distinct clusters of *Ppi* belonging to group II as defined by Sarkar and Guttman, (2004). Phenotypic and genotypic comparisons suggest that most strains classified as MLSA group II possess reduced type III effector suites, but possess the genetic capability to produce at least two separate phytotoxins (syringomycin and syringolin). In contrast, both *Ppi* clusters have lost the capability to produce toxins that are otherwise conserved throughout group II strains, and have also gained type III effectors that probably enable pea pathogenesis (Baltrus *et al*., 2011; Hwang *et al*., 2005; Sarkar *et al*., 2006). Although host range convergence has been demonstrated previously to occur in *P. syringae* (O'Brien *et al*., 2012), the evolutionary scenario in which multiple *Ppi* clades have converged on similar virulence genotypes could provide an important data point to bolster the hypothesis of a trade-off between toxin and effector functions for these MLSA group II strains.

Phylogenetic analysis based on draft genome assemblies can provide a more thorough reconstruction of evolutionary history within *Ppi*. Previously, we have described a draft genome assembly for *Ppi* 1704B (Baltrus *et al*., 2011). Here, we describe the sequencing and draft assemblies for two additional strains of *P. syringae* isolated from diseased pea plants. One of these strains (*Ppi* PP1) is designated as pv. *pisi*, but clusters separately from *Ppi* 1704B on the basis of MLSA data (Hwang *et al*., 2005). The other isolate (*Psy* 1212) is classified as pv. *syringae*, which is a pathovar with a potentially broad host range (Arnold *et al*., 2001). In this report, we compare type III effector and toxin suites across these pathogens, and use phylogenies built from whole-genome information to demonstrate incongruence between whole-genome analyses and previous MLSA-based phylogenies for *Ppi*. These phylogenetic differences alter significantly the interpretations of phytopathogenic evolution within *Ppi*, and highlight that caution should be undertaken when analysing trends on the basis of a limited sample of loci, such as within MLSA.

RESULTS

A total of 14 695 838 reads were generated for *Ppi* PP1, and assembly resulted in 256 contigs that total to 5 949 520 bp with an N50 of 81 886 bp. A total of 14 741 668 reads were generated for *Psy* 1212, and assembly resulted in 338 contigs that total to 6 163 906 bp with an N50 of 61 944 bp.

Psy 1212 and *Ppi* PP1 both contain relatively large numbers of potential type III effectors compared with numerous other strains from *P. syringae* MLSA group II. Of note, type III effector suites from both *Psy* B728a and *Psy* 1212 show substantial overlap with each other, and comparisons within pv. *pisi* also show substantial overlap (Table 1). As shown previously (Hwang *et al*., 2005), *Psy* 1212 contains a pathway to produce syringomycin, in addition to two other known toxins within the group II strains (syringolin,

**Correspondence*: Email: baltrus@email.arizona.edu

Table 1 Type III effector family distribution.

	Psy B728a	*Psy* 1212	*Ppi* 1704B	*Ppi* PP1
avrB3	P	P	P	P
avrE	P	P	P	P
avrpto1	P	P		
avrRpm1	P	P		
avrRps4			P	P
hopA3			P	P
hopC1		P	P	P
hopE1			P, partial or pseudo	P
hopF1				P
hopH1	P	P	P	P
hopI1	P	P	P	P
hopM1	P	P	P, partial or pseudo	P
hopR1			P	P
hopX1	P	P, IS element	P	P
hopZ2		P	P	
hopZ3	P	P		
hopAA1	P	P	P	P
hopAB1	P	P		
hopAE1	P	P	P, partial or pseudo	P
hopAF1	P	P	P	P
hopAG1	P	P	P	P
hopAH1	P	P	P, truncated	P, truncated
hopAH2	P	P	P	P
hopAI1	P, truncated	P, start codon		
hopAL1				P
hopAM1				P
hopAT1			P	
hopAX1			P	P
hopAZ1		P		

The letter P denotes the presence of this particular effector family, whereas the lack of a letter denotes absence. Descriptions after the letter P indicate the state of the effector family. Partial or pseudo indicates that there was only a partial match to an effector family, so that complete presence could not be verified. *Ppi*, *Pseudomonas syringae* pv. *pisi*; *Psy*, *Pseudomonas syringae* pv. *syringae*.

syringopeptin), whereas *Ppi* PP1 lacks these pathways altogether. Interestingly, *Psy* 1212 appears to contain a greater number of type III effectors than other sequenced group II isolates. Indeed, this strain contains all of the type III effectors present within *Psy* B728a, as well as three additional effectors (*hopAZ1*, *hopZ2* and a divergent copy of *hopC1*). Substantial overlap also exists between the effector suites of two sequenced isolates from *Ppi*. In particular, *Ppi* PP1 possesses loci from type III effector families that related strains (based on MLSA phylogenies) lack, but which this strain shares with *Ppi* 1704B. These include *hopR1*, *hopE1*, *avrRps4*, *hopA3*, *hopX1*, *hopAX1*, *hopAF1*, *hopC1*, *hopAG1* and *avrB3*.

We have been able to recapitulate previous MLSA branching patterns (Hwang *et al*., 2005) shown for the group II strains from both *Psy* 1212 and *Ppi* PP1, which demonstrate that *Psy* 1212 clusters with other pathovar *syringae* isolates, whereas *Ppi* PP1 appears to cluster separately from *Ppi* 1704B (Fig. 1). Phylogenies of MLSA sequencing built using maximum likelihood and parsimony also consistently split pv. *pisi* into two clades (data not shown). Furthermore, there was no instance of *Ppi* PP1 and *Ppi* 1704B clustering together for any other highly supported tree for all phylogenetic reconstruction methods used. In whole-genome analyses, comparison of the 12 genomes yielded 88 363 single-nucleotide polymorphisms (SNPs) in the core genome, with 22 567 of these being parsimony informative for phylogenetic reconstruction (Fig. 1). We rooted the tree with *P. syringae* pv. *phaseolicola* 1448a (*Pph* 1448A) as the outgroup based on previous MLSA studies (Hwang *et al*., 2005). All but one branch had 100% bootstrap support, with *Ppi* 1704B and *Ppi* PP1 placed unambiguously into different clades. There was substantial homoplasy within this phylogeny with a consistency index of only 0.55, suggesting substantial recombination within portions of this tree. In contrast with the MLSA results, using phylogenetic data from the whole genome consistently clustered isolates from pv. *pisi* together with strong support. Taking all of this information together, the simplest explanation for the apparent convergence of *Ppi* strains is inaccurate inference of core genome evolution based on MLSA data, rather than convergent loss of toxins and gain of effectors during host jumps by these *Ppi* strains onto pea.

DISCUSSION

Pseudomonas syringae possesses a variety of tools within its virulence arsenal, including both type III effectors and pathways that produce nonribosomal peptide toxins (O'Brien *et al*. 2011). Our previous report hints at a trade-off between virulence mechanisms within diverse *P. syringae* strains designated as MLSA group II (Baltrus *et al*., 2011). Specifically, an ancestor for all group II strains appears to have acquired a set of three distinct toxin pathways (syringolin, syringopeptin, syringomycin). Coupled with this acquisition, extant strains appear to contain a smaller number of type III effectors on average than other *P. syringae* groups. Although this trend is apparent at the genome analysis level, to date there has been little direct insight into trade-offs between the presence of effectors and toxins within this group.

One significant data point that suggests such a trade-off is the observation that pathovar *pisi* (isolated from diseased pea plants) has subsequently lost all three toxins and contains the largest effector suite within group II (Baltrus *et al*., 2011). In addition, phylogenies built using MLSA sequences suggest that a second clade of *Ppi* exists that has independently lost all three toxins (Hwang *et al*., 2005). Evolutionary convergence of this loss only in *Ppi*, coupled with an independent increase in type III effector repertoires, implies that the maintenance of these toxins is detrimental to pea pathogenesis.

To explore the evolutionary relationships between *P. syringae* isolates capable of causing disease on pea, we have generated

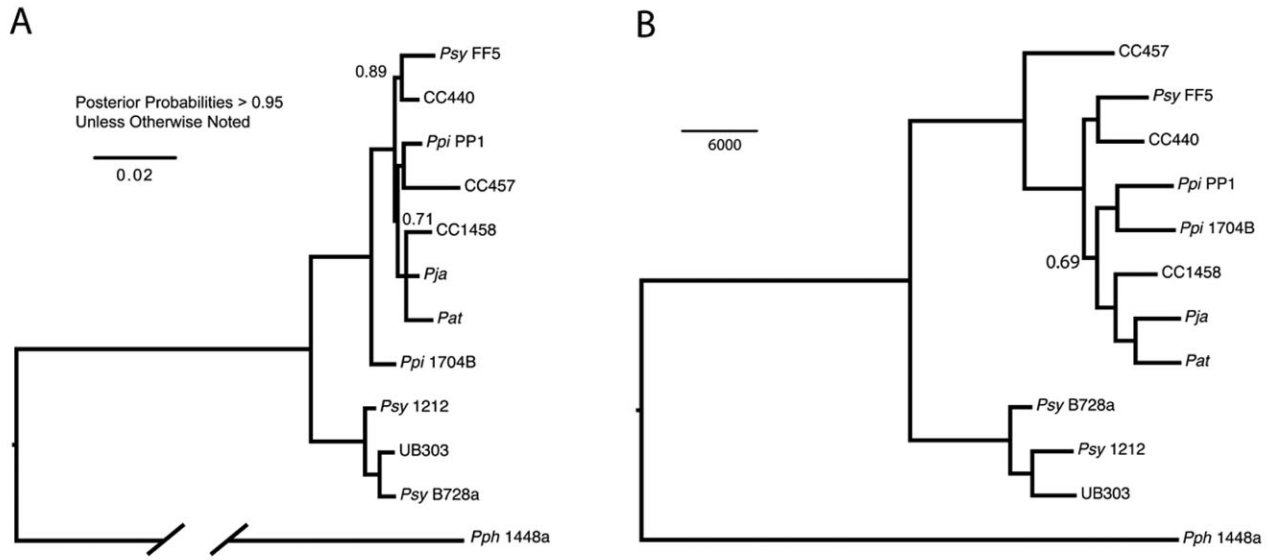

Fig. 1 Comparison of multi-locus sequence analysis (MLSA)-based and whole-genome single-nucleotide polymorphism (SNP)-based phylogenies for *Pseudomonas syringae* strains. (A) A Bayesian phylogeny of *P. syringae* MLSA group II strains using seven concatenated gene sequences as described by Baltrus *et al*. (2011). All posterior probabilities are >0.95, except where shown. Scale bar represents substitutions per site. (B) Maximum parsimony phylogeny built using SNPs determined from whole-genome sequences. Scale bar represents number of SNPs. In each case, *P. syringae* pv. *phaseolicola* 1448a (*Pph* 1448a) was used as an outgroup. *Pat* is *P. syringae* pv. *atrofaciens* DSM5025. Strains CC457, CC440, CC1458 and UB3303 are *P. syringae* isolated by Cindy Morris, but do not have pathovar designations. Assemblies for these strains can be found at NCBI accession numbers AVEB00000000, AVEC00000000, AVEN00000000, and AVDZ00000000. *Pja* is *P. syringae* pv. *japonica* MAFF301072 as described in Baltrus *et al*. (2011). *Psy* FF5 is *P. syringae* pv. *syringae* FF5 as described in Sohn *et al*. (2012). *Ppi* is *P. syringae* pv. *pisi*. All branches had a bootstrap support of 100 unless otherwise noted.

draft genome assemblies for two separate pathovars isolated from diseased peas and have compared suites of known virulence genes. We have performed phylogenetic analyses, comparing both MLSA and SNPs from whole-genome data, in order to identify evolutionary patterns within *Ppi*. Phylogenies built using sequence fragments from MLSA loci within these draft assemblies recapitulate previously reported phylogenetic relationships for *Ppi* 1704B and *Ppi* PP1, where these isolates are found in separate clusters within the *P. syringae* phylogeny (Hwang *et al*., 2005). However, a phylogeny built using SNP data from throughout the genome demonstrates that these two strains actually cluster together within the phylogeny, and thus any apparent convergence in toxin and effector suites is actually the product of shared evolutionary history. This pattern is further reinforced by comparison of type III effector suites, which substantially overlap between pv. *pisi* isolates, and demonstrate that the trade-off between effector number and toxin presence for *Ppi* compared with many other group II strains is consistent for multiple isolates within this pathovar.

What processes could cause incongruence between MLSA and whole-genome phylogenies? There are a limited number of informative loci within the MLSA genes across group II strains, and it appears that *Ppi* 1704B has a large number of unique SNPs compared with other closely related strains. When we investigate the MLSA data more closely, there are twice the number of single-nucleotide variants private to *Ppi* 1704B than in all but one other analysed strain. It is highly unlikely that this divergence has arisen as a result of sequencing errors, because these phylogenetic signals have been observed by multiple laboratories over numerous strains, as well as through whole-genome sequencing. This pattern could stem from recombination between *Ppi* 1704B and strains not represented within the sequence sample, as our whole-genome analyses indicate extensive recombination within this group. Such a result would mirror reports of recombination confounding phylogenetic analyses within pv. *tabaci* (Studholme, 2011). However, phylogenetic analysis of individual MLSA loci suggests that this incongruence is not the result of simple recombination across a single gene, because both *pisi* strains cluster for only three of six loci (*rpoD*, *gyrB* and *acnB*, data not shown).

These differences could also reflect changes to population dynamics within lineages. For instance, if one lineage experiences more population bottlenecks than closely related clades, genetic drift could increase fixation of single-nucleotide variants across loci throughout the genome. In a small sample of data, such as with MLSA loci, these differences could overwhelm

phylogenetic signals and lead to clustering of *Ppi* 1704B outside of other *Ppi* strains. When whole-genome data are used, the strength of this phylogenetic signal relative to noise increases and provides a better estimate of the phylogenetic branching history. With these results in mind and given the cost efficiency of modern sequencing technologies, every effort should be made to utilize whole-genome information before making evolutionary inferences.

In summary, although MLSA schemes work well for assigning *P. syringae* strains to broad groups, as demonstrated by Hwang *et al*. (2005), within-group relationships can be skewed relative to whole-genome approaches, and we should be cautious when making evolutionary inferences from these data.

EXPERIMENTAL PROCEDURES

Genomic DNA for both *Psy* 1212 and *Ppi* PP1 was obtained by similar methods to those described previously from overnight cultures initiated from single colonies (Baltrus *et al*., 2012). Sequence fragments were generated by an Illumina HiSeq (University of Arizona Sequencing Core, Tucson, AZ, USA) using 100-bp paired-end libraries, and assembled using SPAdes version 2.5.0 (Bankevich *et al*., 2012). Reads were not trimmed for quality prior to assembly. This Whole-Genome Shotgun project for *Ppi* PP1 has been deposited at DDBJ/EMBL/GenBank under the accession number AUZR00000000 (Benson *et al*., 2008). The version described in this article is version AUZR02000000. The Whole-Genome Shotgun project for *Psy* 1212 has been deposited at DDBJ/EMBL/GenBank under the accession number AVCR00000000. The version described in this article is version AVCR02000000. For all analyses, we use an additional publicly available genome for pv. *atrofaciens* (GenBank accession number AWUI00000000).

Sequences used in the phylogenetic comparisons, but not described for the first time in this paper, are publicly available (Baltrus *et al*., 2011; Feil *et al*., 2005; Joardar *et al*., 2005; Sohn *et al*., 2012; D. A. Baltrus *et al*., unpublished data). Phylogenetic analyses were performed on concatenated MLSA loci as described previously using concatenated fragments from seven MLSA genes across a diverse array of isolates from *P. syringae* group I (Baltrus *et al*., 2011). Briefly, maximum likelihood and parsimony analyses were performed with PHYLIP by implementing SEQBOOT to generate 100 bootstrapped datasets, dnaML or dnapars, to yield maximum likelihood or parsimony phylogenies, respectively, and CONSENSE to create a majority rules consensus tree from the bootstrapped datasets (Felsenstein, 1989). MRBAYES was used to perform Bayesian phylogenetic analysis with flat priors, with a burn-in period of 250 000 generations after convergence after 1 000 000 total generations (Ronquist *et al*., 2012).

A whole-genome phylogeny was generated using SNPs from loci present in all genomes employing alignments of sequencing reads to a reference genome, followed by SNP determinations and, finally, merging of all the SNP data based on the position in the reference. Shared orthologous SNPs, considered as the core genome, were identified and consolidated using In-Silico Genotyper version 16.8.1, a program that creates a matrix of sequence variant differences (SNPs and indels) among taxa by merging single sample variant files (X. X. Beckstrom-Sternberg *et al*., unpublished data). The processing occurs in several steps and is merged into a pipeline using GATK-Queue (http://www.broadinstitute.org/gatk/index.php). During the alignment step, next-generation sequencing reads for whole genomes were aligned against a multi-FASTA reference, comprising the *P. syringae* B728a genome and five plasmids (pB76-81, pNCPPB880-40, pPDDCC3357-6, pPSR1 and pPT14-32) using BWA-0.6.2 (Li and Durbin, 2009); in addition, one publicly sequenced genome, *Pph* 1448A (including NC_005773.3—chromosome; NC_007274.1—large plasmid; and NC_007275.1—small plasmid), was aligned against the same multi-FASTA reference using nucmer in MUMmer 3.23 (Kurtz *et al*., 2004). Duplicated regions in the reference were detected by aligning the reference to itself with nucmer. Variants falling within these regions were filtered out. SNPs were called from next-generation sequencing alignments with Unified Genotyper in GATK 2.1–13-g1706365 (DePristo *et al*., 2011; McKenna *et al*., 2010), and SNPs in *Pph* 1448A were called with show-snps in MUMmer 3.23. All SNPs were output in variant call format (VCF). Ambiguous SNP calls were marked in the variant files using three criteria, including quality, depth of coverage and allele frequency. Variant files from all taxa were merged, creating a matrix of variants by reference position in all taxa. Any variant positions missing for a particular taxon were given a call of either the reference state or no-coverage (.) based on coverage and quality to complete the matrix. Because of the increased likelihood of alignment errors being close to one another, we had a mismatch cut-off of five nucleotides, meaning that no two SNPs could be closer to each other than five bases. Nonetheless, analysis without a mismatch cut-off yielded 234 853 SNPs, but no difference in tree topology and minimal effect on branch lengths (data not shown). The SNP matrix was analysed using maximum parsimony in PAUP* v4.0b and branch support based on 1000 bootstrap replicates (Swofford, 2002).

Type III effectors and toxins were identified as described by Baltrus *et al*. (2011). Briefly, draft genome assemblies for *Psy* 1212 and *Ppi* PP1 were queried using tBLASTn with queries consisting of protein sequences for known type III effector families or for key loci within toxin pathways. Each BLAST hit was validated by hand for copy number, identity and completeness. Briefly, all BLAST results were visually inspected to make sure that each sequence displayed high similarity to only one region in the assembly, that this similar region was not part of a larger open reading frame, and that the length of this region was greater than 40% of the length of the original query sequence.

ACKNOWLEDGEMENTS

Funding was provided by startup funds to David Baltrus from the University of Arizona. We thank Ryan Sprissler and the University of Arizona Genetics Core Facility for technical help. We also thank Dawn Arnold (University of the West of England) for bacterial strains, Eric Johnson and Paul Etter (University of Oregon) for supplying genomic data that motivated this work, and Jeff Dangl (UNC Chapel Hill, Chapel Hill, NC, USA) for providing unpublished data.

REFERENCES

Arnold, D.L., Gibbon, M.J., Jackson, R.W., Wood, J.R., Brown, J., Mansfield, J.W., Taylor, J.D. and Vivian, A. (2001) Molecular characterization of *avrPphD*, a

widely-distributed gene from *Pseudomonas syringae* pv. *phaseolicola* involved in non-host recognition by pea (*Pisum sativum*). *Physiol. Mol. Plant Pathol.* **58**, 55–62.

Baltrus, D.A., Nishimura, M.T., Romanchuk, A., Chang, J.H., Mukhtar, M.S., Cherkis, K., Roach, J., Grant, S.R., Jones, C.D. and Dangl, J.L. (2011) Dynamic evolution of pathogenicity revealed by sequencing and comparative genomics of 19 *Pseudomonas syringae* isolates. *Plos Pathog.* **7**, e1002132.

Baltrus, D.A., Nishimura, M.T., Dougherty, K.M., Biswas, S., Mukhtar, M.S., Vicente, J., Holub, E.B. and Dangl, J.L. (2012) The molecular basis of host specialization in bean pathovars of *Pseudomonas syringae*. *Mol. Plant–Microbe Interact.* **25**, 877–888.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., Pyshkin, A.V., Sirotkin, A.V., Vyahhi, N., Tesler, G., Alekseyev, M.A. and Pevzner, P.A. (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **19**, 455–477.

Benson, D.B., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J. and Wheeler, D.L. (2008) GenBank. *Nucleic Acids Res.* **36** (Database issue), D25–D30.

DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C., Philippakis, A.A., del Angel, G., Rivas, M.A., Hanna, M., McKenna, A., Fennell, T.J., Kernytsky, A.M., Sivachenko, A.Y., Cibulskis, K., Gabriel, S.B., Altshuler, D. and Daly, M.J. (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491–498.

Feil, H., Feil, W.S., Larimer, F., DiBartolo, G., Copeland, A., Lykidis, A., Trong, S., Nolan, M., Goltsman, E., Thiel, J., Malfatti, S., Loper, J.E., Lapidus, A., Detter, J.C., Land, M., Richardson, P.M., Krypides, N.C., Ivanova, N. and Lindow, S.E. (2005) Comparison of the complete genome sequences of *Pseudomonas syringae* pv. *syringae* B728a and pv. *tomato* DC3000. *Proc. Natl. Acad. Sci. USA*, **102**, 11 064–11 069.

Felsenstein, J. (1989) PHYLIP—Phylogeny inference package (version 3.2). *Cladistics*, **5**, 164–166.

Hwang, M.S., Morgan, R.L., Sarkar, S.F., Wang, P.W. and Guttman, D.S. (2005) Phylogenetic characterization of virulence and resistance phenotypes of *Pseudomonas syringae*. *Appl. Environ. Microbiol.* **71**, 5182–5191.

Joardar, V., Lindeberg, M., Jackson, R.W., Selengut, J., Dodson, R., Brinkac, L.M., Daugherty, S.C., Deboy, R., Durkin, A.S., Giglio, M.G., Madupu, R., Nelson, W.C., Rosovitz, M.J., Sullivan, S., Crabtree, J., Creasy, T., Davidsen, T., Haft, D.H., Zafar, N., Zhou, L., Halpin, R., Holley, T., Khouri, H., Feldblyum, T., White, O., Fraser, C.M., Chatterjee, A.K., Cartinhour, S., Schneider, D.J., Mansfield, J., Collmer, A. and Buell, C.R. (2005) Whole-genome sequence analysis of *Pseudomonas syringae* pv. phaseolicola 1448A reveals divergence among pathovars in genes involved in virulence and transposition. *J. Bacteriol.* **187**, 6488–6498.

Kurtz, S., Phillippy, A., Delcher, A.L., Smoot, M., Shumway, M., Antonescu, C. and Salzberg, S.L. (2004) Versatile and open software for comparing large genomes. *Genome Biol.* **5**, R12.

Li, H. and Durbin, R. (2009) Fast and accurate short read alignment with Burrows–Wheeler Transform. *Bioinformatics*, **25**, 1754–1760.

Mazarei, M. and Kerr, A. (1990) Distinguishing pathovars of *Pseudomonas syringae* on peas: nutritional, pathogenicity, and serological tests. *Plant Pathol.* **39**, 278–285.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M. and DePristo, M.A. (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303.

O'Brien, H.E., Thakur, S. and Guttman, D.S. (2011) Evolution of plant pathogenesis in *Pseudomonas syringae*: a genomics perspective. *Annu. Rev. Phytopathol.* **49**, 269–289.

O'Brien, H.E., Thakur, S., Gong, Y., Fung, P., Zhang, J., Wang, P.W., Yong, C., Scortichini, M. and Guttman, D.S. (2012) Extensive remodeling of the *Pseudomonas syringae* pv. *avellanae* type III secretome associated with two independent host shifts onto hazelnut. *BMC Microbiol.* **12**, 141.

Ronquist, F., Teslenko, M., Van der Mark, P., Ayres, D.L., Darling, A., Hohna, S., Larget, B., Liu, L., Suchard, M.A. and Huelsenbeck, J.P. (2012) MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* **61**, 539–542.

Sarkar, S.F. and Guttman, D.S. (2004) Evolution of the core genome of *Pseudomonas syringae*, a highly clonal, endemic plant pathogen. *Appl. Environ. Microbiol.* **70**, 1999–2012.

Sarkar, S.F., Gordon, J.S., Martin, G.B. and Guttman, D.S. (2006) Comparative genomics of host-specific virulence in *Pseudomonas syringae*. *Genetics*, **174**, 1041–1056.

Sohn, K.H., Jones, J.D. and Studholme, D.J. (2012) Draft genome sequence of *Pseudomonas syringae* pathovar *syringae* strain FF5, causal agent of stem tip dieback disease on ornamental pear. *J. Bacteriol.* **194**, 3733–3734.

Studholme, D.J. (2011) Application of high-throughput genome sequencing to intrapathovar variation in *Pseudomonas syringae*. *Mol. Plant Pathol.* **12**, 829–838.

Swofford, D.L. (2002) *PAUP*. Phylogenetic Analysis Using Parsimony (* and Other Methods), Version 4.0*. Sunderland: Sinauer Associates.