New Phytologist

Comparative genomics of *Pseudomonas syringae* reveals convergent gene gain and loss associated with specialization onto cherry (*Prunus avium*)

Michelle T. Hulin[1,2], Andrew D. Armitage[1], Joana G. Vicente[3], Eric B. Holub[3], Laura Baxter[3], Helen J. Bates[1], John W. Mansfield[4], Robert W. Jackson[2] and Richard J. Harrison[1,2]

[1]NIAB EMR, New Road, East Malling, ME19 6BJ, UK; [2]School of Biological Sciences, University of Reading, Reading, RG6 6AJ, UK; [3]School of Life Sciences, Warwick Crop Centre, University of Warwick, Wellesbourne, CV35 9EF, UK; [4]Faculty of Natural Sciences, Imperial College London, London, SW7 2AZ, UK

Author for correspondence:
Richard J. Harrison
Tel: +44 (0)1732 843833
Email: richard.harrison@emr.ac.uk

Received: 9 *January 2018*
Accepted: 22 *March 2018*

New Phytologist (2018) **219:** 672–696
doi: 10.1111/nph.15182

Key words: avirulence, bacterial canker, comparative genomics, host specialization, prediction, *Pseudomonas*, toxins, type III effectors.

Summary

- Genome-wide analyses of the effector- and toxin-encoding genes were used to examine the phylogenetics and evolution of pathogenicity amongst diverse strains of *Pseudomonas syringae* causing bacterial canker of cherry (*Prunus avium*), including pathovars *P. syringae* pv *morsprunorum* (*Psm*) races 1 and 2, *P. syringae* pv *syringae* (*Pss*) and *P. syringae* pv *avii*.
- Phylogenetic analyses revealed *Psm* races and *P. syringae* pv *avii* clades were distinct and were each monophyletic, whereas cherry-pathogenic strains of *Pss* were interspersed amongst strains from other host species.
- A maximum likelihood approach was used to predict effectors associated with pathogenicity on cherry. *Pss* possesses a smaller repertoire of type III effectors but has more toxin biosynthesis clusters than *Psm* and *P. syringae* pv *avii*. Evolution of cherry pathogenicity was correlated with gain of genes such as *hopAR1* and *hopBB1* through putative phage transfer and horizontal transfer respectively. By contrast, loss of the *avrPto/hopAB* redundant effector group was observed in cherry-pathogenic clades. Ectopic expression of *hopAB* and *hopC1* triggered the hypersensitive reaction in cherry leaves, confirming computational predictions.
- Cherry canker provides a fascinating example of convergent evolution of pathogenicity that is explained by the mix of effector and toxin repertoires acting on a common host.

Introduction

Pseudomonas syringae is a species complex, associated with plants and the water cycle, comprising several divergent clades that frequently recombine (Young, 2010; Parkinson *et al.*, 2011; Berge *et al.*, 2014; Baltrus *et al.*, 2017). It is a globally important pathogen, causing disease on over 180 different host species. *P. syringae* is responsible for recurring chronic diseases in perennial crops, such as cherry canker (Lamichhane *et al.*, 2014), and also sporadic outbreaks on annual crops, such as bacterial speck of tomato (*Solanum lycopersicum*) (Şahin, 2001). Individual strains are reported to be specialized and assigned to over 60 host-specific pathovars; some of these are further divided into races which show host genotype specificity (Joardar *et al.*, 2005). Strains also exist that can infect a variety of crop species, indicating that specialization is not always the norm (Monteil *et al.*, 2013; Bartoli *et al.*, 2015a,b). This complexity makes *P. syringae* an important model to study the evolution of host specificity (O'Brien *et al.*, 2011; Mansfield *et al.*, 2012).

High-throughput sequencing has become a major tool in bacteriology (Edwards & Holt, 2013). With the increasing number of genomes available, population-level analyses allow complex evolutionary questions to be addressed, such as how disease epidemics emerge and what ecological processes drive the evolution of pathogenicity (Guttman *et al.*, 2014; Monteil *et al.*, 2016). Before genomic methods were available, mutational studies of *P. syringae* were used to identify functional virulence factors in pathogenesis, such as type III secretion system effector (T3E) repertoires and toxins (Lindgren, 1997; Bender *et al.*, 1999). Some T3Es were also shown to act as avirulence (*avr*) factors when detected by a corresponding pathogen recognition (R) protein in the host, leading to effector-triggered immunity (ETI) (Jones & Dangl, 2006). ETI is often associated with the hypersensitive response (HR) which is a cell death mechanism important in preventing pathogen spread (Morel & Dangl, 1997). Evidence suggests that *P. syringae* has evolved a functionally redundant repertoire of effectors, which is postulated to allow pathogen populations to lose/modify expendable *avr* elicitors, with minimal impact on overall virulence (Arnold & Jackson, 2011). It is proposed that as pathogen lineages specialize, they fine-tune their effector repertoires to maximize fitness in this niche by ensuring adequate growth and transmission, whilst avoiding detection by the plant immune system. Host range becomes restricted because the pathogen may lose effectors important for disease on other hosts or gain effectors detected by other plant species (Schulze-

Lefert & Panstruga, 2011). Many genomics studies have therefore focused on linking variation in virulence gene complements with particular diseases (Baltrus *et al.*, 2011, 2012; O'Brien *et al.*, 2012).

Much of the understanding of *P. syringae*–plant interactions has been achieved using herbaceous plant models. Woody pathosystems provide a greater challenge (Lamichhane *et al.*, 2014). Population genomics of *P. syringae* pv *actinidiae*, the causal agent of kiwifruit canker, revealed that three pathogenic clades, with distinct effector sets, have arisen during kiwifruit cultivation (McCann *et al.*, 2013, 2017). A study of the olive pathogen *P. syringae* pv *savastanoi* revealed that the *hopBL* effector family is overrepresented in wood-infecting pathovars (Matas *et al.*, 2014). Genes involved in the metabolism of aromatic compounds, phytohormone production, tolerance to reactive oxygen species and sucrose metabolism have also been associated with virulence on woody tissues (Green *et al.*, 2010; Bartoli *et al.*, 2015b; Buonaurio *et al.*, 2015; Nowell *et al.*, 2016).

This study used genomics to examine the evolution of strains that cause bacterial canker on sweet and wild cherry (both *Prunus avium*). Clades of *P. syringae* that constitute the main causal agents of bacterial canker include *P. syringae* pv *morsprunorum* (*Psm*) race 1 and race 2 and *P. syringae* pv *syringae* (*Pss*) (Bull *et al.*, 2010; Bultreys & Kaluzna, 2010). In addition, *P. syringae* pv *avii* causes bacterial canker of wild cherry (Ménard *et al.*, 2003). The cherry-pathogenic clades of *P. syringae* are reported to exhibit differences in virulence, host range and lifestyle (Crosse & Garrett, 1966; Scortichini, 2010), making the *P. syringae*–cherry interaction a good pathosystem to study convergent gain of pathogenicity. The genomic analysis has been coupled with robust pathogenicity testing (Hulin *et al.*, 2018) and functional analysis of potential avirulence (*avr*) genes. This study provides a proof of concept that genomics-based methods can be used to identify candidate genes involved in disease and will likely become the major tool in disease monitoring, diagnostics and host range prediction.

Materials and Methods

Bacteria, plants and pathogenicity tests

Methods used for bacterial culture and sources of plants were as described in Hulin *et al.* (2018) and are detailed in Supporting Information Methods S1. Plant species utilized included *P. avium* L. and *Nicotiana tabacum* L. *Pseudomonas* strains are listed in Table 1. *Escherichia coli* was grown on lysogeny broth (LB) agar plates and grown overnight at 37°C. Antibiotic concentrations: kanamycin, 50 $\mu g\ ml^{-1}$; gentamycin, 10 $\mu g\ ml^{-1}$; spectinomycin, 100 $\mu g\ ml^{-1}$; nitrofurantoin, 100 $\mu g\ ml^{-1}$. X-gal was used at a concentration of 80 $\mu g\ ml^{-1}$. Tables S1–S3 list the *P. syringae* mutants, plasmids and primers used in this study.

Pathogenicity tests, performed on detached cherry leaves and analysed as in Hulin *et al.* (2018) are described in Methods S1.

Genome sequencing, assembly and annotation

Bioinformatics commands for analyses performed in this paper are available on Github (https://github.com/harrisonlab/pseudomonas/). Genome sequencing using Illumina and genome assembly were performed as in Hulin *et al.* (2018). For long-read sequencing, PacBio (Pacific Biosystems, Menlo Park, CA, USA) and MinION (Oxford Nanopore, Oxford, UK) were used. High molecular weight DNA was extracted using a cetyltrimethylammonium bromide method (Feil *et al.*, 2012). For the PacBio sequencing of strains R1-5244, R2-leaf and syr9097, DNA samples were sent to the Earlham Institute (Norwich) for PacBio RSII sequencing.

For MinION sequencing of *Psm* R1-5300, the DNA library was prepared using the RAD001 rapid-prep kit (Oxford Nanopore) and run on the Oxford Nanopore MinION, flowcell vR9.5 followed by basecalling using Metrichor (Oxford Nanopore). MinION reads were extracted from Fast5 files using Poretools (Loman & Quinlan, 2014). The sequencing reads for both long-read technologies were trimmed and assembled using Canu (Berlin *et al.*, 2015), and Circlator was used to circularize contigs (Hunt *et al.*, 2015). The assemblies were polished using error-corrected Illumina reads with Bowtie2, Samtools and Pilon 1.17 (Li *et al.*, 2009; Langmead & Salzberg, 2013; Walker *et al.*, 2014).

Plasmid profiling was performed using an alkaline-lysis method and gel electrophoresis (Moulton *et al.*, 1993; Neale *et al.*, 2013). Genomes were submitted to GenBank and accession numbers are listed in Table 1.

Orthology analysis

OrthoMCL (Li *et al.*, 2003) was used to identify orthologous genes. All genomes were re-annotated using Rast (Aziz *et al.*, 2008) to ensure similar annotation quality. For this reason, the Illumina short-read assemblies of the four long-read sequenced genomes (R1-5244, R1-5300, R2-leaf and syr9097) were used in orthology analysis. OrthoMCL was run with default settings and a 50 residue cut-off length. All Rast annotated protein files used in this analysis are available on Github (https://github.com/harrisonlab/pseudomonas/).

Phylogenetic and genomic analysis of *Pseudomonas syringae*

Nucleotide sequences of single-copy genes present in all genomes were aligned using ClustalW (Larkin *et al.*, 2007) and trimmed using Gblocks (Castresana, 2000). Gene alignments were concatenated using Geneious (Kearse *et al.*, 2012). RAxML-avx v.8.1.15 (Stamatakis, 2014) was used in partitioned mode to build the maximum likelihood phylogeny, with a general time reversible (GTR) gamma model of substitution and 100 nonparametric bootstrap replicates. To detect core genes that may have undergone recombination, the program Geneconv (Sawyer, 1989) was used as in Yu *et al.* (2016). Whole-genome alignments were performed using progressiveMauve (Darling *et al.*, 2010).

Virulence and mobility gene identification

All T3E-encoding protein sequences were downloaded from pseudomonas-syringae.org, including the recent classification of

Table 1 List of bacterial strains used in this study, including a range of cherry pathogens and nonpathogens

Strain	Pathovar	Race	PG	Isolation source	Isolator	*Prunus* cv	Sequenced	Pathogenicity tested on cherry (*Prunus avium*)	Accession number
avii5271	***avii***		**1**	***Prunus avium***	**Garrett, 1990, UK**	**Wild cherry**	**This work**	**Vicente *et al.* (2004)**	**NBAO00000000**
R1-5270	***morsprunorum***	**1**	**3**	***Prunus avium***	**Garrett, 1990, UK**	**Wild cherry**	**This work**	**Vicente *et al.* (2004)**	**NBAN00000000**
R2-7968A	***morsprunorum***	**2**	**1**	***Prunus avium***	**Vicente, 2000, UK**	**Wild cherry**	**This work**	**Vicente *et al.* (2004)**	**NBAI00000000**
R2-9095	***morsprunorum***	**2**	**1**	***Prunus avium***	**Roberts, 2010, UK**	**Wild cherry**	**This work**	**M. Hulin, pers. obs.**	**MLED00000000**
syr5264	***syringae***		**2**	***Prunus avium***	**Garrett, 1990, UK**	**Wild cherry**	**This work**	**Vicente *et al.* (2004)**	**NBAQ00000000**
syr5275	***syringae***		**2**	***Prunus avium***	**Garrett, 1990, UK**	**Wild cherry**	**This work**	**Vicente *et al.* (2004)**	**NBAP00000000**
syr7928A	***syringae***		**2**	***Prunus avium***	**Vicente, 2000, UK**	**Wild cherry**	**This work**	**Vicente *et al.* (2004)**	**NBAL00000000**
syr8094A	***syringae***		**2**	***Prunus avium***	**Vicente, 2001, UK**	**Wild cherry**	**This work**	**Vicente *et al.* (2004)**	**NBAK00000000**
Ps-7928C	Unknown		2	*Prunus avium*	Vicente, 2000, UK	Wild cherry	This work	Vicente *et al.* (2004)	NBAM00000000
Ps-7969	Unknown		2	*Prunus avium*	Vicente, 2000, UK	Wild cherry	This work	Vicente *et al.* (2004)	NBAJ00000000
R1-5244	***morsprunorum***	**1**	**3**	***Prunus avium***	**Crosse, 1960, UK**	**Unknown**	**This work**	**Hulin *et al.* (2018)**	**CP026557–CP026561**
R1-5300	*morsprunorum*	1	3	*Prunus domestica*	Prunier, UK	Victoria	This work	**Hulin *et al.* (2018)**	MLEN00000000
R2-leaf	***morsprunorum***	**2**	**1**	***Prunus avium***	**Hulin, 2014, UK**	**Napoleon**	**This work**	**Hulin *et al.* (2018)**	**CP026562–CP026567**
syr9097	***syringae***		**2**	***Prunus avium***	**Roberts, 2010, UK**	**Unknown**	**This work**	**Hulin *et al.* (2018)**	**CP026568**
syr2675	*syringae*		2	*Phaseolus vulgaris*	1965, Kenya		This work	This work	MLEX00000000
syr2676	*syringae*		2	*Phaseolus vulgaris*	1990, Lesotho		This work	nt	MLEY00000000
syr2682	*syringae*		2	*Phaseolus vulgaris*	1990, Lesotho		This work	nt	MLFA00000000
syr3023	*syringae*		2	*Syringa vulgaris*	1950, UK		This work	This work	MLFD00000000
syr100	*syringae*		2	*Phaseolus lunatus*	1962, Kenya		This work	This work	MLEV00000000
R1-9326	*morsprunorum*	1	3	*Prunus domestica*	Roberts, 2011, UK	Victoria	Hulin *et al.* (2018)	Hulin *et al.* (2018)	MLEO00000000
R1-9629	*morsprunorum*	1	3	*Prunus domestica*	Roberts, 2012, UK	Victoria	Hulin *et al.* (2018)	Hulin *et al.* (2018)	MLEP00000000
R1-9646	***morsprunorum***	**1**	**3**	***Prunus avium***	**Roberts, 2012, UK**	**Stella**	**Hulin *et al.* (2018)**	**Hulin *et al.* (2018)**	**MLEE00000000**
R1-9657	*morsprunorum*	1	3	*Prunus avium*	Roberts, 2012, UK	Kiku-Shidare	Hulin *et al.* (2018)	Hulin *et al.* (2018)	MLEF00000000
R2-5255	***morsprunorum***	**2**	**1**	***Prunus avium***	**Prunier, UK**	**Napoleon**	**Hulin *et al.* (2018)**	**Hulin *et al.* (2018)**	**MLEC00000000**
R2-5260	***morsprunorum***	**2**	**1**	***Prunus avium***	**Garrett, UK**	**Roundel**	**Hulin *et al.* (2018)**	**Hulin *et al.* (2018)**	**MLEG0000000**
R2-SC214	***morsprunorum***	**2**	**1**	***Prunus avium***	**Roberts, 1983, UK**	**Wild cherry**	**Hulin *et al.* (2018)**	**Hulin *et al.* (2018)**	**MLEI00000000**
syr9293	***syringae***		**2**	***Prunus domestica***	**Roberts, 2011, UK**	**Victoria**	**Hulin *et al.* (2018)**	**Hulin *et al.* (2018)**	**MLEQ00000000**
syr9630	***syringae***		**2**	***Prunus domestica***	**Roberts, 2012, UK**	**Victoria**	**Hulin *et al.* (2018)**	**Hulin *et al.* (2018)**	**MLER00000000**
syr9644	***syringae***		**2**	***Prunus avium***	**Roberts, 2012, UK**	**Stella**	**Hulin *et al.* (2018)**	**Hulin *et al.* (2018)**	**MLEK00000000**
syr9654	***syringae***		**2**	***Prunus domestica***	**Roberts, 2012, UK**	**Victoria**	**Hulin *et al.* (2018)**	**Hulin *et al.* (2018)**	**MLES00000000**
syr9656	***syringae***		**2**	***Prunus avium***	**Roberts, 2012, UK**	**Kiku-Shidare**	**Hulin *et al.* (2018)**	**Hulin *et al.* (2018)**	**MLEM00000000**
syr9659	***syringae***		**2**	***Prunus avium***	**Roberts, 2012, UK**	**Kiku-Shidare**	**Hulin *et al.* (2018)**	**Hulin *et al.* (2018)**	**MLEL00000000**

Table 1 (Continued)

Strain	Pathovar	Race	PG	Isolation source	Isolator	*Prunus* cv	Sequenced	Pathogenicity tested on cherry (*Prunus avium*)	Accession number
Ps-9643	Unknown		1	*Prunus domestica*	Roberts, 2012, UK	Victoria	Hulin *et al.* (2018)	Hulin *et al.* (2018)	MLET00000000
avii3846	***avii***		**1**	***Prunus avium***	**1991, France**	**Wild cherry**	**Nowell *et al.* (2016)**	**Ménard *et al.* (2003)**	**LIIJ00000000**
R1-2341	***morsprunorum***	**1**	**3**	***Prunus cerasus***	**1988, Hungary**	**Unknown**	**Nowell *et al.* (2016)**	**nt**	**LIIB00000000**
R1-5269	***morsprunorum***	**1**	**3**	***Prunus cerasus***	**Garrett, 1990, UK**	**Wild cherry**	**Nowell *et al.* (2016)**	**Vicente *et al.* (2004)**	**LIHZ00000000**
R2-5261	***morsprunorum***	**2**	**1**	***Prunus avium***	**Garrett, UK**	**Roundel**	**Nowell *et al.* (2016)**	**Vicente *et al.* (2004)**	**LIIA00000000**
R2-302280	***morsprunorum***		**1**	***Prunus domestica***	**USA**	**Unknown**	**Baltrus *et al.* (2011)**	**Gilbert *et al.* (2009)***	**AEAE00000000**
syr2339	***syringae***		**2**	***Prunus avium***	**1984, Hungary**	**Unknown**	**Nowell *et al.* (2016)**	**nt**	**LIHU00000000**
syr7872	***syringae***		**2**	***Prunus domestica***	**Lewis, 2000, UK**	**Opal**	**Nowell *et al.* (2016)**	**Vicente *et al.* (2004)**	**LIHS00000000**
syr7924	***syringae***		**2**	***Prunus avium***	**Vicente, 2000, UK**	**Wild cherry**	**Nowell *et al.* (2016)**	**Vicente *et al.* (2004)**	**LIHR00000000**
acer302273	*aceris*		2	*Acer* sp.	USA		Baltrus *et al.* (2011)	nt	AEAO00000000
acti18884	*actinidiae*		1	*Actinidia deliciosa*	2010, New Zealand		McCann *et al.* (2013)	nt	AOKO00000000
acti19073	*actinidiae*		1	*Actinidia deliciosa*	1998, Korea		McCann *et al.* (2013)	nt	AOJR00000000
acti212056	*actinidiae*		1	*Actinidia deliciosa*	2012, Japan		Sawada *et al.* (2015)	nt	BBWG00000000
acti302091	*actinidiae*		1	*Actinidia deliciosa*	1984, Japan		Baltrus *et al.* (2011)	nt	AEAL00000000
actiCRAFRU	*actinidiae*		1	*Actinidia deliciosa*	2010, Italy		Butler *et al.* (2013)	nt	ANGD00000000
actiNCPPB3871	*actinidiae*		1	*Actinidia deliciosa*	1992, Italy		Marcelletti *et al.* (2011)	nt	ANGD00000000
aes089323	*aesculi*		3	*Aesculus hippocastanum*	India, 1980		Baltrus *et al.* (2011)	nt	AEAD00000000
aes2250	*aesculi*		3	*Aesculus hippocastanum*	2008, UK		Green *et al.* (2010)	nt	ACXT00000000
aes3681	*aesculi*		3	*Aesculus hippocastanum*	1969, India		Green *et al.* (2010)	nt	ACXS00000000
amy3205	*amygdali*		3	*Prunus dulcis*	1967, Greece		Bartoli *et al.* (2015a)	nt	JYHB00000000
amyICMP3918	*amygdali*		3	*Prunus dulcis*	Panagopoulos, 1967, Greece		Thakur *et al.* (2016)	nt	LJPQ00000000
avelBP631	*avellanae*		1	*Corylus avellana*	1976, Greece		O'Brien *et al.* (2012)	Hulin *et al.* (2018)	AKBS00000000
avelVe037	*avellanae*		2	*Corylus avellana*	1990, Italy		O'Brien *et al.* (2012)	nt	AKCJ00000000
BRIP34876	Unknown		2	*Hordeum vulgare*	1971, Australia		Gardiner *et al.* (2013)	nt	AMXK00000000
castCFBP4217	*castaneae*		3	*Castanea crenata*	1977, Japan		Nowell *et al.* (2016)	nt	LIIH00000000
CC1416	Unknown		1	Epilithon	USA		Baltrus *et al.* (2014b)	nt	AVEP00000000
CC1544	Unknown		1	Lake water	France		Baltrus *et al.* (2014b)	nt	AVEI00000000
CC1559	Unknown		1	Snow	France		Baltrus *et al.* (2014b)	nt	AVEG00000000
CC94	Unknown		2	*Cantaloupe*	France		Baltrus *et al.* (2014b)	nt	AVEA00000000

Table 1 (Continued)

Strain	Pathovar	Race	PG	Isolation source	Isolator	*Prunus* cv	Sequenced	Pathogenicity tested on cherry (*Prunus avium*)	Accession number
cera6109	*cerasicola*		3	*Prunus yedoensis*	1995, Japan		Nowell *et al.* (2016)	nt	LIIG00000000
ceraICMP17524	*cerasicola*		3	*Prunus yedoensis*	Japan		Thakur *et al.* (2016)	nt	LJQA00000000
ciccICMP5710	*ciccaronei*		3	*Ceratonia siliqua*	Italy		Thakur *et al.* (2016)	nt	LJPY00000000
cunnICMP11894	*cunninghamiae*		3	*Cunninghamia lanceolata*	China		Thakur *et al.* (2016)	nt	LJQE00000000
daphICMP9757	*daphniphylli*		3	*Daphniphyllum teijsmannii*	Japan		Thakur *et al.* (2016)	nt	LJQF00000000
delphi569	*delphinii*		1	*Delphinium* sp.	New Zealand		Thakur *et al.* (2016)	nt	LJQH00000000
dendro3226	*dendropanacis*		3	*Dendropanax trifidus*	1979, Japan		Bartoli *et al.* (2015a)	nt	JYHG00000000
dendro4219	*dendropanacis*		3	*Dendropanax trifidus*	1981, Japan		Bartoli *et al.* (2015a)	nt	JYHD00000000
dendro9150	*dendropanacis*		3	*Dendropanax trifidus*	Japan		Thakur *et al.* (2016)	nt	LJQG00000000
erio4455	*eriobotryae*		3	*Eriobotrya japonica*	USA		Thakur *et al.* (2016)	nt	LJQI00000000
glyR4	*glycinea*		3	*Glycine max*	Cross, 1960, USA		Qi *et al.* (2011)	nt	AEGH00000000
ICMP19498	Unknown		3	*Actinidia deliciosa*	2010, New Zealand		Visnovsky *et al.* (2016)	nt	LKCH00000000
lach301315	*lachrymans*		3	*Cucumis sativus*	Japan		Baltrus *et al.* (2011)	nt	AEAF00000000
lach302278	*lachrymans*		1	*Cucumis sativus*	USA		Baltrus *et al.* (2011)	nt	AEAM00000000
lapsaICMP3947	*lapsa*		2	*Zea* sp.	Unknown		Thakur *et al.* (2016)	nt	LJQQ00000000
meli6289	*meliae*		3	*Melia azedarach*	Japan		Thakur *et al.* (2016)	nt	LJQT00000000
morsU7805	*morsprunorum*		3	*Prunus mume*	Unknown		Mott *et al.* (2016)	nt	LGLQ00000000
myriAZ8448	*myricae*		3	*Myrica rubra*	Japan		Thakur *et al.* (2016)	nt	LGLA00000000
neriiICMP16943	*savastanoi*		3	*Olea europea*	Spain		Thakur *et al.* (2016)	nt	LJQW00000000
paniLMG2367	*panici*		2	*Panicum miliaceum*	Unknown		Liu *et al.* (2012)	nt	ALAC00000000
papu1754	*papulans*		2	*Malus sylvestris*	1973, Canada		Nowell *et al.* (2016)	nt	JYHI00000000
persNCPPB2254	*persicae*		1	*Prunus persica*	1972, France		Zhao *et al.* (2015)	nt	LAZV00000000
photICMP7840	*photiniae*		3	*Photinia glabra*	Japan		Thakur *et al.* (2016)	nt	LJQO00000000
pisiPP1	*pisi*		2	*Pisum sativum*	Japan		Baltrus *et al.* (2014a)	nt	AUZR00000000
phas1448a	*phaseolicola*		3	*Phaseolus vulgaris*	Teverson, 1965, Ethiopia		Joardar *et al.* (2005)	Hulin *et al.* (2018)	CP000058
rhapCFBP4220	*rhaphiolepidis*		3	*Rhaphiolepis umbellata*	1980, Japan		Nowell *et al.* (2016)	nt	LIHV00000000
RMA1	Unknown		1	*Aquilegia vulgaris*	Jackson, 2012, UK		Hulin *et al.* (2018)	Hulin *et al.* (2018)	MLEU00000000
sava3335	*savastanoi*		3	*Olea europea*	Stead, France		Rodríguez-Palenzuela *et al.* (2010)	nt	ADMI00000000
sava4352	*savastanoi*		3	*Olea europea*	Yugoslavia		Thakur *et al.* (2016)	nt	LGKR00000000

Table 1 (Continued)

Strain	Pathovar	Race	PG	Isolation source	Isolator	*Prunus* cv	Sequenced	Pathogenicity tested on cherry (*Prunus avium*)	Accession number
savaDAPP-PG722	*savastanoi*		3	*Olea europea*	Italy		Moretti *et al.* (2014)	nt	JOJV00000000
savaPseNe107	*savastanoi*		3	*Olea europea*	Balestra, Nepal		Bartoli *et al.* (2015a)	nt	JYHF00000000
soliICMP16925	*solidagae*		2	*Solidago altissima*	Japan		Thakur *et al.* (2016)	nt	JYHF00000000
syr1212	*syringae*		2	*Pisum sativum*	UK		Baltrus *et al.* (2014a)	This study	AVCR00000000
syr2340	*syringae*		2	*Pyrus* sp.	1985, Hungary		Nowell *et al.* (2016)	nt	LIHT00000000
syr41a	*syringae*		2	*Prunus armeniaca*	2011, France		Bartoli *et al.* (2015a)	nt	JYHJ00000000
syrB301D	*syringae*		2	*Pyrus communis*	Garrett, 1959, UK		Ravindran *et al.* (2015)	nt	CP005969
syrB64	*syringae*		2	*Triticum aestivum*	Wilcoxson, USA		Dudnik & Dudler (2014)	nt	ANZF00000000
syrB728a	*syringae*		2	*Phaseolus vulgaris*	1987, USA		Feil *et al.* (2005)	This work	CP000075
syrHS191	*syringae*		2	*Panicum miliaceum*	Hayward, Australia, 1969		Ravindran *et al.* (2015)	nt	CP006256
syrUMAF0158	*syringae*		2	*Mangifera indica*	Cazorla, 2010, Spain		Martínez-García *et al.* (2015)	nt	CP005970
thea3923	*theae*		1	*Camelia sinensis*	1974, Japan		Mazzaglia *et al.* (2012)	nt	AGNN00000000
tomaDC3000	*tomato*		1	*Solanum lycopersicum*	1960, UK		Buell *et al.* (2003)	nt	AE016853
tomaT1	*tomato*		1	*Solanum lycopersicum*	1986, Canada		Almeida *et al.* (2009)	nt	ABSM00000000
UB303	Unknown		2	Lake water	France		Baltrus *et al.* (2014b)	nt	AVDZ00000000
ulmiICMP3962	*ulmi*		3	*Ulmus* sp.	Yugoslavia		Thakur *et al.* (2016)	nt	LJRQ00000000
USA007	Unknown		1	Stream water	USA		Baltrus *et al.* (2014b)	nt	AVDY00000000
USA011	Unknown		1	Stream water	USA		Baltrus *et al.* (2014b)	nt	AVDX00000000

Pathovar designation, phylogroup, isolation information, cherry pathogenicity (reference for when tested; nt, not tested), publication of genome sequence and NCBI accession numbers are listed. Strains in bold were considered pathogenic in cherry. cv, cultivar of sweet cherry or plum. Long-read sequenced genomes are highlighted with shading. Strains are ordered, first with those sequenced in this study, followed by other *Pseudomonas* strains from cherry and plum used in plasmid profiling analysis and previously pathogenicity tested in Hulin *et al.* (2018). Next, further strains isolated from cherry and plum sequenced elsewhere are listed. Finally, the remaining strains were only used in comparative analysis. Note that all 108 genomes were used in initial orthology analysis but only 102 were used in the final phylogeny and comparative genomics.

*The pathogenic status of MAFF302280 on cherry is debated. This strain is reported to be the pathotype strain of *P. syringae* pv *morsprunorum* (*Psm*; Sawada *et al.*, 1999), so is assumed to be equivalent to CFBP 2351, NCPPB2995, ICMP5795 and LMG5075. The strain NCPPB2995 was reported to be potentially nonpathogenic (Gardan *et al.*, 1999). Whilst, the 'same' strain LMG5075 tested positive for pathogenicity in a recent publication (Gilbert *et al.*, 2009). There is no definite link showing that MAFF302280 is the same strain as the others listed as it is not linked to them in online databases (http://www.straininfo.net/) or taxonomy-focused publications (Bull *et al.*, 2010). It is assumed to be putatively pathogenic in this study owing to its close relatedness to other *Psm* R2 strains; however, further pathogenicity tests would be required to fully confirm this.

HopF effectors into four alleles (Lo *et al.*, 2016). tBLASTN (Altschul *et al.*, 1990) was used to search each genome for homologues with a score of ≥ 70% identity and ≥ 40% query length to at least one sequence in each effector family. Nucleotide sequences were extracted and manually examined for frameshifts or truncations. Disrupted genes were classed as pseudogenes. A heatmap of effector presence was generated using heatmap.2 in gplots (Warnes *et al.*, 2016). Interproscan (Quevillon *et al.*, 2005) identified protein domains, and Illustrator for Biological Sequences was used for visualization (Liu *et al.*, 2015). Genomic regions containing effectors were aligned using MAFFT (Katoh *et al.*, 2002).

A similar analysis was performed for phytotoxin and auxin biosynthesis, wood degradation, ice nucleation and plasmid-associated genes. Protein sequences were obtained from NCBI (Table S4) and blasted against the genome sequence as noted earlier. Prophage identification was performed using PHASTER (Arndt *et al.*, 2016).

Gain and loss analysis

Gain loss mapping engine (GLOOME) was used to plot the gain and loss of genes on the core-genome phylogeny (Cohen *et al.*, 2010). Effector genes were considered present even if predicted to be pseudogenes, as these can still be gained and lost. The optimization level was set to 'very high', a mixture model allowing variable gain and loss distributions was used and the distribution type was GENERAL_GAMMA_PLUS_INV. Highly probable events ($P \geq 0.80$) on the branches leading to cherry-pathogenic strains were extracted.

BayesTraits analysis

BAYESTRAITS v.2 was used to correlate T3E gene evolution with pathogenicity (Pagel, 2004). A binary matrix was created of effector family presence and pathogenicity of each strain. The effector matrix was collapsed into effector families, as the different alleles likely perform similar biological functions *in planta* (Cunnac *et al.*, 2011). Putative pseudogenes were considered absent, as they may be nonfunctional. The BAYESTRAITS methodology followed an approach as in Press *et al.* (2013) and is described in detail in Methods S1.

Horizontal gene transfer analysis

For each effector family, best-hit nucleotide sequences were aligned using CLUSTALW (Larkin *et al.*, 2007). RAxML was used to build a phylogenetic tree with a GTR model of evolution and 1000 bootstrap replicates. Incongruence with the core-genome tree was examined visually. To further assess horizontal transfer, a species–gene tree reconciliation method RANGER-DTL v.2 (Bansal *et al.*, 2012) was used, as in Bruns *et al.* (2018). Full details are described in Methods S1.

Identification of genomic islands

Genomic islands (GIs) were identified in the PacBio-sequenced cherry pathogenic strains using ISLANDVIEWER3 (Dhillon *et al.*, 2015). Islands were manually delimited as in McCann *et al.* (2013). BLASTN was utilized to determine if these GIs were present in other *P. syringae* genomes. As most GIs exceeded 10 kb and most genome assemblies were highly fragmented, the islands were split into 0.5 kb sections before analysis to prevent false negatives due to contig breaks. An island was concluded as likely to be present if all sections produced hits with a query length > 30%. To validate this approach, the Illumina-sequenced genome assemblies of the PacBio-sequenced strains were searched for their own islands.

General DNA manipulations and bacterial transformations

Cloning and other molecular biology techniques, including ectopic expression of potential *avr* genes, were as described in earlier studies (Staskawicz *et al.*, 1984; Arnold *et al.*, 2001; Kvitko & Collmer, 2011). Details are provided in Methods S1.

Results

Genome assembly and sequencing statistics

Eighteen *P. syringae* strains isolated from cherry and plum were phenotyped for pathogenicity and genome sequenced in a previous study (Hulin *et al.*, 2018). To increase this sample, nine strains isolated from wild cherry and five additional non-*Prunus* out-groups were genome sequenced using the Illumina MiSeq. The genomes of eight cherry strains sequenced elsewhere (Baltrus *et al.*, 2011; Nowell *et al.*, 2016) were also downloaded from NCBI.

Information on the origin and pathogenicity of each strain is summarized in Table 1. Twenty-eight were considered pathogenic to cherry, including all *Pss* isolated from cherry and plum. By contrast, three *Psm* race 1 strains from plum (R1-5300, R1-9326 and R1-9629) and one from cherry, strain R1-9657, failed to induce canker on cherry following tree inoculations; and three strains of unknown taxonomy isolated from plum and cherry (Ps-9643, Ps-7928C and Ps-7969) were also non-pathogenic (references in Table 1). The cherry pathogens are referred to as their described pathovar names throughout this study. To simplify figures, cherry pathogens are highlighted and the first few letters of the pathovar name were used. '*Pss*' becomes '*syr*', as otherwise *Pss* could refer to other pathovars beginning with 's' (e.g. *savastanoi*).

All strains included in this study were sequenced using Illumina MiSeq. Three representative strains (R1-5244, R2-leaf and syr9097) were sequenced using PacBio and the nonpathogenic *Psm* R1 strain R1-5300 was sequenced using Oxford Nanopore, to obtain more complete genomes. Table 2 summarizes the genome assembly statistics for all strains sequenced in this study and Hulin *et al.* (2018). Illumina genomes assembled into 23–352 contigs, whilst the long-read sequenced genomes assembled into one to six contigs. The number of plasmids in each strain was determined by plasmid profiling (Fig. S1). *Psm* R1 and R2 strains possessed between two and six plasmids, *P. syringae* pv *avii* 5271 possessed six plasmids, whereas, apart from three strains (syr5275, syr7928A, syr9644) with one plasmid each, most cherry-pathogenic strains of *Pss* did not possess plasmids. The strain syr9097, which was sequenced using PacBio, lacked plasmids. The genomes sequenced with long-read technology all assembled into the correct number of contigs corresponding to chromosome and plasmids, apart from R1-5300. The chromosome of this strain was separated into two contigs (tig0 and tig75), based on a whole-genome alignment with *Psm* R1 strain R1-5244 (Fig. S2).

The *Psm* R1 (R1-5244, R1-5300) and *Psm* R2 (R2-leaf) long-read assemblies revealed putatively complete plasmid contigs

Table 2 Assembly statistics for all strains sequenced in this study and Hulin *et al.* (2018)

Assembly	Sequencing	Reference	No. of contigs	Plasmids	Total length	GC%	N50	Coverage	Features
R1-5270	**Illumina**	**This work**	**185**	**3**	**6258 313**	**58.10**	**202 152**	**134**	**5770**
R1-9326	Illumina	Hulin *et al.* (2018)	268	4	6353 636	57.91	142 021	81	5874
R1-9629	Illumina	Hulin *et al.* (2018)	216	3	6341 664	57.94	142 021	172	5856
R1-9646	**Illumina**	**Hulin *et al.* (2018)**	**171**	**3**	**6302 776**	**58.03**	**235 429**	**180**	**5801**
R1-9657	Illumina	Hulin *et al.* (2018)	191	4	6317 852	57.91	145 272	158	5848
R2-5255	**Illumina**	**Hulin *et al.* (2018)**	**206**	**2**	**6448 834**	**58.38**	**102 760**	**112**	**5966**
R2-7968A	**Illumina**	**This work**	**278**	**6**	**6498 711**	**58.42**	**91 262**	**134**	**6016**
R2-5260	**Illumina**	**Hulin *et al.* (2018)**	**223**	**3**	**6495 620**	**58.41**	**101 794**	**458**	**5995**
R2-9095	**Illumina**	**This work**	**304**	**2**	**6418 849**	**58.48**	**92 453**	**100**	**5887**
R2-SC214	**Illumina**	**Hulin *et al.* (2018)**	**203**	**3**	**6253 818**	**58.56**	**108 341**	**180**	**5747**
avii5271	**Illumina**	**This work**	**352**	**6**	**6243 644**	**58.56**	**56 064**	**127**	**5809**
Ps-9643	Illumina	Hulin *et al.* (2018)	58	1	5937 102	58.78	243 355	212	5386
syr9293	**Illumina**	**Hulin *et al.* (2018)**	**73**	**0**	**6135 031**	**58.84**	**557 853**	**196**	**5302**
syr9630	**Illumina**	**Hulin *et al.* (2018)**	**57**	**0**	**5940 819**	**59.33**	**347 701**	**206**	**5175**
syr9644	**Illumina**	**Hulin *et al.* (2018)**	**75**	**1**	**6173 193**	**59.13**	**251 053**	**208**	**5334**
syr9654	**Illumina**	**Hulin *et al.* (2018)**	**49**	**0**	**5941 610**	**59.37**	**245 023**	**147**	**5148**
syr9656	**Illumina**	**Hulin *et al.* (2018)**	**39**	**0**	**5980 728**	**59.10**	**1007 808**	**205**	**5184**
syr9659	**Illumina**	**Hulin *et al.* (2018)**	**51**	**0**	**5943 090**	**59.37**	**235 830**	**116**	**5148**
syr5264	**Illumina**	**This work**	**59**	**0**	**6029 896**	**59.08**	**380 149**	**114**	**5314**
syr5275	**Illumina**	**This work**	**64**	**1**	**5994 091**	**59.30**	**371 492**	**145**	**5207**
syr7928A	**Illumina**	**This work**	**59**	**1**	**6129 363**	**59.26**	**371 492**	**141**	**5338**
syr8094A	**Illumina**	**This work**	**71**	**0**	**5942 438**	**59.33**	**265 238**	**106**	**5184**
syr7928C	Illumina	This work	49	1	5994 455	59.17	325 175	124	5318
syr7969	Illumina	This work	92	0	6185 932	59.01	164 374	151	5476
RMA1	Illumina	Hulin *et al.* (2018)	95	1	6306 889	58.73	187 448	320	5825
syr100	Illumina	This work	23	0	5872 916	59.33	893 822	83	5140
syr2675	Illumina	This work	65	0	5994 384	59.34	227 612	83	5177
syr2676	Illumina	This work	90	1	6158 476	59.30	259 660	78	5387
syr2682	Illumina	This work	185	1	6259 099	59.21	242 212	84	5405
syr3023	Illumina	This work	228	0	6203 212	58.90	456 738	88	5365
R1-5244	**PacBio**	**This work**	**5**	**4**	**6445 963**	**58.05**	**6109 228**	**82**	**6024**
R2-leaf	**PacBio**	**This work**	**6**	**5**	**6576 340**	**58.41**	**6242 845**	**141**	**6093**
syr9097	**PacBio**	**This work**	**1**	**0**	**5929 959**	**59.30**	**5929 959**	**100**	**5147**
R1-5300	MinION	**This work**	6	4	645 601	57.87	5688 034	100	6449

Cherry pathogens are in bold. Long-read sequenced genomes are highlighted with shading. N50, the weighted median contig size in the assembly; Features, the number of protein encoding and RNA sequences in the RAST annotated genome.

containing plasmid-associated genes (Tables S5–S7). All three strains (R1-5244, R1-5300 and R2-leaf) possessed plasmids with *repA* homologues, indicating they may belong to common plasmid family pPT23A (Zhao *et al.*, 2005). Several plasmids also contained T4SS conjugational machinery (VirB/D), so may be conjugative.

Core-genome phylogenetic analysis

To examine the relatedness of strains, an analysis of core genes was carried out using 108 genomes of strains from the well-studied phylogroups 1–3 isolated from both plants and aquatic environments. A maximum likelihood phylogeny based on 1035 concatenated core genes was constructed (Fig. S3). There was low support for certain P2 and P3 clades based on bootstrap analysis. To determine if particular taxa were causing low support, the analysis was systematically repeated for the two phylogroups, with non-cherry strains removed. Support and tree likelihood values were compared (Table S8). Within P3, the removal of *P. syringae* pv *eriobotryae* or *P. syringae* pv *daphniphylli* improved support, whilst the removal of *P. syringae* pv *syringae* 1212 improved support values in P2 (Figs S4, S5). The global analysis was then repeated with these taxa removed (Figs S6–S9). The final phylogeny (Fig. 1), with the highest mean branch support (92.8%), lacked *P. syringae* pv *eriobotryae*. The phylogeny, built using a 611 888 bp alignment, contained 102 taxa due to the removal of strains found to be identical to others (dendro4219, syr9630, R1-9629, R1-9326 and R1-5269). Most support values exceeded 70%, with good support for branches leading to cherry-pathogenic clades.

One explanation for the low support within P2 and P3 was that these clades have undergone core-genome recombination. The program GENECONV (Sawyer, 1989) showed that 140 genes had putatively recombined (127 288 bp total length, 20.8% of the alignment). Table S9 lists the number of recombination events per phylogroup. The most frequent core gene recombination occurred in P3 (73 genes affected), followed by 31 genes in P2 and only 13 in P1.

Cherry pathogens were found in all three phylogroups. The two *Psm* races (R1 in P3, R2 in P1) and *P. syringae* pv *avii* (P1)

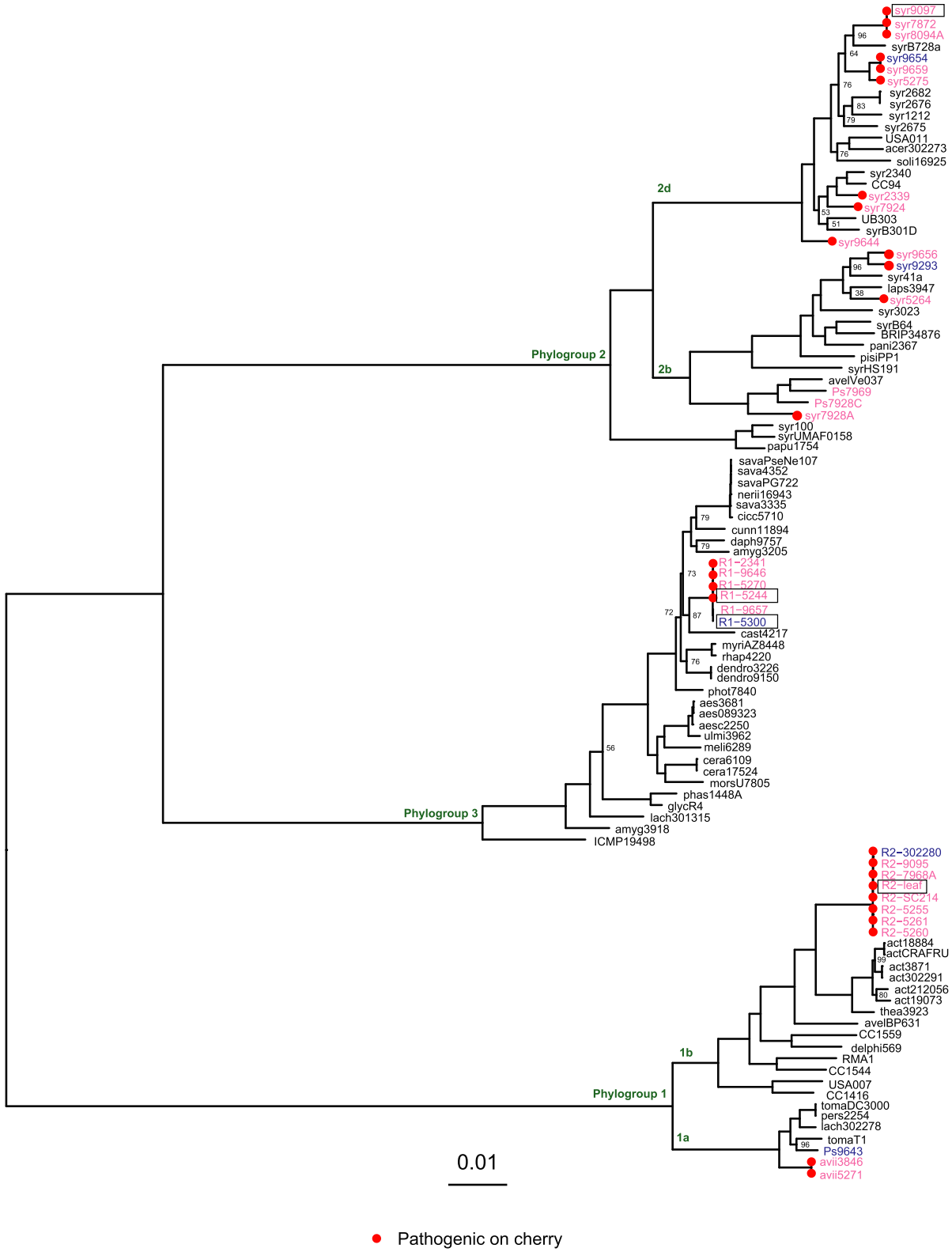

Fig. 1 Core-genome phylogenetic tree. Multi-locus phylogeny based on 1035 genes which represent the core genome of *Pseudomonas syringae*. Strains from cherry and plum are highlighted in pink and blue respectively. Strains pathogenic to cherry (assessed in Hulin *et al.*, 2018; Vicente *et al.*, 2004) are labelled with red circles. Strains with long-read sequenced genomes are in black boxes. Phylogroups are also labelled for reference. Percentage bootstrap support values below 99% are shown for each node. The bar is nucleotide substitutions per site.

formed monophyletic clades. Within *Psm* R1, strains pathogenic to cherry formed a clade distinct from previously classified non-pathogenic strains (Hulin *et al.*, 2018), indicating that there has been divergence in their core genomes. By contrast, *Prunus*-infecting strains of *Pss* were found across P2, interspersed with strains isolated from other plants and aquatic environments. To ensure that genomic comparisons between P2 strains were based on differential pathogenicity, several related non-*Prunus* strains were pathogenicity tested on cherry leaves. *In planta* bacterial populations of non-*Prunus* strains were reduced compared with *Prunus Pss* strains (Fig. S10; Table S10).

Search for virulence factors

The *hrp* pathogenicity island All sequenced strains contained the *hrp* pathogenicity island required for conventional Type III secretion. Core effector genes from the adjacent conserved effector locus (Alfano *et al.*, 2000), such as *avrE1*, *hopM1* and *hopAA1*, were present (Fig. S11), However, *hopAA1* was truncated in both *Psm* R1 and R2 due to inversion events. The *hopAA1* gene was truncated in *Psm* R2, whilst in *Psm* R1 both *hopAA1* and *hopM1* were truncated (Fig. S12).

Type III effectors and other virulence genes Genomes were then scanned for known virulence genes and a heatmap of presence, absence and pseudogenization was constructed (Fig. 2). In terms of T3Es, there was variation both between and within the different cherry-pathogenic clades. Notably, *Psm* R1, which contained strains pathogenic and nonpathogenic on cherry, showed clear differentiation in effector repertoires (Table S11). *Psm* R1, R2 and *P. syringae* pv *avii* possessed 24–34 effector genes, whereas *Pss* strains possessed nine to 15. The reduced effector repertoire of *Pss* was representative of P2 strains as previously noted (Baltrus *et al.*, 2011; Dudnik & Dudler, 2014). Table 3 lists the effectors in each long-read genome assembly in order of appearance.

Non-T3 virulence factors were identified. All pathogenic *Psm* R1 strains possessed the coronatine biosynthesis clusters, which were plasmid borne in *Psm* R1-5244. All cherry-pathogenic *Pss* strains possessed at least one biosynthesis gene cluster for the toxins syringomycin, syringolin and syringopeptin, with several strains possessing all three. Strains within clade P2b possessed the biosynthesis genes for mangotoxin. The nonpathogenic cherry P2b strains Ps7928C and Ps7969 lacked all toxin biosynthesis clusters.

A cluster of genes named WHOP (woody hosts and *Pseudomonas*) thought to be involved in aromatic compound (lignin) degradation (Caballo-Ponce *et al.*, 2016) was present in *Psm* R1 and R2, whilst *P. syringae* pv *avii* and most *Pss* strains contained no WHOP homologues. Two cherry P2d strains (syr2339 and syr7924) did, however, possess the catechol *catBCA* cluster. The genomes were also searched for the ice nucleation gene cluster. Members of *Psm* R1, *Pss* and *P. syringae* pv *avii* strains all possessed genes involved in ice nucleation (Fig. 2), whilst *Psm* R2 lacked the complete set of ice nucleation genes.

Associating type III effector evolution with host specificity

T3E evolution was statistically correlated with cherry pathogenicity, using BayesTraits and Gloome (Pagel, 2004; Cohen *et al.*, 2010). BayesTraits takes a binary matrix of two traits within a phylogeny and determines if changes in the two characters (effector gene and pathogenicity) have evolved independently or dependently. Fig. 3(a) shows the likelihood ratio of cherry pathogenicity being correlated with each effector family's evolution, with significantly associated effectors highlighted.

BayesTraits analysis using the core-genome phylogeny predicted the evolution of six T3E families was linked to cherry pathogenicity. These were *hopBF*, *hopAB*, *hopH*, *hopAR*, *avrPto* and *hopBB*. To account for any phylogenetic uncertainty, the program was also run 100 times on the full set of 100 bootstrapped trees generated by RAxML. The evolution of T3Es *hopBF*, *hopAR* and *hopAB* was always associated with pathogenicity for all 100 trees in all runs, indicating strong association. However, the T3E genes *avrPto*, *hopBB* and *hopH* were only significantly correlated for 88%, 77% and 62% of trees respectively, averaged across the different runs (Fig. S13). To determine how these genes had been gained or lost across the phylogeny, the program Gloome was used (Cohen *et al.*, 2010). Fig. 3(b) illustrates the predicted gain and loss of these T3Es on the branches leading to clades pathogenic to cherry. Those putatively associated with pathogenicity (high probability of gain in cherry-pathogenic clades) included *hopAR1*, *hopBB1*, *hopBF1* and *hopH1*. The T3Es *hopAB1* and *avrPto1* were found to be lost from cherry pathogenic *Psm* R1, whilst the *hopAB1* and *hopAB3* alleles were pseudogenized in *Psm* R2 and *P. syringae* pv *avii* (Fig. 3b). All effector gain and loss events are presented in Fig. S14 and Table S12. Fig. S15 shows the phylogeny with branch labels used in Gloome.

Gloome predicted that key effectors have been gained in multiple clades. The *hopAR1* gene has been gained in *Psm* R1, *Psm* R2, *Pss* and *P. syringae* pv *avii*. The T3E *hopBB1* was present in the majority of strains within *Psm* R1, R2 and *P. syringae* pv *avii* but was absent from *Pss* strains. It showed high probability of gain on branches leading to both *Psm* R2 and *P. syringae* pv *avii*. However, Gloome predicted loss in two *Psm* R1 strains, indicating that the gene may have experienced dynamic evolution in cherry pathogens. The *hopBB1* effector is closely related to members of the *hopF* family and *avrRpm2* (Lo *et al.*, 2016). In addition to the significant acquisition of *hopBB1* homologues, the *hopF* family was expanded in cherry pathogens. Pathogenic strains in *Psm* R1 and R2 all possessed two *hopF* alleles each (*hopF3* and *hopF4*, and *hopF2* and *hopF4*; see Fig. 2). *P. syringae* pv *avii* did not possess any *hopF* homologues, but had gained *hopBB1*. By contrast, *Pss* strains lacked all *hopF* members.

Origins of key effectors in cherry pathogens

To understand the origins of key effectors, gene phylogenies were produced. Incongruence with the core-genome phylogeny indicated that effector sequences had likely experienced horizontal

New Phytologist

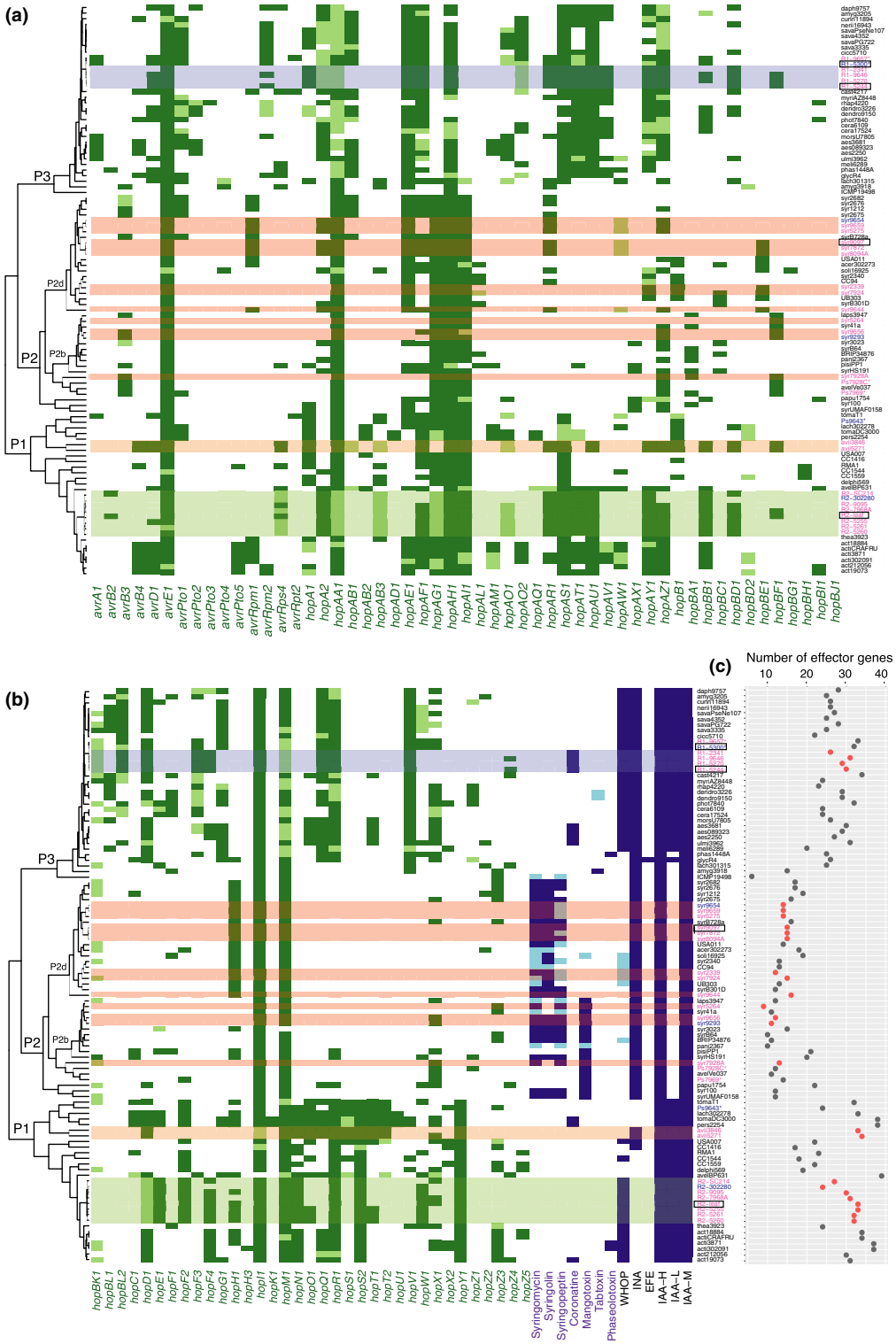

Fig. 2 Virulence gene identification. (a) Heatmap of virulence gene presence and absence across *Pseudomonas syringae* (*avrA1*-*hopBJ1*). The dark green squares indicate presence of a full-length type III effector (T3E) homologue, whereas light green squares indicate that the gene is disrupted or truncated in some way. Strains isolated from cherry and plum are highlighted in pink and blue respectively. Asterisks indicate nonpathogenic to cherry in controlled pathogenicity tests. Strains with long-read sequenced genomes are in black boxes. The cherry-pathogenic clades are illustrated via horizontal shading of cells with *P. syringae* pv *morsprunorum* (*Psm*) R1 in blue, *Psm* R2 in light green, *P. syringae* pv *syringae* (*Pss*) in pink and *P. syringae* pv *avii* in orange. Strains are ordered based on the core-genome phylogenetic tree, which is represented by the dendrogram, with phylogroups labelled (P1–P3). (b) Continuation of (a) for T3Es *hopBK1*–*hopZ5* and additional non-T3 virulence factors, which are coloured in dark blue (all genes full-length and present) and light blue (not all genes present/truncation of genes). (c) The total number of full-length and pseudogenized T3E genes plotted for each strain, with cherry pathogenic strains in red and other strains in grey.

Table 3 Order of effectors in genomes sequenced using PacBio/MinION methods

	Contig	Length	Effector
R1-5244			
Chromosome	tig0	6109 228	*hopAZ1*, *hopA2**, *avrE1*, *hopM1**, *hopAA1**[†], *hopZ4*, *hopAT1*, *hopQ1*, *hopD1*, *hopR1*, *hopF4*, *hopBL2*, *hopAV1*, *hopAO2**, *hopAY1*, *hopF3*, *hopAS1*, *hopI1*, *hopAE1*, *hopAF1-2*, *hopAU1*, *hopAH1*, *hopV1*, *hopAR1*, *hopBK1**
Plasmid	tig3	168 854	*hopAF1-1*, *hopBF1*, *avrD1*, *avrRpm2*, *hopBD1*
Plasmid	tig4	81 536	*hopA1*
Plasmid	tig5	45 535	–
Plasmid	tig6	40 810	–
R1-5300			
Chromosome	tig0	5688 034	*hopV1*, *hopAZ1*, *avrA1*, *hopQ1-2*, *hopA2**, *avrE1*, *hopM1**, *hopAA1*[†], *hopAB1*, *hopQ1*, *hopD1*, *hopR1*, *hopAO2**, *avrRpm2*, *avrPto1*, *hopAS1*, *hopAT1**, *hopBL2**, *hopI1*, *hopAE1*, *hopAF1-2*, *hopF3*, *hopAY1**, *hopAU1*, *hopAH1*
Chromosome	tig75	697 453	*hopW1*, *hopBK1**, *hopAR1*
Plasmid	tig46	52 059	–
Plasmid	tig65	47 809	*hopX1*, *hopBB1*, *hopG1*
Plasmid	tig84	57 689	*hopAO1**
Plasmid	tig113	102 557	*avrD1*
R2-leaf			
Chromosome	tig0	6242 845	*hopY1*, *hopAS1*, *hopAT1*, *hopH1*, *hopF4*, *hopW1*, *hopR1*, *hopAG1**, *hopAH1-2*, *hopAI1*, *hopN1*, *hopAA1**, *hopM1*, *avrE1*[†], *hopF2*, *hopE1*, *hopA2*, *hopAH1-1*, *hopAH1-1*, *hopAB3**, *avrRps4*, *hopS2*, *hopI1*, *hopAR1*
Plasmid	tig5	102 862	*hopAO1**, *hopAZ1*, *hopAY1*
Plasmid	tig4	97 840	*hopD1**, *hopAU1*
Plasmid	tig6	69 519	*hopAF1-1*, *hopBF1*
Plasmid	tig8	42 783	*hopBB1*, *hopBD1*
Plasmid	tig9	20 491	*avrB2*, *hopX1*
syr9097			
Chromosome	tig0	5929 959	*hopAG1*, *hopAH1*, *hopA1I*, *avrRpm1*, *hopAR1*, *hopI1*, *hopAE1*, *hopBE1*, *hopAF1*, *hopAH1*, *hopAW1**,*hopH1*, *hopA2*, *avrE1*, *hopM1*, *hopAA1*[†]

Effectors are listed in order of appearance on each assembly contig (labelled as chromosomal or plasmid). Where effectors could be considered as linked (within 10 kb of each other) they are underlined.

*Effector gene is disrupted and is labelled as a pseudogene.

[†]Effectors within the conserved effector locus.

gene transfer (HGT) between the pathogenic clades, as their sequences clustered together. There has been possible effector exchange between *Psm* R1, R2 and *P. syringae* pv *avii*. To predict precisely where transfers had occurred on the phylogeny, the program RANGER-DTL was utilized (Bansal *et al.*, 2012). Table 4 reports T3Es that exhibited evidence of HGT between cherry pathogens (gene trees are presented in Figs S16, S17). Full transfer events are listed in Table S13, and Fig. S18 shows the phylogeny with branch labels used in RANGER-DTL. The BAYESTRAITS pathogenicity-correlated T3Es, *hopBB* and *hopBF*, both showed evidence of HGT. Fig. 4(a) shows examples of T3Es putatively undergoing HGT between cherry pathogenic clades highlighted in red. Alignments of the flanking regions (Fig. 4b) showed homology between the cherry pathogens and included mobile elements likely involved in recombination events. Putatively transferred effectors were mostly plasmid encoded in the long-read genomes (Table 3). In R1-5244, several of these genes were encoded on one plasmid (Contig 3), whilst in R2-leaf they were found on two plasmids (Contig 6 and 8).

The pathogenicity-associated T3E gene *hopAR1* was present in 23/28 cherry pathogens and showed probable gain in pathogenic clades. Phylogenetic analysis of this T3E (Fig. 5a) showed that the sequences for the different cherry pathogenic clades did not cluster with each other, indicating convergent acquisition. Prophage identification (Table S14) did, however, reveal that this T3E is predicted to have been gained in *Psm* R1 and R2 within different phage sequences, whilst in *Pss* it is on a genomic island (Fig. 5b), and so has been acquired via distinct mechanisms. The *Psm* R1 phage is 51.5 kb, described as intact, and contains both *hopAR1* and a truncated version of *hopBK1*. The *Psm* R2 phage sequence was 37.1 kb and was described as 'incomplete', indicating it did not have all the components of an active prophage. Further analysis of this region in *Psm* R2 and P2 strains revealed a shared adjacent tRNA-Thr gene (Fig. S19a,b). Within P2, although cherry *Pss* strains lacked the phage, several strains isolated from bean (syr2675, syr2676 and syr2682) possessed the *hopAR1* gene within a phage homologous to that in *Psm* R2. The syr2675 *hopAR1* sequence was also the most closely related homologue of *Psm* R2 *hopAR1* (Fig. 5a). This evidence suggests that this effector gene may have been transferred via phage between phylogroups.

Many T3Es are mobilized between bacteria on GIs. GIs were identified for the three PacBio-sequenced strains of *Psm* R1, *Psm* R2 and *Pss* (Tables S15–S17). R1-5244 GIs contained the coronatine biosynthesis cluster and six T3Es. In R2-leaf, eight T3E genes were located on GIs, whilst in syr9097 three T3Es were

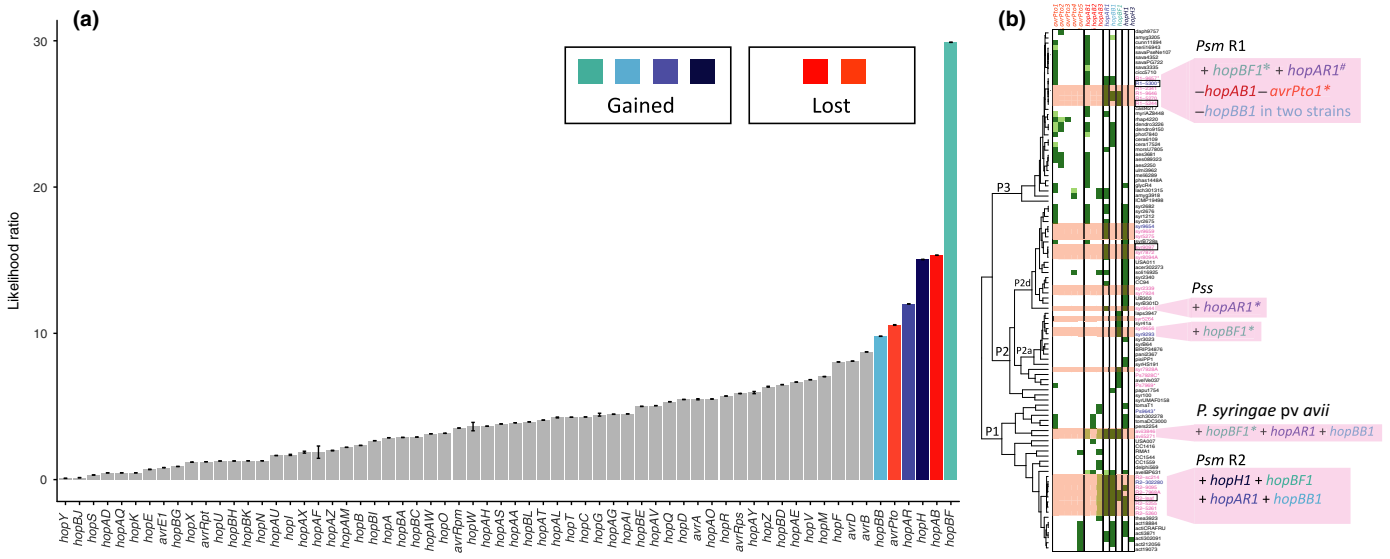

Fig. 3 Association of type III effector evolution with cherry pathogenicity. (a) Barplot showing the likelihood ratio for the correlation of each effector gene family with cherry pathogenicity based on BayesTraits analysis using the core-genome phylogeny. The values are obtained from means of 100 independent runs of the program with error bars showing ± SE above and below the mean. Those effectors that were not significantly associated with pathogenicity are coloured in grey. Coloured bars were associated with pathogenicity ($P \leq 0.05$ in >90% of runs). Whilst those genes that were hypothesized to be gained in cherry pathogens (from gain loss mapping engine (Gloome) analysis) are coloured in shades of blue, whilst where the significant gene was absent in cherry pathogenic clades the bar is coloured in shades of red. (b) Gain and loss of BayesTraits-associated T3Es in cherry-pathogenic clades on the core-genome phylogeny predicted using Gloome ($P \geq 0.8$). The phylogeny and heatmap of these effector genes is presented (heatmap as in Fig. 2, effector gene names are colour coded based on the bar colours in (a) and cherry-pathogenic strains are highlighted by pink horizontal shading of columns). Phylogroups (P1–P3) are labelled. Strains with long-read sequenced genomes are in black boxes. *The probability of this effector being gained/lost was slightly < 0.8, but exceeded 0.65 (see Supporting Information Table S12 for details). #The *hopAR1* gene has been gained on the branch leading to *P. syringae* pv *morsprunorum* (*Psm*) R1 (including pathogens and nonpathogens).

found on genomic islands. These GIs were then searched for in other *P. syringae* genomes to identify potential sources of transfer, and Fig. S20 shows heatmaps of GI presence. The *Psm* R1 GIs included several found only in pathogenic *Psm* R1 strains, differentiating them from the nonpathogens. These included the coronatine biosynthesis cluster (GI1), *hopF3* (GI6) and *hopAT1* (GI14). Most *Psm* R1 GIs produced hits across *P. syringae*, particularly in P1 and P3. *Psm* R2 GIs were most commonly found in P1. Several were shared with other cherry-pathogenic clades, including those containing *hopAF1* (GI36), *hopAT1* (GI3) and *hopD1* (GI6). Finally, although most islands identified in syr9097 were commonly found across the species complex, those containing T3Es (GI30, GI23 and GI26) appeared to be P2 specific, indicating that cherry-pathogenic strains likely gained these islands from other members of P2.

Functional analysis of potential *avr* genes

To validate predictions from genome analysis, cloning was used to identify avirulence factors active in cherry. The effector genes *avrPto* and *hopAB* were absent from cherry pathogens, and their evolution was theoretically linked to pathogenicity. Several other candidate avirulence effectors were identified that were absent from cherry pathogens but present in close out-groups (Fig. 6). Avirulence-gene identification focused on *Psm* R1, as any T3E variation within this clade may be due to differences in host specificity rather than phylogenetic distance. Potential avirulence T3E genes included *avrA1*, *avrPto1*, *hopAA1*, *hopAB1*, *hopAO2* and *hopG1*, which had full-length homologues in nonpathogenic *Psm* R1 strains but were absent from or truncated in pathogens. These genes were cloned from the strain R1-5300 (except *hopAO2*, which was cloned from R1-9657).

The effector *avrRps4* was also cloned from *P. syringae* pv *avellanae* (*Psv*) BPIC631, a close relative of *Psm* R2. This effector was absent from most cherry-pathogenic strains (Fig. 6). Several pathogens possessed the full-length gene (R2-leaf, R2-9095 and *P. syringae* pv *avii*), but lacked the KRVY domain that functions *in planta* (Fig. S21) (Sohn *et al.*, 2009). The *hopAW1* gene was cloned from *Pph*1448A as this T3E has undergone two independent mutations in *Pss* strains, disrupting the beginning of the gene (Fig. S22). Finally, *hopC1* was cloned from the *Aquilegia vulgaris* pathogen RMA1, which is basal to the *Psm* R2 clade as it is absent from all cherry-pathogenic strains.

Nine effectors were cloned into pBBR1MCS-5 and conjugated into three pathogenic strains (R1-5244, R2-leaf and syr9644). The presence of the plasmids did not affect multiplication *in vitro*. Knock-out strains for the T3SS gene *hrpA* were obtained for R1-5244 and R2-leaf to act as nonpathogenic controls that could not secrete T3Es and failed to cause the HR on tobacco (Fig. S23).

Bacterial multiplication experiments were conducted in cherry leaves. The transconjugants expressing HopAB1 or HopC1 failed to multiply to the same levels as the pathogenic empty vector (EV) controls or produce disease lesions. The ectopic expression

Table 4 List of putative horizontal gene transfer events that have occurred between cherry-infecting clades within *Pseudomonas syringae*

Effector	Putative transfers	Region	Plasmid location	Predicted in RANGER-DTL
avrD1	R1/R2/*P. syringae* pv *avii*	Plasmid	tig3 (R1-5244)	Y
*avrRps4**	R2/*P. syringae* pv *avii*	Next to cluster of mobile elements	–	Y
hopAF1	R1/R2/*P. syringae* pv *avii*	Plasmid	tig3 (R1-5244), tig6 (R2)	Y
*hopAO1**	R1/R2/*P. syringae* pv *avii*	Plasmid	tig5(R2), tig84 (R1-5300)	N
hopAT1	R1/R2	Genomic island	–	N
hopAU1	R2/*P. syringae* pv *avii*	Plasmid	tig4 (R2)	Y
hopAY1	R2/*P. syringae* pv *avii*	Plasmid	tig5 (R2)	Y
hopBB1	R1/R2/*P. syringae* pv *avii*	Plasmid	tig8 (R2), tig65 (R1-5300)	Y
hopBD1	R2/*P. syringae* pv *avii*	Plasmid	tig3 (R1-5244), tig8 (R2)	Y
hopBF1	R1/R2/*P. syringae* pv *avii*	Plasmid	tig3 (R1-5244), tig6 (R2)	Y
*hopD1**	R2/*P. syringae* pv *avii*	Plasmid	tig4(R2)	N
hopO1	R2/*P. syringae* pv *avii*	Next to cluster of mobile elements (next to *hopT1*)	–	Y
hopT1	R2/*P. syringae* pv *avii*	Next to cluster of mobile elements (next to *hopO1*)	–	Y

Where the effector gene is present in the PacBio- or MinION-sequenced strains, its chromosomal or plasmid location is indicated. The type III effector genes *hopO1* and *hopT1* were not present in the PacBio-sequenced strains and therefore it is uncertain if they are on plasmids or chromosomal.
*Effector gene is disrupted in some strains and is labelled as a pseudogene.

of AvrA1, AvrRps4 and HopAW1 also caused significant reductions in growth, but this reduction was not consistently seen across all three pathogenic strains (Fig. 7a). As the addition of the *hopAB1* gene reduced pathogenicity, full-length *hopAB2* and *hopAB3* genes were also cloned from *Psv*BPIC631 and RMA1, and were also found to reduce pathogen multiplication (Fig. 7b).

To investigate the induction of the HR by the HopAB family and HopC1, inoculations were performed at high concentrations (2×10^8 colony-forming units (CFU) ml^{-1}) as in Hulin *et al.* (2018). In *Psm* R1 and R2, the addition of these T3Es led to more rapid tissue collapse than observed in EV controls, indicative of HR induction (Fig. 7c,d); HopC1 and HopAB1 were particularly effective. With *Pss*, however, EV transconjugants themselves caused rapid tissue collapse, making it impossible to recognize an induced HR as symptom development was not significantly different.

The *hopAB1* gene is found in a mobile-element-rich *c.* 40 kb region in the nonpathogenic *Psm* R1-5300, missing from the pathogen *Psm* R1-5244 (Fig. 8a). Meanwhile, *Psm* R2 and *P. syringae* pv *avii* possessed putatively pseudogenized *hopAB3* alleles (Fig. 8b), and *P. syringae* pv *avii* also possessed a truncated *hopAB1* gene (Fig. S24). *hopAB3* is truncated in *Psm* R2 due to a 2 bp insertion (GG at position 1404 bp) leading to a premature stop codon, whilst in *P. syringae* pv *avii* a 218 bp deletion has disrupted the C-terminus. If expressed, the E3-ubiquitin ligase domain is completely absent from the *Psm* R2 protein and disrupted in *P. syringae* pv *avii* (Fig. 8c). Both HopAB3 alleles were also divergent enough that the Pto-interacting domain (PID) was not identified by Interproscan. To determine if the truncated *Psm* R2 HopAB3 allele induced any resistance response in cherry leaves, the gene was expressed in *Psm* R1-5244 and population growth measured. The addition of this gene did not lead to a significant reduction in growth compared with the EV control, unlike other *hopAB* alleles (Fig. 8d), and the transconjugant was still able to induce disease symptoms 10 d post inoculation (dpi) (Fig. 8e).

Overall, the data supported the conclusion that expressing alleles of *hopAB* and *hopC* reduced bacterial multiplication in cherry and were consistent with HR induction by *Psm* R1 and R2. However, it should be noted that any growth changes exhibited might have been influenced by aberrant transcription or translation of these effectors in the plant due to expression *in trans*.

Discussion

Core-genome phylogenetics

Phylogenetic analysis confirmed that cherry pathogenicity has evolved multiple times within *P. syringae*. *Psm* R1, R2 and *P. syringae* pv *avii* each formed distinct monophyletic clades, whereas cherry-pathogenic *Pss* strains were distributed across the P2 clade, indicating that cherry pathogenicity has either evolved multiple times within P2 or that this clade is not particularly specialized. To confirm this genomic prediction of pathogenicity, several additional P2 strains isolated from bean, pea and lilac were tested for pathogenicity in cherry. They each produced lower population levels in cherry leaves than cherry pathogens, suggesting that strains isolated from cherry and plum are more pathogenic to their hosts of origin (Fig. S10). Many P2 strains have previously been named *Pss* on the basis of lilac pathogenicity, despite being pathogenic to other plant species (Young, 1991). A new naming system within this phylogroup is desirable.

Search for candidate effectors involved in cherry pathogenicity

Gains and losses of T3Es were closely associated with pathogenicity. Virulence-associated effectors *hopAR1*, *hopBB1*, *hopH1* and *hopBF1* had been gained in multiple cherry-pathogenic clades. The *hopAR1* effector has been studied in the bean pathogen *P. syringae* pv *phaseolicola* R3 (1302A), as a GI-located *avr* gene (*avrPphB*) whose protein is detected by the corresponding R3 resistance protein *in planta* (Pitman *et al.*, 2005; Neale *et al.*, 2016). HopAR1 also acts as a virulence factor as a cysteine protease which targets receptor-like kinases to interfere with plant

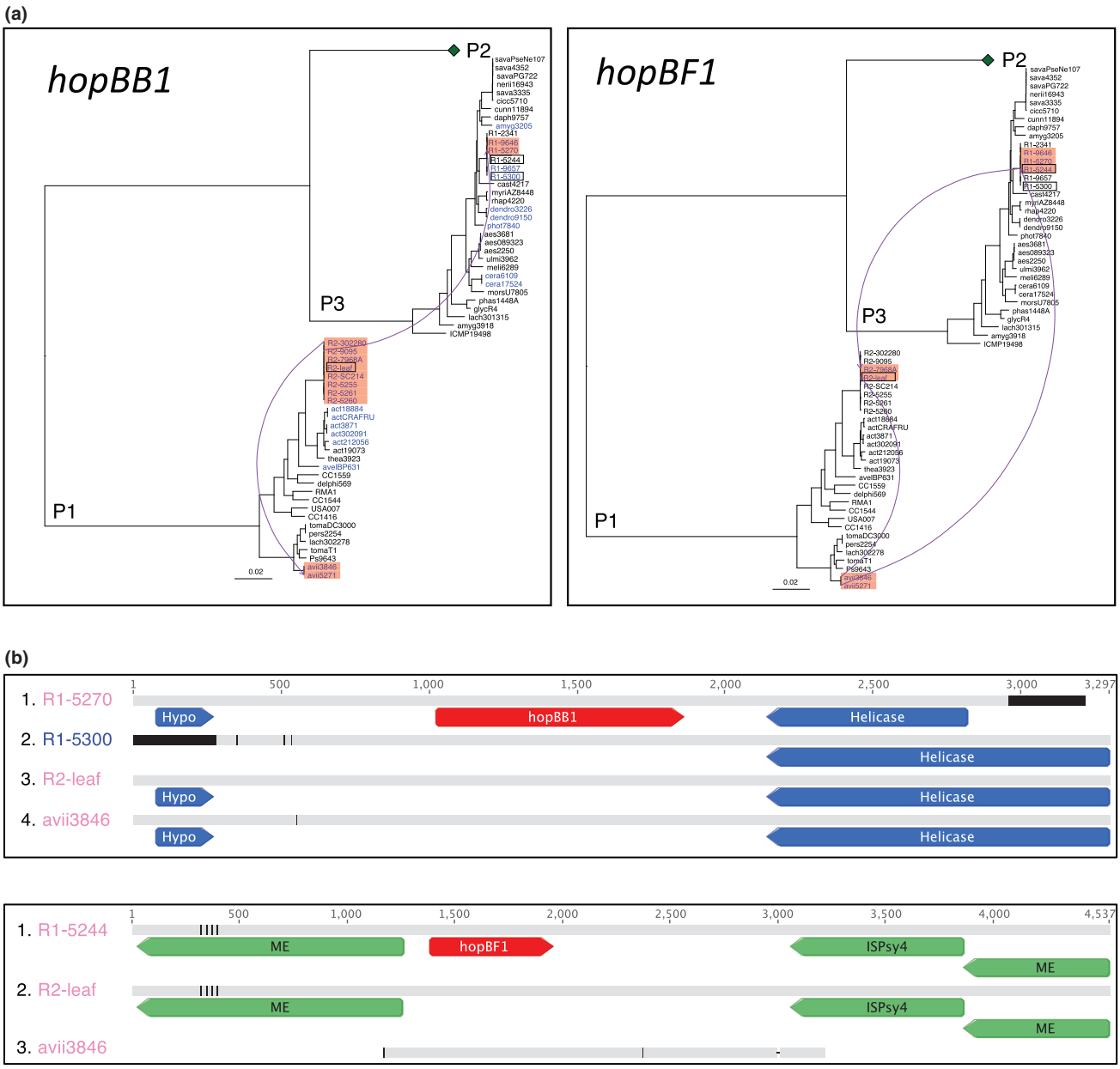

Fig. 4 Horizontal gene transfer has played a key role in the evolution of cherry pathogenicity. (a) Putative horizontal transfer of the *hopBB1* and *hopBF1* genes between different cherry-pathogenic clades based on the *Pseudomonas syringae* core-genome phylogeny. The different phylogroups are labelled (P1–P3), with P2 collapsed to concentrate on the other phylogroups. Strains that possess the effector gene are coloured in blue, and those that are cherry-pathogenic are highlighted in red. Strains with long-read sequenced genomes are in black boxes. The transfer events predicted by RANGER-DTL are shown by purple arrows. The bar shows substitutions per site. (b) DNA alignments of genomic regions containing these two effector genes, showing similar flanking regions between cherry pathogens. Alignments are colour coded based on similarity; identical residues are in grey, whereas dissimilar residues appear in black. The effector gene is coloured in red, mobile element genes are in green and other coding sequences are in blue. Cherry-isolated strains are named in pink, whilst the nonpathogenic plum strain R1-5300 is in blue. Gene name abbreviations: Hypo, hypothetical protein gene; ISPsy4, insertion sequence; ME, mobile element.

PAMP-triggered immunity (PTI) responses (Zhang *et al.*, 2010). This effector could play a similar role in PTI suppression in cherry.

HopBB1 and other members of the HopF family were abundant in cherry pathogens. All HopF members share an N-terminus and myristoylation sites for plant cell membrane localization (Lo *et al.*, 2016) and interfere with PTI and ETI in model plants (Wang *et al.*, 2010; Wu *et al.*, 2011; Hurley *et al.*, 2014). The presence of multiple *hopF* homologues in cherry pathogens and specific gain of *hopBB1* suggested the importance of their function. In comparison, HopH1 and HopBF1 are understudied. HopH1 is a protease, homologous to the *Ralstonia*

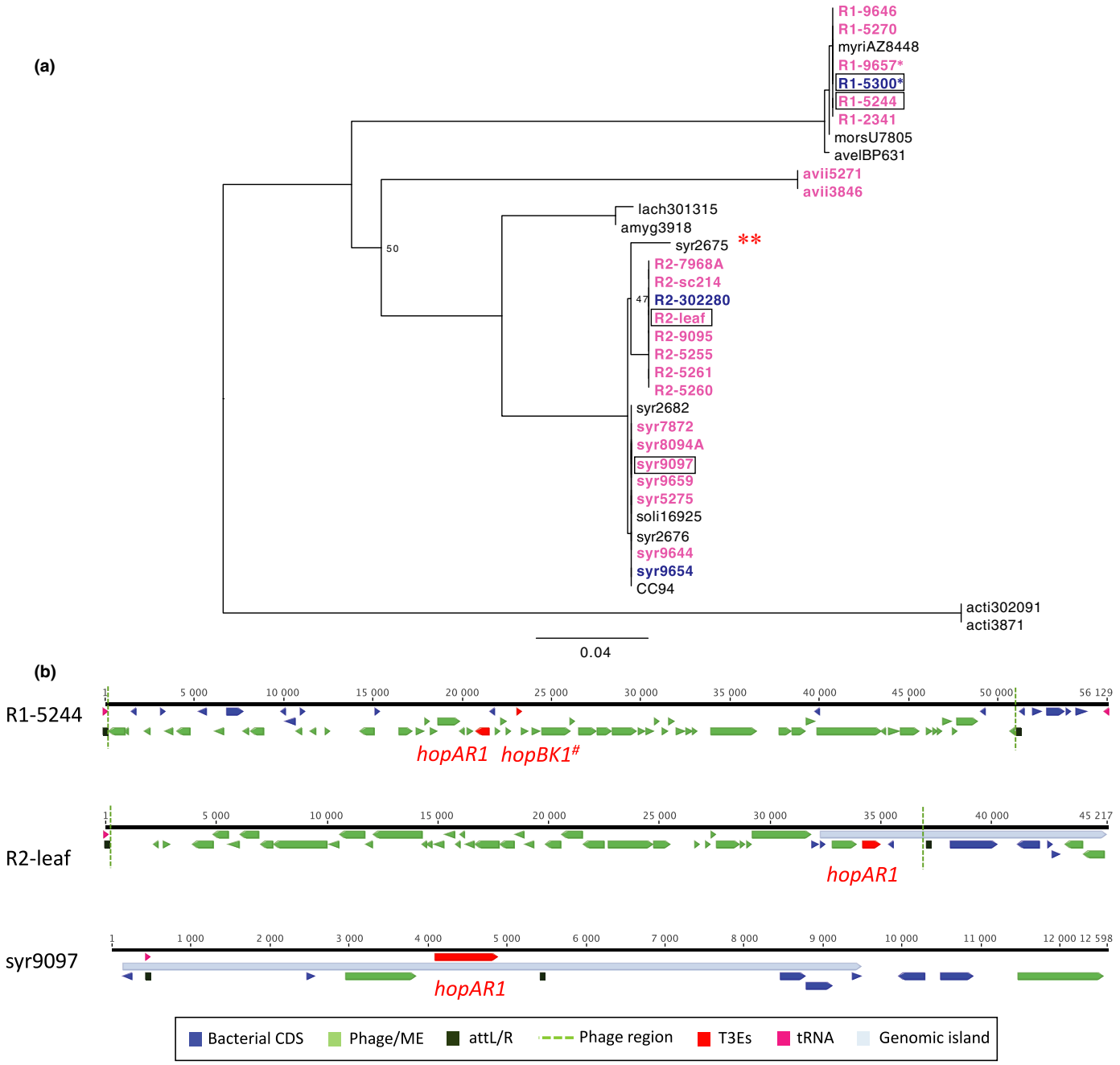

Fig. 5 Evolution of *hopAR1* in different clades of *Pseudomonas syringae* containing cherry pathogens. (a) Maximum likelihood phylogenetic tree built using the nucleotide sequences of the *hopAR1* gene. Cherry and plum isolated strains are highlighted in pink and blue respectively; those names followed by single asterisks were nonpathogenic on cherry in controlled pathogenicity tests. Strains with long-read sequenced genomes are in black boxes. Bootstrap supports < 99% are shown. The bar is nucleotide substitutions per site. Double asterisks point to the clustering of *P. syringae* pv *morsprunorum* (*Psm*) R2 sequences with syr2675. (b) Genomic locations of the *hopAR1* gene in the three PacBio-sequenced cherry pathogens. The gene is located within prophage sequences in *Psm* R1 and R2 (see Table S14 for details), whereas in syr9097 it is on a genomic island (GI) adjacent to a transfer RNA (tRNA) gene. Effector genes are coloured in red, other coding sequences in blue, phage genes predicted by PHASTER and mobile element genes are in green, tRNA genes in pink and GIs predicted (GI14 in *Psm* R2 and GI23 in *P. syringae* pv *syringae* (*Pss*)) in light blue. Predicted phage *att* sites are in dark green, with sites homologous to R2-leaf in *Pss* 9097 also indicated even though a phage is not predicted here. The ends of predicted prophage sequences are denoted with dashed green lines. #*hopBK1* is a pseudogene in this strain. CDS, coding sequence; ME, mobile element; T3E, type III effector.

solanacearum Rip36 protein (Nahar *et al.*, 2014). This T3E gene was found on GI37 in *Psm* R2-leaf and was within 3 kb of *hopF4* (Fig. S25), indicating that these two T3Es may have been gained together. HopBF1 was first discovered in *P. syringae* pv *aptata* and *oryzae* (Baltrus *et al.*, 2011), but its role is undetermined. This study therefore identified candidate T3Es important for cherry pathogenicity that should be the focus of future functional studies.

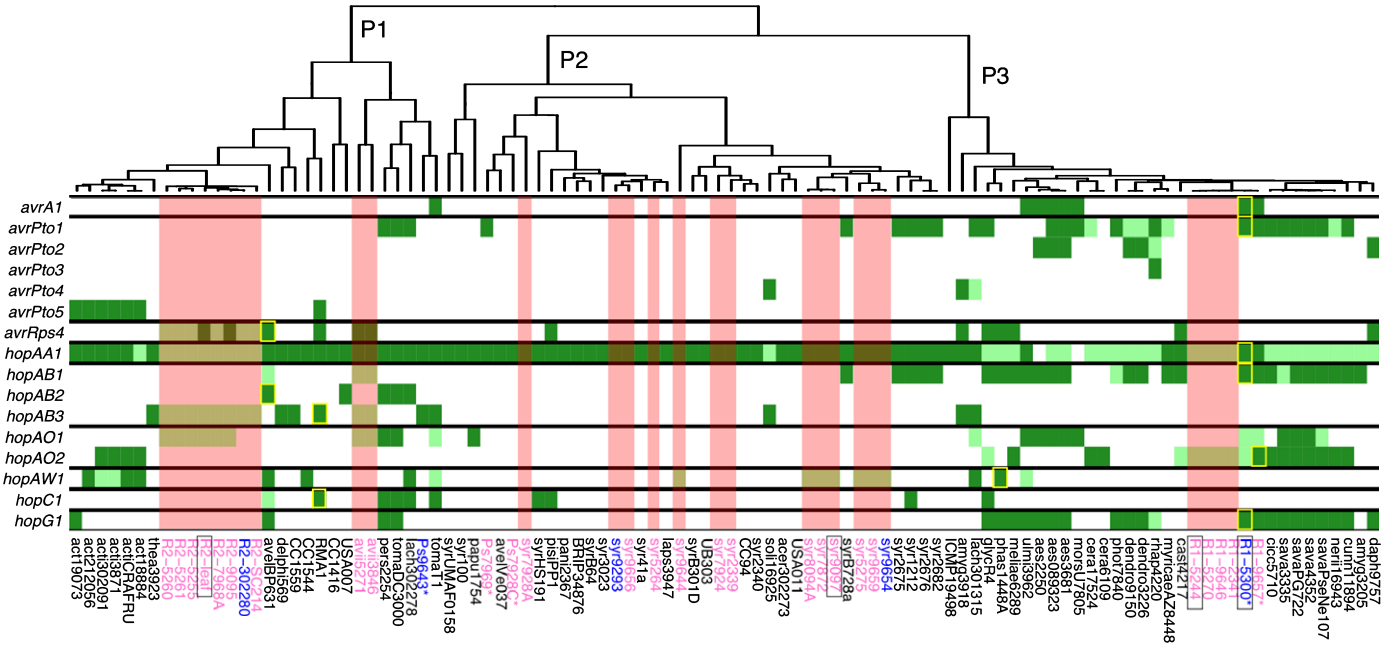

Fig. 6 Distribution of putative avirulent type III effector (T3E) genes across the *Pseudomonas syringae* phylogeny. The heatmap shows the presence and absence of T3Es studied in the functional analysis. This was constructed as in Fig. 2. The cherry-pathogenic clades are illustrated via vertical shading of cells in pink. Full-length putative *avr* genes used in cloning work are outlined with yellow boxes.

Phytotoxin biosynthesis gene clusters were also identified. Coronatine is present on a plasmid in pathogenic *Psm* R1 and may be one of the factors that differentiate pathogens from non-pathogens in this clade. Coronatine functions in virulence by downregulating salicylic acid defence signalling (Grant & Jones, 2009). Necrosis-inducing lipodepsipeptide toxins were common in P2. All cherry-pathogenic *Pss* strains possessed at least one biosynthesis cluster. The ability of *Pss* strains to cause necrosis on cherry fruits has been linked to toxins (Scholz-Schroeder *et al.*, 2001). Interestingly, two nonpathogenic P2b cherry strains lacked all phytotoxins, a deficiency that probably contributes to their lack of pathogenicity.

All cherry-pathogenic *Pss* strains had reduced effector repertoires. This observation supports the hypothesis that a phenotypic trade-off exists, with strains retaining few T3Es, whilst relying more on phytotoxins for pathogenicity (Baltrus *et al.*, 2011; Hockett *et al.*, 2014). If this pathogenic strategy has evolved in the P2 clade, it raises the question as to how it affects host specificity and virulence. P2 strains often infect more than one host species (Rezaei & Taghavi, 2014). These strains probably possess fewer ETI-inducing avirulence factors that restrict effector-rich strains to particular hosts, so may be more successful generalists. The reduction in T3E repertoire, however, may be limiting, as strains may be less capable of the long-term disease suppression required at the start of a hemi-biotrophic interaction.

Most cherry-pathogenic clades possessed genes involved in aromatic compound degradation, shown to be important in virulence on olive (Caballo-Ponce *et al.*, 2016), and ice nucleation genes that stimulate frost damage (Lamichhane *et al.*, 2014). The fact that not all cherry-pathogenic clades possessed these genes suggests they are not essential requirements for bacterial canker; however, they may contribute to niche persistence. For example, Crosse & Garrett (1966) observed that *Psm* R1 survived in cankers for longer than *Pss*. Increased persistence might be linked to genes involved in woody-tissue adaptation.

Horizontal gene transfer has been important in the acquisition of key effectors

HGT is important for effector shuffling within *P. syringae* (Arnold & Jackson, 2011). Pathogenicity-associated T3Es *hopBB1* and *hopBF1* were plasmid encoded and showed evidence of HGT between the cherry-pathogenic clades in P1 and P3. Plasmid profiling revealed that cherry pathogens in these phylogroups possessed native plasmids, some of which were putatively conjugative, indicating the importance of plasmids in gene exchange. By contrast, most cherry-pathogenic *Pss* strains lacked plasmids.

The T3E *hopAR1* was chromosomal in all long-read sequenced genomes. This gene was found within distinct prophage sequences in *Psm* R1 and R2. To our knowledge this is the first reported example of a plant pathogen T3E located within a prophage sequence. Interestingly, the *Psm* R2 *hopAR1* gene homologue was most similar to *hopAR1* from a P2 bean strain syr2675, which is a close relative of cherry *Pss*. This strain possessed a homologous phage to *Psm* R2, indicating that HGT of this T3E between phylogroups may have been phage mediated. This striking example of convergent acquisition of *hopAR1* in the cherry pathogens, putatively through distinct prophages in *Psm* R1 and R2, and a GI in *Pss* indicates that this T3E may have important roles in virulence. The well-characterized

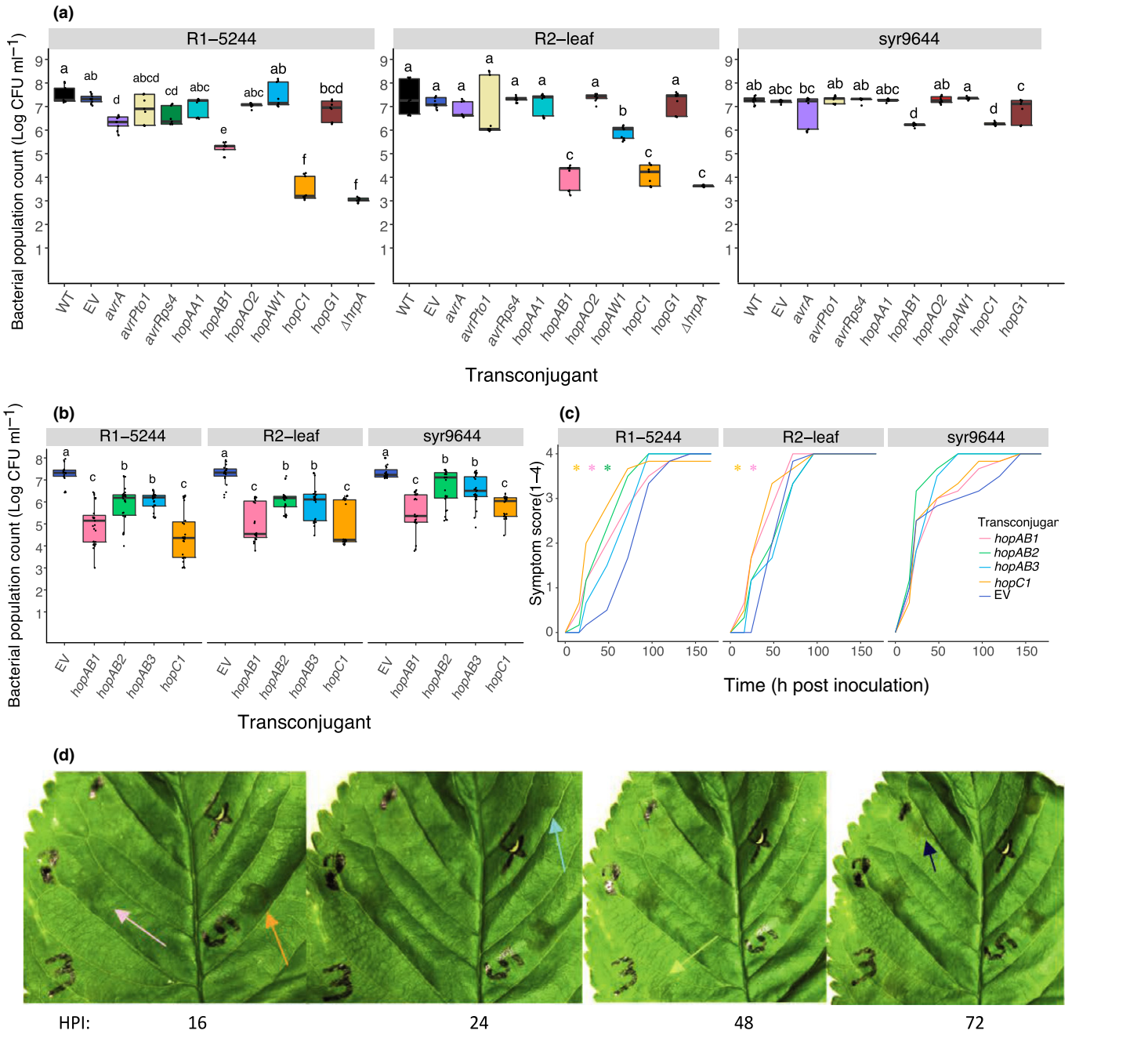

Fig. 7 Identification of avirulence factors activating effector-triggered immunity in cherry. (a) Boxplot of an initial 10-d population count analysis of cherry pathogens (R1-5244, R2-leaf and syr9644) transconjugants expressing candidate avirulence genes. The data presented are based on one experiment, with three leaf replicates and three nested technical replicates ($n = 9$). Boxplots show median and interquartile range (IQR) and whiskers extend to values 1.5× IQR above and below the median. All data points are plotted with circles. Controls included the wild-type strain, a strain containing the empty pBBR1MCS-5 vector and a Δ*hrpA* deletion mutant (for R1-5244 and R2-leaf). A separate ANOVA was performed for each cherry pathogen (R1-5244, R2-leaf and syr9644) and the Tukey-HSD significance groups ($P = 0.05$; confidence level: 0.95) for each strain are presented above each boxplot. (b) Boxplot of 10-d population counts of cherry pathogens (R1-5244, R2-leaf and syr9644) expressing different HopAB alleles and HopC1. The data presented are based on three independent experiments ($n = 27$). Tukey-HSD significance groups are presented above each boxplot. (c) Symptom development of R1-5244, R2-leaf, syr9644 transconjugants. Mean symptom score values are presented and represent two independent experiments ($n = 6$). Symptoms assessed as degree of browning of the infiltration site: 1, limited browning; 2, < 50%; 3, > 50%; 4, 100% of the infiltrated area brown. Analysis was based on area under disease progress curve (AUDPC) values (0–48 h). An ANOVA was performed on AUDPC values, with asterisks indicating significantly different from the empty vector (EV) control. (d) Symptom development over time on a representative leaf inoculated with R1-5244 transconjugants. HPI, hours post inoculation. The order of strains: 1, EV; 2, *hopAB1*; 3, *hopAB2*; 4, *hopAB3*; 5, *hopC1*. Arrows indicate the first appearance of symptoms associated with each strain and are coloured based on the graph in (c). ANOVA tables for all statistical analyses are presented in Tables S18–S24, and AUDPC values are in Table S25. CFU, colony-forming units.

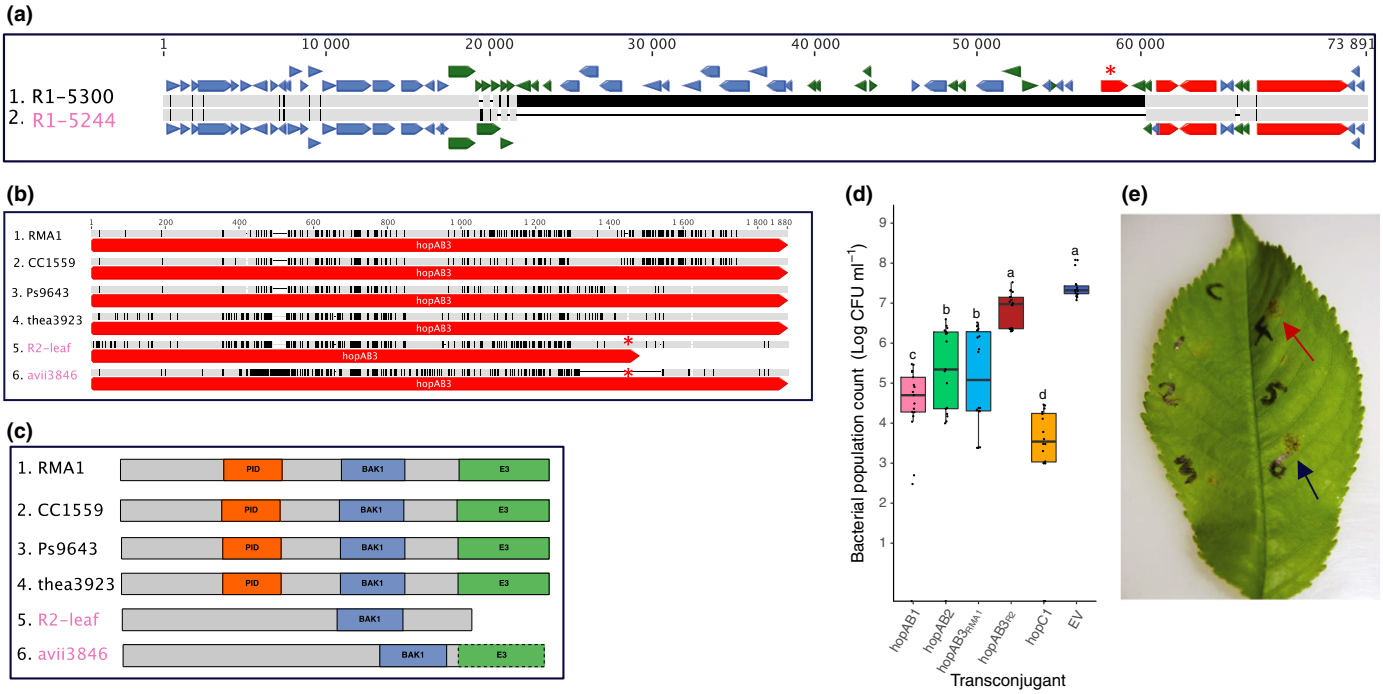

Fig. 8 *hopAB* alleles have been both lost and truncated in cherry pathogens. (a) Alignment of the DNA region surrounding *hopAB1* in *Pseudomonas syringae* pv *morsprunorum* (*Psm*) R1 strains. Grey indicates sequence identity, whereas black indicates divergence. The effector genes are coloured in red, whereas other coding sequences are in blue and mobile element genes are in green. Asterisk indicates the location of *hopAB1* in R1-5300, whilst the upstream effectors are *hopQ1*, *hopD1* and *hopR1*. (b) DNA alignment of the *hopAB3* gene of *Psm* R2 and close out-groups. Asterisks indicate where the *hopAB3* gene has been truncated due to a GG insertion at 1404 bp leading to a frameshift in *Psm* R2, whilst in *P. syringae* pv *avii* (avii3846) there is a deletion at the end of the gene. (c) Diagrams showing the location of key domains in the HopAB3 protein including the Pto-interaction domain (PID), BAK1-interacting domain (BAK1) and E3 ubiquitin ligase (E3). The E3 domain is lost completely from the *Psm* R2 protein, whereas in avii3846 the beginning of this domain is lost. The PID domain was not detected in the cherry pathogen sequences. (d) Boxplot of 10-d population counts of R1-5244 transconjugants expressing three different full-length *hopAB* alleles, truncated $hopAB3_{R2\text{-}leaf}$ and *hopC1*. The boxplots were constructed as in Fig. 7. The data presented are based on two independent experiments ($n = 18$). Tukey-HSD significance groups ($P = 0.05$; confidence level: 0.95) are presented above each boxplot (full statistical analysis is in Table S26). (e) Representative image of symptoms 10 d post inoculation (dpi) with the different R1-5244 transconjugants when inoculated at a low concentration (2×10^6 colony-forming units (CFU) ml^{-1}) to observe pathogenicity. Arrows point to pathogenic symptoms in the strain expressing $hopAB3_{R2\text{-}leaf}$ and the empty vector (EV) strain, colour coded as in (d).

P. syringae pv *phaseolicola* R3 homologue is not associated with a phage, but has been shown to undergo dynamic evolution on a mobile genomic island *in planta* in resistant bean cultivars (Neale *et al.*, 2016).

Several T3Es in *Psm* R1, R2 and *Pss* were located on GIs. To determine the likely source of GIs in cherry strains, all other *P. syringae* strains were searched for homologous sequences. There was evidence of *Psm* R1 and R2 islands being shared between cherry pathogen clades indicative of HGT events occurring between strains occupying the same ecological niche.

Functional genomics revealed convergent loss of an *avr* factor

Genes from the *hopAB* and *avrPto* families form a redundant effector group (REG) vital for early PTI suppression in herbaceous species (Jackson *et al.*, 1999; Lin & Martin, 2005; Kvitko *et al.*, 2009). Both effectors also trigger ETI by interacting with the serine-threonine kinase R protein Pto in tomato (Kim *et al.*, 2002).

Across the *P. syringae* complex, the REG was common (Fig. S26), but cherry pathogens all lacked full-length genes. The *hopAB1* gene has been lost from *Psm* R1, whilst the *Psm* R2 and *P. syringae* pv *avii* predicted HopAB3 proteins lacked the PID and E3-ubiquitin ligase domains through contrasting mutations. *P. syringae* pv *avii* also possessed a truncated *hopAB1* gene (Fig. S24), lacking the PID domain. The lack of a PID in cherry pathogen HopAB proteins suggested that they could have diverged to avoid a Pto-based recognition system in cherry.

Full-length members of this REG were expressed in cherry pathogens to determine their role *in planta*. The addition of HopAB alleles (HopAB1–3) consistently reduced population growth of pathogenic strains *in planta* and triggered a response consistent with the HR. If this effector does trigger immunity in cherry, there may have been selection pressure for its loss or pseudogenization in cherry pathogens in order to reduce avirulence activity. The truncated version of HopAB3 in R2-leaf was found not to exhibit avirulence activity as its expression did not reduce the growth of R1-5244 *in planta*. Although AvrPto is part of the same REG, its expression had no effect on the ability of cherry pathogens to multiply in leaves. The absence of AvrPto from cherry pathogens is therefore unlikely to be driven by avirulence, but could be due to the lack of HopAB virulence targets *in planta*.

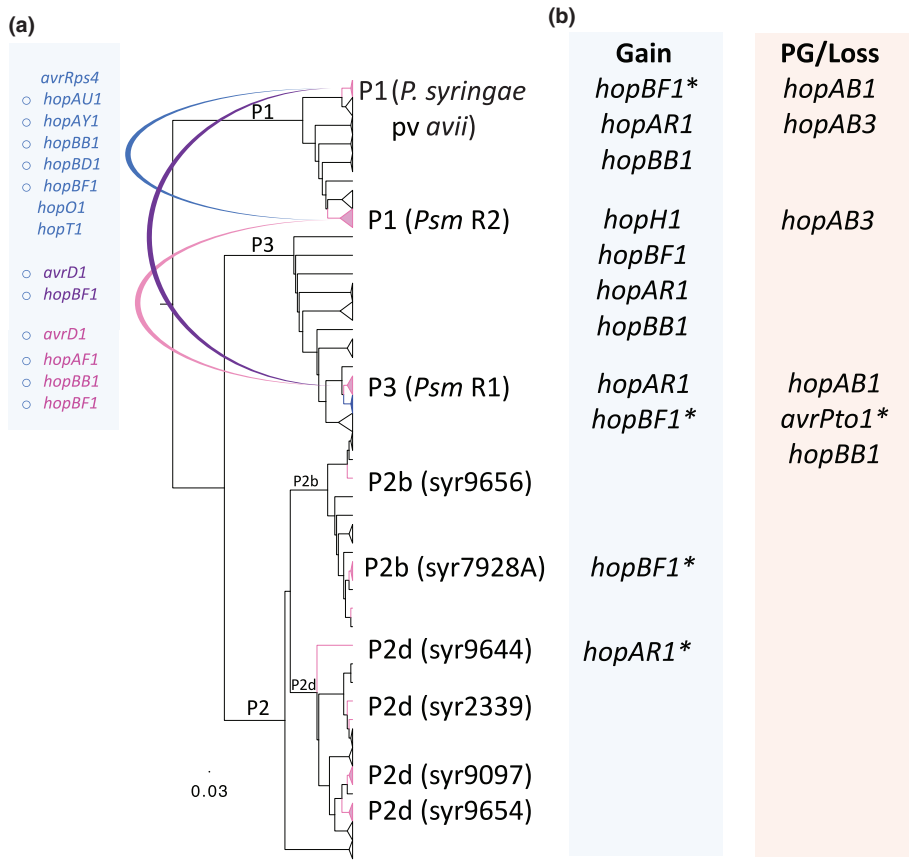

Fig. 9 Model highlighting genomic events that have led to the evolution of pathogenicity towards cherry. (a) The core-genome phylogeny is presented with phylogroups (P1–P3) labelled. Bar shows substitutions per site. For visualization, clades within the phylogenetic tree have been collapsed with clades containing cherry pathogens in pink (*Pseudomonas syringae* pv *morsprunorum* (*Psm*) R1 nonpathogenic strains in blue). Examples of cherry pathogens within each clade of P2 are named. Horizontal gene transfer (HGT) events predicted using Ranger-dtl are shown, with open circles representing those upon plasmids in long-read sequenced strains. (b) The key gains and losses of associated virulence genes in strains pathogenic to cherry are described based on gain loss mapping engine (Gloome) analysis. Asterisks indicate the probability of this effector being gained/lost predicted using Gloome was slightly lower than the threshold of 0.8, but exceeded 0.65 (see Table S12 for details).

As this REG is vital for early disease suppression in model strains, cherry pathogens must rely on other T3Es to fulfil this role.

The variation in *hopAB1* presence in *Psm* R1 is intriguing. *Psm* R1 strains may be pathogenic on both cherry and plum (Δ*hopAB1*) or just pathogenic on plum (possessed *hopAB1*) as recorded in Hulin *et al.* (2018). This suggests that the host proteins in cherry that detect the presence of HopAB are not present/functioning in plum. Future studies may determine how the two host immune responses diverged and could examine *hopAB* diversity across *Prunus* pathogens. This study focused on bacterial canker of *P. avium*; however, strains isolated from additional *Prunus* spp. that cause other diseases were included, such as *P. syringae* pv *cerasicola* (bacterial gall of hybrid cherry *Prunus* × *yedoensis*; Kamiunten *et al.*, 2000), *P. syringae* pv *morsprunorum* FTRSU7805 (canker of apricot), *P. syringae* pv *amygdali* (canker of almond) and *P. syringae* pv *persicae* (decline and canker of peach) (Table 1). All apart from *P. syringae* pv *amygdali* 3205 and *P. syringae* pv *persicae* lacked HopAB (Fig. 2), indicating that there may be a conserved resistance mechanism regulating ETI activated by this effector family in *Prunus* species.

Linking genomics to the evolution of cherry pathogenicity

Cherry pathogenicity has arisen independently within *P. syringae*, with strains using both shared and distinctive virulence strategies. Cherry-pathogenic clades in P1 and P3 have large effector repertoires. Cherry *Pss* were found across P2 with reduced T3Es and several phytotoxin gene clusters. Key events in the evolution of cherry pathogenicity (Fig. 9) appear to be the acquisition of virulence-associated effectors, often through HGT. Putatively important T3Es included *hopAR1*, members of the *hopF* family such as *hopBB1* and the other T3Es *hopBF1* and *hopH1*. Significantly, the loss/pseudogenization of HopAB effectors has also occurred in multiple clades. Within P2, the different cherry-infecting *Pss* clades have slight differences in their virulence factor repertoires that may reflect their convergent gain of pathogenicity. Clades differed in T3E content, phytotoxin genes and possession of genes for catechol degradation (Fig. 2), and thus pathogenicity was achieved with variable virulence factor repertoires. This study demonstrates that populations genomics can be used to examine a complex disease of a perennial plant species. A huge dataset was

narrowed down to several candidate host-specificity-associated genes, two of which (*hopAB* and *hopC1*) encode proteins that had putative avirulence functions *in planta*.

Acknowledgements

We acknowledge the East Malling Trust, University of Reading and BBSRC for funding (BB/P006272/1). Library preparation for PacBio sequencing was performed at the Earlham Institute, whilst for wild cherry strains Illumina MiSeq sequencing, libraries were prepared at the Genomics Facility, University of Warwick. We thank Steve Roberts, Helen Neale, Mateo San José and David Guttman for providing bacterial strains. We thank the East Malling Farm and Glass staff for plant maintenance. The authors declare no conflict of interest.

Author contributions

M.T.H., J.W.M., R.W.J. and R.J.H. conceived and designed the study as well as writing the manuscript. M.T.H. performed bioinformatics, statistical analysis and laboratory work. J.G.V. isolated some of the strains used in this study, J.G.V. and E.B.H. selected representative strains and prepared DNA for MiSeq sequencing of nine strains and L.B. assembled these nine sequences. H.J.B. performed the MinION library preparation and sequencing. A.D.A. assisted in bioinformatics pipeline development. All authors read and reviewed the final manuscript.

ORCID

Richard J. Harrison http://orcid.org/0000-0002-3307-3519

References

Alfano JR, Charkowski A, Deng W, Badel J, Petnicki-Ocwieja T, van Dijk K, Collmer A. 2000. The *Pseudomonas syringae* Hrp pathogenicity island has a tripartite mosaic structure composed of a cluster of type III secretion genes bounded by exchangeable effector and conserved effector loci that contribute to parasitic fitness and pathogenicity in plants. *Proceedings of the National Academy of Sciences, USA* **97**: 4856–4861.

Almeida NF, Yan S, Lindeberg M, Studholme DJ, Schneider DJ, Condon B, Liu H, Viana CJ, Warren A, Evans C *et al.* 2009. A draft genome sequence of *Pseudomonas syringae* pv. *tomato* T1 reveals a type III effector repertoire significantly divergent from that of *Pseudomonas syringae* pv. *tomato* DC3000. *Molecular Plant–Microbe Interactions: MPMI* **22**: 52–62.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *Journal of Molecular Biology* **215**: 403–410.

Arndt D, Grant JR, Marcu A, Sajed T, Pon A, Liang Y, Wishart DS. 2016. PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Research* **44**: W16–W21.

Arnold DL, Gibbon MJ, Jackson RW, Wood JR, Brown J, Mansfield JW, Taylor JD, 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*). *Physiological and Molecular Plant Pathology* **58**: 55–62.

Arnold DL, Jackson RW. 2011. Bacterial genomes: evolution of pathogenicity. *Current Opinion in Plant Biology* **14**: 385–391.

Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M *et al.* 2008. The RAST Server: rapid annotations using subsystems technology. *BMC Genomics* **9**: 75.

Baltrus DA, Dougherty K, Beckstrom-Sternberg SM, Beckstrom-Sternberg JS, Foster JT. 2014a. Incongruence between multi-locus sequence analysis (MLSA) and whole-genome-based phylogenies: *Pseudomonas syringae* pathovar *pisi* as a cautionary tale. *Molecular Plant Pathology* **15**: 461–465.

Baltrus DA, McCann HC, Guttman DS. 2017. Evolution, genomics and epidemiology of *Pseudomonas syringae*. *Molecular Plant Pathology* **18**: 152–168.

Baltrus DA, Nishimura MT, Dougherty KM, Biswas S, Mukhtar MS, Vicente JG, Holub EB, Dangl JL. 2012. The molecular basis of host specialization in bean pathovars of *Pseudomonas syringae*. *Molecular Plant–Microbe Interactions: MPMI* **25**: 877–888.

Baltrus DA, Nishimura MT, Romanchuk A, Chang JH, Mukhtar MS, Cherkis K, Roach J, Grant SR, Jones CD, Dangl JL. 2011. Dynamic evolution of pathogenicity revealed by sequencing and comparative genomics of 19 *Pseudomonas syringae* isolates. *PLoS Pathogens* **7**: 22.

Baltrus DA, Yourstone S, Lind A, Guilbaud C, Sands DC, Jones CD, Morris CE, Dangl JL. 2014b. Draft genome sequences of a phylogenetically diverse suite of *Pseudomonas syringae* strains from multiple source populations. *Genome Announcements* **2**: e01195-13.

Bansal MS, Alm EJ, Kellis M. 2012. Efficient algorithms for the reconciliation problem with gene duplication, horizontal transfer and loss. *Bioinformatics* **28**: 283–291.

Bartoli C, Carrere S, Lamichhane R, Varvaro L, Morris CE. 2015a. Whole-genome sequencing of 10 *Pseudomonas syringae* strains representing different host range spectra. *Genome Announcements* **3**: 2–3.

Bartoli C, Lamichhane JR, Berge O, Guilbaud C, Varvaro L, Balestra GM, Vinatzer BA, Morris CE. 2015b. A framework to gauge the epidemic potential of plant pathogens in environmental reservoirs: the example of kiwifruit canker. *Molecular Plant Pathology* **16**: 137–149.

Bender CL, Alarcón-Chaidez F, Gross DC. 1999. *Pseudomonas syringae* phytotoxins: mode of action, regulation, and biosynthesis by peptide and polyketide synthetases. *Microbiology and Molecular Biology Reviews* **63**: 266–292.

Berge O, Monteil CL, Bartoli C, Chandeysson C, Guilbaud C, Sands DC, Morris CE. 2014. A user's guide to a data base of the diversity of *Pseudomonas syringae* and its application to classifying strains in this phylogenetic complex. *PLoS ONE* **9**: e105547.

Berlin K, Koren S, Chin C-S, Drake JP, Landolin JM, Phillippy AM. 2015. Assembling large genomes with single-molecule sequencing and locality-sensitive hashing. *Nature Biotechnology* **33**: 623–630.

Bruns H, Crüsemann M, Letzel A-C, Alanjary M, McInerney JO, Jensen PR, Schulz S, Moore BS, Ziemert N. 2018. Function-related replacement of bacterial siderophore pathways. *The ISME Journal* **12**: 320–329.

Buell CR, Joardar V, Lindeberg M, Selengut J, Paulsen IT, Gwinn ML, Dodson RJ, Deboy RT, Durkin AS, Kolonay JF *et al.* 2003. The complete genome sequence of the *Arabidopsis* and tomato pathogen *Pseudomonas syringae* pv. *tomato* DC3000. *Proceedings of the National Academy of Sciences, USA* **100**: 10181–10186.

Bull CT, de Boer SH, Denny TP, Firrao G, Fischer-Le Saux M, Saddler GS, Scortichini M, Stead DE, Takikawa Y. 2010. Comprehensive list of names of plant pathogenic bacteria, 1980–2007. *Journal of Plant Pathology* **92**: 551–592.

Bultreys A, Kaluzna M. 2010. Bacterial cankers caused by *Pseudomonas syringae* on stone fruit species with special emphasis on the pathovars *syringae* and *morsprunorum* Race 1 and Race 2. *Journal of Plant Pathology* **92**: S1.21–S1.33.

Buonaurio R, Moretti C, da Silva DP, Cortese C, Ramos C, Venturi V. 2015. The olive knot disease as a model to study the role of interspecies bacterial communities in plant disease. *Frontiers in Plant Science* **6**: 434.

Butler MI, Stockwell PA, Black MA, Day RC, Lamont IL, Poulter RTM. 2013. *Pseudomonas syringae* pv. *actinidiae* from recent outbreaks of kiwifruit bacterial canker belong to different clones that originated in China. *PLoS ONE* **8**: e57464.

Caballo-Ponce E, van Dillewijn P, Wittich R, Ramos C. 2016. WHOP, a genomic region associated with woody hosts in the *Pseudomonas syringae* complex contributes to the virulence and fitness of *Pseudomonas savastanoi* pv. *savastanoi* in olive plants. *Molecular Plant–Microbe Interactions* **30**: 113–126.

Castresana J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular Biology and Evolution* **17**: 540–552.

Cohen O, Ashkenazy H, Belinky F, Huchon D, Pupko T. 2010. GLOOME: gain loss mapping engine. *Bioinformatics* **26**: 2914–2915.

Crosse JE, Garrett CME. 1966. Bacterial canker of stone-fruits. VII. Infection experiments with *Pseudomonas morsprunorum* and *Ps. syringae*. *Annals of Applied Biology* **58**: 31–41.

Cunnac S, Chakravarthy S, Kvitko BH, Russell AB, Martin GB, Collmer A. 2011. Genetic disassembly and combinatorial reassembly identify a minimal functional repertoire of type III effectors in *Pseudomonas syringae*. *Proceedings of the National Academy of Sciences, USA* **108**: 2975–2980.

Darling AE, Mau B, Perna NT. 2010. Progressivemauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS ONE* **5**: e11147.

Dhillon BK, Laird MR, Shay JA, Winsor GL, Lo R, Nizam F, Pereira SK, Waglechner N, McArthur AG, Langille MGI *et al.* 2015. IslandViewer 3: more flexible, interactive genomic island discovery, visualization and analysis. *Nucleic Acids Research* **43**: W104–W108.

Dudnik A, Dudler R. 2014. Genomics-based exploration of virulence determinants and host-specific adaptations of *Pseudomonas syringae* strains isolated from grasses. *Pathogens* **3**: 121–148.

Edwards DJ, Holt KE. 2013. Beginner's guide to comparative bacterial genome analysis using next-generation sequence data. *Microbial Informatics and Experimentation* **3**: 2.

Feil H, Feil WS, Chain P, Larimer F, DiBartolo G, Copeland A, Lykidis A, Trong S, Nolan M, Goltsman E *et al.* 2005. Comparison of the complete genome sequences of *Pseudomonas syringae* pv. *syringae* B728a and pv. *tomato* DC3000. *Proceedings of the National Academy of Sciences, USA* **102**: 11064–11069.

Feil WS, Feil H, Copeland A. 2012. *Bacterial genomic DNA isolation using CTAB*. [WWW document] URL http://1ofdmq2n8tc36m6i46scovo2e.wpengine.netdna-cdn.com/wp-content/uploads/2014/02/JGI-Bacterial-DNA-isolation-CTAB-Protocol-2012.pdf [accessed 28 December 2017]

Gardan L, Shafik H, Belouin S, Broch R, Grimont F, Grimont P. 1999. DNA relatedness among the pathovars of *Pseudomonas syringae* and description of *Pseudomonas tremae* sp. nov. and *Pseudornonas cannabina* sp. nov. (ex Sutic and Dowson 1959). *International Journal of Systematic Bacteriology* **49**: 469–478.

Gardiner DM, Stiller J, Covarelli L, Lindeberg M, Shivas RG, Manners JM. 2013. Genome sequences of *Pseudomonas* spp. isolated from cereal crops. *Genome Announcements* **1**: e00209–e00213.

Gilbert V, Planchon V, Legros F, Maraite H, Bultreys A. 2009. Pathogenicity and aggressiveness in populations of *Pseudomonas syringae* from Belgian fruit orchards. *European Journal of Plant Pathology* **126**: 263–277.

Grant MR, Jones JD. 2009. Hormone (dis)harmony moulds plant health and disease. *Science* **324**: 750–752.

Green S, Studholme DJ, Laue BE, Dorati F, Lovell H, Arnold D, Cottrell JE, Bridgett S, Blaxter M, Huitema E *et al.* 2010. Comparative genome analysis provides insights into the evolution and adaptation of *Pseudomonas syringae* pv. *aesculi* on *Aesculus hippocastanum*. *PLoS ONE* **5**: e10224.

Guttman DS, McHardy AC, Schulze-Lefert P. 2014. Microbial genome-enabled insights into plant–microorganism interactions. *Nature Reviews Genetics* **15**: 797–813.

Hockett KL, Nishimura MT, Karlsrud E, Dougherty K, Baltrus DA. 2014. *Pseudomonas syringae* CC1557: a highly virulent strain with an unusually small type III effector repertoire that includes a novel effector. *Molecular Plant–Microbe Interactions* **27**: 923–932.

Hulin MT, Mansfield JW, Brain P, Xu X, Jackson RW, Harrison RJ. 2018. Characterisation of the pathogenicity of strains of *Pseudomonas syringae* towards cherry and plum. *Plant Pathology*. doi: 10.1111/ppa.12834

Hunt M, De Silva N, Otto TD, Parkhill J, Keane JA, Harris SR. 2015. Circlator: automated circularization of genome assemblies using long sequencing reads. *Genome Biology* **16**: 294.

Hurley B, Lee D, Mott A, Wilton M, Liu J, Liu YC, Angers S, Coaker G, Guttman DS, Desveaux D. 2014. The *Pseudomonas syringae* type III effector HopF2 suppresses *Arabidopsis* stomatal immunity. *PLoS ONE* **9**: e114921.

Jackson RW, Athanassopoulos E, Tsiamis G, Mansfield JW, Sesma A, Arnold DL, Gibbon MJ, Murillo J, Taylor JD, Vivian A. 1999. Identification of a pathogenicity island, which contains genes for virulence and avirulence, on a large native plasmid in the bean pathogen *Pseudomonas syringae* pathovar *phaseolicola*. *Proceedings of the National Academy of Sciences, USA* **96**: 10875–10880.

Joardar V, Lindeberg M, Jackson RW, Selengut J, Dodson R, Brinkac LM, Daugherty SC, Deboy R, Durkin AS, Giglio MG *et al.* 2005. Whole-genome sequence analysis of *Pseudomonas syringae* pv. *phaseolicola* 1448A reveals divergence among pathovars in genes involved in virulence and transposition. *Journal of Bacteriology* **187**: 6488–6498.

Jones JDG, Dangl JL. 2006. The plant immune system. *Nature* **444**: 323–329.

Kamiunten H, Nakaol T, Oshida S. 2000. Agent of bacterial gall of cherry tree. *Journal of General Plant Pathology* **66**: 219–224.

Katoh K, Misawa K, Kuma K, Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research* **30**: 3059–3066.

Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C *et al.* 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**: 1647–1649.

Kim YJ, Lin NC, Martin GB. 2002. Two distinct *Pseudomonas* effector proteins interact with the Pto kinase and activate plant immunity. *Cell* **109**: 589–598.

Kvitko BH, Collmer A. 2011. Construction of *Pseudomonas syringae* pv. *tomato* DC3000 mutant and polymutant strains. In: McDowell J, ed. *Plant immunity. Methods in molecular biology (methods and protocols), Vol. 712*. New York, NY, USA: Humana Press.

Kvitko BH, Park DH, Velásquez AC, Wei C-F, Russell AB, Martin GB, Schneider DJ, Collmer A. 2009. Deletions in the repertoire of *Pseudomonas syringae* pv. *tomato* DC3000 type III secretion effector genes reveal functional overlap among effectors. *PLoS Pathogens* **5**: e1000388.

Lamichhane JR, Varvaro L, Parisi L, Audergon J-M, Morris CE. 2014. Chapter four – Disease and frost damage of woody plants caused by *Pseudomonas syringae*: seeing the forest for the trees. In: Sparks DL, ed. *Advances in agronomy, Vol. 126*. San Diego, CA, USA: Academic Press, 235–295.

Langmead B, Salzberg S. 2013. Fast gapped-read alignment with Bowtie2. *Nature Methods* **9**: 357–359.

Larkin M, Blackshields G, Brown NP, Chenna R, Mcgettigan P a, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R *et al.* 2007. Clustal W and Clustal X version 2.0. *Bioinformatics* **23**: 2947–2948.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**: 2078–2079.

Li L, Stoeckert CJJ, Roos DS. 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Research* **13**: 2178–2189.

Lin NC, Martin GB. 2005. An *avrPto*/*avrPtoB* mutant of *Pseudomonas syringae* pv. *tomato* DC3000 does not elicit Pto-mediated resistance and is less virulent on tomato. *Molecular Plant–Microbe Interactions* **18**: 43–51.

Lindgren PB. 1997. The role of *hrp* genes during plant–bacterial interactions. *Annual Review of Phytopathology* **35**: 129–152.

Liu H, Qiu H, Zhao W, Cui Z, Ibrahim M, Jin G, Li B, Zhu B, Xie GL. 2012. Genome sequence of the plant pathogen *Pseudomonas syringae* pv. *panici* LMG 2367. *Journal of Bacteriology* **194**: 5693–5694.

Liu W, Xie Y, Ma J, Luo X, Nie P, Zuo Z, Lahrmann U, Zhao Q, Zheng Y, Zhao Y *et al.* 2015. IBS: an illustrator for the presentation and visualization of biological sequences. *Bioinformatics* **31**: 3359–3361.

Lo T, Koulena N, Seto D, Guttman DS, Desveaux D. 2016. The HopF family of *Pseudomonas syringae* type III secreted effectors. *Molecular Plant Pathology* **18**: 1–12.

Loman NJ, Quinlan AR. 2014. Poretools: a toolkit for analyzing nanopore sequence data. *Bioinformatics* **30**: 3399–3401.

Mansfield JW, Genin S, Magori S, Citovsky V, Sriariyanum M, Ronald P, Dow M, Verdier V, Beer SV, Machado MA *et al.* 2012. Top 10 plant pathogenic bacteria in molecular plant pathology. *Molecular Plant Pathology* **13**: 614–629.

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**: e27297.

Martínez-García PM, Rodríguez-Palenzuela P, Arrebola E, Carrión VJ, Gutièrrez-Barranquero JA, Pèrez-García A, Ramos C, Cazorla FM, De Vicente A. 2015. Bioinformatics analysis of the complete genome sequence of the mango tree pathogen *Pseudomonas syringae* pv. *syringae* UMAF0158 reveals traits relevant to virulence and epiphytic lifestyle. *PLoS ONE* **10**: 1–26.

Matas IM, Castañeda-Ojeda MP, Aragón IM, Antúnez-Lamas M, Murillo J, Rodriquez-Palenzuela P, Lopez-Solanilla E, Ramos C. 2014. Translocation and functional analysis of *Pseudomonas savastanoi* pv. *savastanoi* NCPPB 3335 type III secretion system effectors reveals two novel effector families of the *Pseudomonas syringae* complex. *Molecular Plant–Microbe Interactions* **27**: 424–436.

Mazzaglia A, Studholme DJ, Taratufolo MC, Cai R, Almeida NF, Goodman T, Guttman DS, Vinatzer BA, Balestra GM. 2012. *Pseudomonas syringae* pv. *actinidiae* (PSA) isolates from recent bacterial canker of kiwifruit outbreaks belong to the same genetic lineage. *PLoS ONE* **7**: 1–11.

McCann HC, Li L, Liu Y, Li D, Pan H, Zhong C, Rikkerink EHA, Templeton MD, Straub C, Colombi E *et al.* 2017. Origin and evolution of the kiwifruit canker pandemic. *Genome Biology and Evolution* **9**: 932–944.

McCann HC, Rikkerink EHA, Bertels F, Fiers M, Lu A, Rees-George J, Andersen MT, Gleave AP, Haubold B, Wohlers MW *et al.* 2013. Genomic analysis of the kiwifruit pathogen *Pseudomonas syringae* pv. *actinidiae* provides insight into the origins of an emergent plant disease. *PLoS Pathogens* **9**: e1003503.

Ménard M, Sutra L, Luisetti J, Prunier JP, Gardan L. 2003. *Pseudomonas syringae* pv. *avii* (pv. nov.), the causal agent of bacterial canker of wild cherries (*Prunus avium*) in France. *European Journal of Plant Pathology* **109**: 565–576.

Monteil CL, Cai R, Liu H, Llontop MEM, Leman S, Studholme DJ, Morris CE, Vinatzer BA. 2013. Nonagricultural reservoirs contribute to emergence and evolution of *Pseudomonas syringae* crop pathogens. *New Phytologist* **199**: 800–811.

Monteil CL, Yahara K, Studholme DJ, Mageiros L, Méric G, Swingle B, Morris CE, Vinatzer BA, Sheppard SK. 2016. Population-genomic insights into emergence, crop-adaptation, and dissemination of *Pseudomonas syringae* pathogens. *Microbial Genomics* **21**: e000089.

Morel JB, Dangl JL. 1997. The hypersensitive response and the induction of cell death in plants. *Cell Death & Differentiation* **4**: 671–683.

Moretti C, Cortese C, Passos da Silva D, Venturi V, Ramos C, Firrao G, Buonaurio R. 2014. Draft genome sequence of *Pseudomonas savastanoi* pv. *savastanoi* strain DAPP-PG 722, isolated in Italy from an olive plant affected by knot disease. *Genome Announcements* **2**: e00864-14.

Mott GA, Thakur S, Smakowska E, Wang PW, Belkhadir Y, Desveaux D, Guttman DS. 2016. Genomic screens identify a new phytobacterial microbe-associated molecular pattern and the cognate *Arabidopsis* receptor-like kinase that mediates its immune elicitation. *Genome Biology* **17**: 98.

Moulton J, Vivian A, Hunter P, Taylor JD. 1993. Changes in cultivar-specificity toward pea can result from transfer of plasmid RP4 and other incompatibility group P1 replicons to *Pseudomonas syringae* pv. *pisi*. *Journal of General Microbiology* **39**: 3149–3155.

Nahar K, Matsumoto I, Taguchi F, Inagaki Y, Yamamoto M, Toyoda K, Shiraishi T, Ichinose Y, Mukaihara T. 2014. *Ralstonia solanacearum* type III secretion system effector Rip36 induces a hypersensitive response in the nonhost wild eggplant *Solanum torvum*. *Molecular Plant Pathology* **15**: 297–303.

Neale HC, Laister R, Payne J, Preston G, Jackson RW, Arnold DL. 2016. A low frequency persistent reservoir of a genomic island in a pathogen population ensures island survival and improves pathogen fitness in a susceptible host. *Environmental Microbiology* **18**: 4144–4152.

Neale HC, Slater RT, Mayne L-M, Manoharan B, Arnold DL. 2013. *In planta* induced changes in the native plasmid profile of *Pseudomonas syringae* pathover *phaseolicola* strain 1302A. *Plasmid* **70**: 420–424.

Nowell RW, Laue BE, Sharp PM, Green S. 2016. Comparative genomics reveals genes significantly associated with woody hosts in the plant pathogen *Pseudomonas syringae*. *Molecular Plant Pathology* **17**: 1409–1424.

O'Brien HE, Thakur S, Gong Y, Fung P, Zhang J, Yuan L, Wang PW, Yong C, Scortichini M, Guttman DS. 2012. Extensive remodeling of the *Pseudomonas syringae* pv. *avellanae* type III secretome associated with two independent host shifts onto hazelnut. *BMC Microbiology* **12**: 141.

O'Brien HE, Thakur S, Guttman DS. 2011. Evolution of plant pathogenesis in *Pseudomonas syringae*: a genomics perspective. *Annual Review of Phytopathology* **49**: 269–289.

Pagel M. 2004. Detecting correlated evolution on phylogenies: a general method for the comparative analysis of discrete characters. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **255**: 37–45.

Parkinson N, Bryant R, Bew J, Elphinstone J. 2011. Rapid phylogenetic identification of members of the *Pseudomonas syringae* species complex using the *rpoD* locus. *Plant Pathology* **60**: 338–344.

Pitman AR, Jackson RW, Mansfield JW, Kaitell V, Thwaites R, Arnold DL. 2005. Exposure to host resistance mechanisms drives evolution of bacterial virulence in plants. *Current Biology* **15**: 2230–2235.

Press MO, Li H, Creanza N, Kramer G, Queitsch C, Sourjik V, Borenstein E. 2013. Genome-scale co-evolutionary inference identifies functions and clients of bacterial Hsp90. *PLoS Genetics* **9**: e1003631.

Qi M, Wang D, Bradley CA, Zhao Y. 2011. Genome sequence analyses of *Pseudomonas savastanoi* pv. *glycinea* and subtractive hybridization-based comparative genomics with nine pseudomonads. *PLoS ONE* **6**: e16451.

Quevillon E, Silventoinen V, Pillai S, Harte N, Mulder N, Apweiler R, Lopez R. 2005. InterProScan: protein domains identifier. *Nucleic Acids Research* **33**: 116–120.

Ravindran A, Jalan N, Yuan JS, Wang N, Gross DC. 2015. Comparative genomics of *Pseudomonas syringae* pv. *syringae* strains B301D and HS191 and insights into intrapathovar traits associated with plant pathogenesis. *MicrobiologyOpen* **4**: 553–573.

Rezaei R, Taghavi SM. 2014. Host specificity, pathogenicity and the presence of virulence genes in Iranian strains of *Pseudomonas syringae* pv. *syringae* from different hosts. *Archives of Phytopathology and Plant Protection* **47**: 2377–2391.

Rodríguez-Palenzuela P, Matas IM, Murillo J, López-Solanilla E, Bardaji L, Pérez-Martínez I, Rodríguez-Moskera ME, Penyalver R, López MM, Quesada JM *et al.* 2010. Annotation and overview of the *Pseudomonas savastanoi* pv. *savastanoi* NCPPB 3335 draft genome reveals the virulence gene complement of a tumour-inducing pathogen of woody hosts. *Environmental Microbiology* **12**: 1604–1620.

Şahin F. 2001. Severe outbreak of bacterial speck, caused by *Pseudomonas syringae* pv. *tomato*, on field-grown tomatoes in the eastern Anatolia region of Turkey. *Plant Pathology* **50**: 799.

Sawada H, Shimizu S, Miyoshi T, Shinozaki T, Kusumoto S, Noguchi M, Naridomi T, Kikuhara K, Kansako M, Fujikawa T *et al.* 2015. *Pseudomonas syringae* pv. *actinidiae* biovar 3. *Japanese Journal of Phytopathology* **81**: 111–126.

Sawada H, Suzuki F, Matsuda I, Saitou N. 1999. Phylogenetic analysis of *Pseudomonas syringae* pathovars suggests the horizontal gene transfer of *argK* and the evolutionary stability of *hrp* gene cluster. *Journal of Molecular Evolution* **49**: 627–644.

Sawyer S. 1989. Statistical tests for detecting gene conversion. *Molecular Biology and Evolution* **6**: 526–538.

Scholz-Schroeder BK, Hutchison ML, Grgurina I, Gross DC. 2001. The contribution of syringopeptin and syringomycin to virulence of *Pseudomonas syringae* pv. *syringae* strain B301D on the basis of *sypA* and *syrB1* biosynthesis mutant analysis. *Molecular Plant–Microbe Interactions* **14**: 336–348.

Schulze-Lefert P, Panstruga R. 2011. A molecular evolutionary concept connecting nonhost resistance, pathogen host range, and pathogen speciation. *Trends in Plant Science* **16**: 117–125.

Scortichini M. 2010. Epidemiology and predisposing factors of some major bacterial diseases of stone and nut fruit trees species. *Journal of Plant Pathology* **92**: 73–78.

Sohn KH, Zhang Y, Jones JDG. 2009. The *Pseudomonas syringae* effector protein, AvrRPS4, requires in planta processing and the KRVY domain to function. *Plant Journal* **57**: 1079–1091.

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

Staskawicz BJ, Dahlbeck D, Keen NT. 1984. Cloned avirulence gene of *Pseudomonas syringae* pv. *glycinea* determines race-specific incompatibility on *Glycine max* (L.) Merr. *Proceedings of the National Academy of Sciences, USA* **81**: 6024–6028.

Thakur S, Weir BS, Guttman DS. 2016. Phytopathogen genome announcement: draft genome sequences of 62 *Pseudomonas syringae* type and pathotype strains. *Molecular Plant–Microbe Interactions* **29**: 243–246.

Vicente JG, Alves JP, Russell K, Roberts SJ. 2004. Identification and discrimination of *Pseudomonas syringae* isolates from wild cherry in England. *European Journal of Plant Pathology* **110**: 337–351.

Visnovsky SB, Fiers M, Lu A, Panda P, Taylor R, Pitman AR. 2016. Draft genome sequences of 18 strains of *Pseudomonas* isolated from kiwifruit plants in New Zealand and overseas. *Genome Announcements* **4**: e00061-16.

Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK *et al.* 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS ONE* **9**: e112963.

Wang Y, Li J, Hou S, Wang X, Li Y, Ren D, Chen S, Tang X, Zhou J-M. 2010. A *Pseudomonas syringae* ADP-ribosyltransferase inhibits Arabidopsis mitogen-activated protein kinase kinases. *The Plant Cell* **22**: 2033–2044.

Warnes G, Bolker B, Bonebakker L, Gentleman R, Huber W, Liaw A, Lumley T, Maechler M, Magnusson A, Moeller S *et al.* 2016. *gplots: various R programming tools for plotting data.* [WWW document] URL https://cran.r-project.org/web/packages/gplots/index.html [accessed 28 December 2017].

Wu S, Lu D, Kabbage M, Wei H-L, Swingle B, Records AR, Dickman M, He P, Shan L. 2011. Bacterial effector HopF2 suppresses *Arabidopsis* innate immunity at the plasma membrane. *Molecular Plant–Microbe Interactions: MPMI* **24**: 585–593.

Young JM. 1991. Pathogenicity and identification of the lilac pathogen, *Pseudomonas syringae* pv. *syringae* van Hall 1902. *Annals of Applied Biology* **118**: 283–298.

Young JM. 2010. Taxonomy of *Pseudomonas syringae*. *Journal of Plant Pathology* **92**: S1.5–S1.14.

Yu D, Yin Z, Li B, Jin Y, Ren H, Zhou J, Zhou W, Liang L, Yue J, Xu S. 2016. Gene flow, recombination, and positive selection in *Stenotrophomonas maltophilia*: mechanisms underlying the diversity of the widespread opportunistic pathogen. *Genome* **59**: 1063–1075.

Zhang J, Li W, Xiang T, Liu Z, Laluk K, Ding X, Zou Y, Gao M, Zhang X, Chen S *et al.* 2010. Receptor-like cytoplasmic kinases integrate signaling from multiple plant immune receptors and are targeted by a *Pseudomonas syringae* effector. *Cell Host & Microbe* **7**: 290–301.

Zhao W, Jiang H, Tian Q, Hu J. 2015. Draft genome sequence of *Pseudomonas syringae* pv. *persicae* NCPPB 2254. *Genome Announcements* **3**: e00555-15.

Zhao Y, Ma Z, Sundin GW. 2005. Comparative genomic analysis of the pPT23A plasmid family of *Pseudomonas syringae*. *Journal of Bacteriology* **187**: 2113–2126.

Supporting Information

Additional Supporting Information may be found online in the Supporting Information tab for this article:

Fig. S1 Plasmid profiles of all *Prunus*-infecting sequenced strains and some out-groups for comparison.

Fig. S2 Whole-genome alignment of R1-5244 and R1-5300 using ProgressiveMauve.

Fig. S3 Initial core genome phylogeny of 108 *Pseudomonas syringae* strains.

Fig. S4 Core genome phylogenetics of phylogroup 3 with removal of different out-group strains.

Fig. S5 Core genome phylogenetics of phylogroup 2 with removal of different out-group strains.

Fig. S6 Core genome phylogenetic tree with removal of the out-group strain *Pseudomonas syringae* pv *daphniphylli* 569.

Fig. S7 Core genome phylogenetic tree with removal of the out-group strains *Pseudomonas syringae* pv *daphniphylli* 569 and *P. syringae* pv *syringae* 1212.

Fig. S8 Core genome phylogenetic tree with removal of the out-group strain *Pseudomonas syringae* pv *eriobotryae* 4455.

Fig. S9 Core genome phylogenetic tree with removal of the out-group strains *Pseudomonas syringae* pv *eriobotryae* 4455 and *P. syringae* pv *syringae* 1212.

Fig. S10 Boxplot of 10-d population counts (Log CFU ml^{-1}) of *Pseudomonas syringae* pv *syringae* strains isolated from different host plant species, when inoculated into detached cherry leaves.

Fig. S11 Organisation of the structural T3SS and conserved effector locus (CEL) of the three PacBio-sequenced pathogens (R2-leaf, syr9097 and R1-5244).

Fig. S12 Alignment of the CEL in *Psm* R1.

Fig. S13 Percentage of phylogenetic trees that supported the dependant model of evolution in the BayesTraits analysis for T3E families that were significantly associated with pathogenicity.

Fig. S14 Predicted gain and loss of effector genes occurring on each branch leading to cherry pathogenic clades on the core genome phylogenetic tree.

Fig. S15 Core genome phylogenetic tree used in Gloome gain and loss analysis with the branch labels added.

Fig. S16 Maximum-likelihood phylogenetic trees for effector genes (*avrD-hopAY*) showing evidence of horizontal gene tranfer between cherry pathogen clades.

Fig. S17 Maximum-likelihood phylogenetic trees for effector genes (*hopBB-hopT*) showing evidence of horizontal gene tranfer between cherry pathogen clades.

Fig. S18 Core genome phylogenetic tree used in Ranger-dtl analysis with the branch labels added.

Fig. S19 *hopAR1* gene region in *Psm* R2 and phylogroup 2 strains.

Fig. S20 Genomic islands characteristic of cherry pathogens are found across *Pseudomonas syringae*.

Fig. S21 AvrRps4 alignment.

Fig. S22 HopAW1 alignment.

Fig. S23 Representative images of the tobacco hypersensitive response assay for R1-5244 and R2-leaf Δ*hrpA* mutants.

Fig. S24 Protein alignment of members of the HopAB effector family.

Fig. S25 Genomic region containing the *hopH1* gene in *Psm* R2-leaf.

Fig. S26 Heatmap showing the presence and absence of the *avrPto*/*hopAB* REG across the *Pseudomonas syringae* complex.

Table S1 All *Pseudomonas syringae* transconjugants and gene deletion mutants generated in this study

Table S2 All vectors/*E. coli* strains used in this study with antibiotic resistance information and reference

Table S3 All primers used in this study

Table S4 Protein sequences used in BLAST (Altschul *et al.*, 1990) analysis to identify virulence factors with NCBI accession number, gene name and abbreviation and source organism

Table S5 Plasmid-encoding gene identification in *Psm* R1-5244

Table S6 Plasmid-encoding gene identification in *Psm* R1-5300

Table S7 Plasmid-encoding gene identification in *Psm* R2-leaf

Table S8 Statistical information for different phylogenetic trees constructed from the core genome in this study

Table S9 Number of recombining genes in the core genome of each phylogroup estimated using GENECONV

Table S10 ANOVA table of 10 dpi population count of *Pseudomonas syringae* pv *syringae* outgroups on detached cherry leaves

Table S11 Effector presence comparisons between cherry pathogenic and nonpathogenic *Psm* R1 strains

Table S12 GLOOME output for all events occurring on branches leading to cherry pathogens and tips of the phylogeny

Table S13 Putative horizontal gene transfer events predicted using RANGER-DTL

Table S14 Output of PHASTER identification of prophages in the three cherry pathogenic PacBio-sequenced strains

Table S15 Genomic islands predicted in *Psm* R1-5244

Table S16 Genomic islands predicted in *Psm* R2-leaf

Table S17 Genomic islands predicted in *Pss* 9097

Table S18 ANOVA table of 10 dpi population count of *Psm* R1-5244 transconjugants (all candidate avirulence genes) on detached cherry leaves

Table S19 ANOVA table of 10 dpi population count of *Psm* R2-leaf transconjugants (all candidate avirulence genes) on detached cherry leaves

Table S20 ANOVA table of 10 dpi population count of *syr*9644 transconjugants (all candidate avirulence genes) on detached cherry leaves

Table S21 ANOVA table of 10 dpi population count of *Psm* R1-5244 transconjugants of candidate avirulence genes *hopAB* alleles and *hopC1* on detached cherry leaves

Table S22 ANOVA table of 10 dpi population count of *Psm* R2-leaf transconjugants of candidate avirulence genes *hopAB* alleles and *hopC1* on detached cherry leaves

Table S23 ANOVA table of 10 dpi population count of *syr*9644 trans-conjugants of candidate avirulence genes *hopAB* alleles and *hopC1* on detached cherry leaves

Table S24 ANOVA table of AUDPC symptom score analysis on detached cherry leaves

Table S25 AUDPC (Area Under the Disease Progress Curve) values for 0–48 h symptom development of pathogenic clades expressing different candidate avirulence genes

Table S26 ANOVA table of 10 dpi population count of *Psm* R1-5244 transconjugants of candidate avirulence genes *hopAB* alleles and *hopC1* as well as the *hopAB3* allele from cherry pathogen R2-leaf on detached cherry leaves

Methods S1 Detailed descriptions of methodology used.

Please note: Wiley Blackwell are not responsible for the content or functionality of any Supporting Information supplied by the authors. Any queries (other than missing material) should be directed to the New Phytologist Central Office.

See also the Commentary on this article by Baltrus & Orth, **219**: 482–484.