

RESEARCH ARTICLE

# Bioinformatic characterisation of the effector repertoire of the strawberry pathogen *Phytophthora cactorum*

Andrew D. Armitage<sup>1</sup>, Erik Lysøe<sup>2</sup>, Charlotte F. Nellist<sup>1</sup>, Laura A. Lewis<sup>1</sup>, Liliana M. Cano<sup>3,4</sup>, Richard J. Harrison<sup>1\*</sup>, May B. Brurberg<sup>2,5</sup>

**1** NIAB EMR, Kent, United Kingdom, **2** Norwegian Institute of Bioeconomy Research (NIBIO), Division of Biotechnology and Plant Health, Ås, Norway, **3** University of Florida, UF/IFAS Indian River Research and Education Center, Fort Pierce, Florida, United States of America, **4** The Sainsbury Laboratory, Norwich, United Kingdom, **5** Norwegian University of Life Sciences (NMBU), Department of Plant Sciences, Ås, Norway

\* [richard.harrison@emr.ac.uk](mailto:richard.harrison@emr.ac.uk)



**OPEN ACCESS**

**Citation:** Armitage AD, Lysøe E, Nellist CF, Lewis LA, Cano LM, Harrison RJ, et al. (2018) Bioinformatic characterisation of the effector repertoire of the strawberry pathogen *Phytophthora cactorum*. PLoS ONE 13(10): e0202305. <https://doi.org/10.1371/journal.pone.0202305>

**Editor:** Zonghua Wang, Fujian Agriculture and Forestry University, CHINA

**Received:** June 2, 2018

**Accepted:** June 23, 2018

**Published:** October 2, 2018

**Copyright:** © 2018 Armitage et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** This Whole Genome Shotgun project has been deposited at DDBJ/ENA/GenBank under the accession no. MJFZ00000000. The version described in this paper is version MJFZ01000000.

**Funding:** AA, CFN, LAL and RJH were funded by BBSRC grant BB/K017071/1 and BB/K017071/2, <https://bbsrc.ukri.org/research/grants/grants/AwardDetails.aspx?FundingReference=BB/K017071/1> and <https://bbsrc.ukri.org/research/>

## Abstract

The oomycete pathogen *Phytophthora cactorum* causes crown rot, a major disease of cultivated strawberry. We report the draft genome of *P. cactorum* isolate 10300, isolated from symptomatic *Fragaria x ananassa* tissue. Our analysis revealed that there are a large number of genes encoding putative secreted effectors in the genome, including nearly 200 RxLR domain containing effectors, 77 Crinklers (CRN) grouped into 38 families, and numerous apoplastic effectors, such as phytotoxins (PcF proteins) and necrosis inducing proteins. As in other *Phytophthora* species, the genomic environment of many RxLR and CRN genes differed from core eukaryotic genes, a hallmark of the two-speed genome. We found genes homologous to known *Phytophthora infestans* avirulence genes including *Avr1*, *Avr3b*, *Avr4*, *Avrblb1* and *AvrSmira2* indicating effector sequence conservation between *Phytophthora* species of clade 1a and clade 1c. The reported *P. cactorum* genome sequence and associated annotations represent a comprehensive resource for avirulence gene discovery in other *Phytophthora* species from clade 1 and, will facilitate effector informed breeding strategies in other crops.

## Introduction

The oomycetes are a diverse class of eukaryotic microorganisms that include pathogens of plants, animals and fungi [1]. The causal agents of plant diseases are well represented in this phylogenetic class, with over 60% of known oomycetes characterised as plant pathogens [2]. Of these, the *Phytophthora* genus is responsible for some of the most economically and culturally significant diseases, including potato late blight caused by the pathogen *Phytophthora infestans*, stem rot of soybean caused by *Phytophthora sojae*, Sudden Oak Death caused by *Phytophthora ramorum* and blight of peppers and cucurbits caused by *Phytophthora capsici* [3].

[grants/grants/AwardDetails.aspx?FundingReference=BB/K017071/2](https://doi.org/10.1371/journal.pone.0202305). MB, EL and sequencing costs were financed by a strategic NIBIO project (basic funding).

**Competing interests:** The authors have declared that no competing interests exist.

The hemi-biotrophic oomycete pathogen *Phytophthora cactorum* (Lebert and Cohn) was identified as the causal agent of strawberry crown rot disease in 1952 [4] and is now considered a major disease of strawberry in temperate regions, leading to plant losses of up to 40% [5]. *P. cactorum* is homothallic and produces oospores (resting spores) in diseased plant tissue. These can persist in the soil for many years and are an important source of infection in field production systems. *P. cactorum* is also a problem in the propagation of plants, risking rapid spread of the disease upon distribution [6]. Chemical control via soil fumigation with chloropicrin, 1,3-dichloropropene, dazomet and methyl bromide have proved effective in management of the pathogen [7]. However, the phasing out of chemical fumigants in accordance with stricter European regulations (e.g. 91/414/EEC), has led to increased incidence of historically well-controlled soilborne diseases. This has elevated the importance of integrating disease resistance into modern breeding germplasm. However, the functionality and durability of resistance is determined by pathogen encoded secreted effector proteins that can alter plant processes to aid infection [8]. Genome sequencing of *Phytophthora* spp. pathogens and subsequent functional characterisation of putative effector candidates from predicted gene models has provided a framework for study of *Phytophthora* diseases [9–11].

For *P. infestans*, characterisation of effector genes, including study of their interaction with host resistance genes [12–17] has provided information about the durability of deploying specific potato resistance genes. Similar suites of resistance genes have been identified against the soybean pathogen *P. sojae*, with fourteen major resistance genes at eight genomic loci determining a race structure within *P. sojae* [18]. This highlights the importance of understanding pathogen populations in the field and the associated genetic variation in effector complements. In contrast to these *Phytophthora* pathosystems, strawberry resistance to *P. cactorum* appears to be quantitative [19–22], and no race structure has been reported to date. As such, resistance is not determined by a single gene-for-gene recognition, as often associated with RxLR effectors [23]. In soybean, quantitative (partial) resistance is observed alongside race-specific resistance and has been linked to the accumulation of PR1a (a matrix metalloproteinase), a basic peroxidase and a  $\beta$ -1,3-endoglucanase at the inoculation site [24] and to the accumulation of suberin in the roots [25]. For this reason, a range of effector candidates need to be considered when studying quantitative resistance in the strawberry pathosystem.

*Phytophthora* produce apoplastic effectors that are secreted to the extracellular space of the host and cytoplasmic effectors that are translocated to the host cytoplasm or intracellular compartments. Cytoplasmic RxLR's are typified by an N-terminal signal peptide sequence allowing secretion of the protein, followed by an RxLR-EER motif that may be cleaved prior to secretion [26], and a variable C-terminal domain, often containing WY domains [27]. RxLR effectors typically modulate host defense by suppressing host cell death [23]. The recognition of the RxLR (arginine, any amino-acid, leucine, arginine) class of effectors is mediated by plant resistance proteins, most often (but not exclusively) NB-LRR containing genes [13,16, 28].

Another major class of cytoplasmic effectors in *Phytophthora* pathogens are the Crinklers (CRN, for CRinkling and Necrosis), named due to the response observed when *P. infestans* CRNs were ectopically expressed in plants [29]. CRNs have been shown to promote Pattern-Triggered Immunity (PTI), a process that is suppressed by RxLR effectors, indicating their functions may be associated with the necrotrophic stage of a hemi-biotrophic lifecycle [30–32]. Resistance is yet to be shown to this family of effectors but evidence has been presented for a heightened resistance response in *Nicotiana benthamiana* when infected with a *P. sojae* mutant overexpressing *PsCRN161* and in tomato plants infected with PVX vector containing *P. infestans crn2* [29,33]. CRNs characteristically possess an N-terminal LxLFLAK-motif connected with translocation, a DWL domain, the conserved recombination HVLVVVP-motif C-terminal domain. In some cases a DI domain is present between the LFLAK and DWL domain

[9,32]. Functional studies have shown the LFLAK domain to be involved in entry into the host cell, following this CRNs target host nuclear processes, but the mechanisms of trafficking into the nucleus, remain unknown [34]. Interestingly, most CRNs effectors do not have predicted signal peptides or if they have, these exhibit lower SignalP scores (HMM models) compared to the RXLRs proteins, the other class of host-translocated effectors. These weak *in silico* predictions of signal peptides in CRNs proteins could be due to these CRNs being non-functional or due to non-classical methods of secretion from the pathogen [34].

A diverse range of other secreted effectors are deployed during infection by *Phytophthora* spp. including plant cell wall degrading enzymes, protease inhibitors and phytotoxins of the PcF Toxin Family [35–37]. Furthermore, secreted non-effector proteins have been implicated in triggering HR in non-host species, such as elicitor INF1 [38]. Elicitors are secreted sterol binding and carrier proteins, an essential protein family for *Phytophthora* spp., which are unable to produce sterols themselves due to an inability to produce oxidosqualene [39,40].

With this work we aim to develop new genetic resources tools for the study of *Phytophthora* crown rot disease on cultivated strawberry including the first strawberry pathogen genome for *P. cactorum*, as well as the identification of candidate effectors from apoplastic and cytoplasmic families. *P. cactorum* pathogen has a diverse host range, infecting over 200 plant species [41]. This includes beech, for which a draft genome assembly was recently released [42]. New approaches to identify *Phytophthora* CRNs are described, including their use to identify novel CRN families in *P. cactorum*, as well as highlighting additional CRNs in reference *Phytophthora* spp. genomes. These data represent valuable new resources for study of host adaptation within *P. cactorum* and enable the study of effector complements within *P. cactorum* and their comparison to Clade 1 *Phytophthora* spp. *P. infestans* and *P. parasitica* as well as the more distant species *P. sojae* and *P. capsici* [43].

## Materials and methods

### Pathogen isolate

We sequenced the genome of *P. cactorum* Bioforsk isolate ID number 10300, isolated from symptomatic *Fragaria x ananassa* from Ås, Norway in 2006. Routine culturing was performed on V8 media at 20°C. Genomic DNA was extracted using the OmniPrep™ kit for High Quality Genomic DNA Extraction, following the manufacturer's protocol, using mycelium, cultured in liquid Plich medium.

### Pathogen sequencing and genome assembly

Genomic libraries from *P. cactorum* were prepared for Illumina short read sequencing with insert sizes of 300 bp, 1 kb and 5 kb. Libraries with inserts of 300 bp and 1 kb were prepared using Illumina Truseq LT (FC-121-2001), whereas 5 kb mate-pair genomic libraries were prepared using Nextera Mate Pair gel-plus and gel-free protocols. Illumina sequencing was performed on the libraries using 2 x 75 bp reads for 300 bp and 1 kb insert libraries and 2 x 300 bp reads for the 5 kb insert library. Sequencing resulted in 42.86, 57.76 and 10.84 million reads from the 300 bp, 1 kb and 5 kb insert libraries, respectively. Removal of low quality and adapter sequences using fastq-mcf, resulted in 41.29, 51.15 and 4.17 million reads from the 300 bp, 1 kb and 5 kb insert libraries, respectively.

*De-novo* genome assembly was performed using ABySS software version 1.3.7 [44], using a kmer length of 53 bp. Contigs shorter than 500 bp were discarded and assembly statistics of remaining contigs were summarised using QUASt software version 3.0 [45]. BUSCO software version 3.0.2 was used to assess the completeness of the assembly using the associated dataset of 303 core Eukaryotic genes as database for BUSCO analyses [46]. RepeatModeler software

version 1.0.8 and RepeatMasker software version 4.04 were used to identify repetitive elements and low complexity regions within the genome assembly (available at: <http://repeatmasker.org>).

### Gene models and ORF prediction

Gene prediction was performed on the softmasked *P. cactorum* genome using BRAKER1 software version 2 [47], a pipeline for automated training and gene prediction of AUGUSTUS version 3 [48]. Evidence for gene models was generated using publically available *P. cactorum* RNAseq reads [49], which were downloaded and aligned to *P. cactorum* assembly using STAR software version 2.5.3a [50]. Gene models were also called using CodingQuarry software version 2.0 [51], which was run using the “pathogen” flag parameter. CodingQuarry gene models were used to supplement BRAKER gene models, when individual CodingQuarry gene models were predicted in intergenic regions between Braker gene models.

Gene models were also supplemented with additional effector candidates from open reading frames (ORFs) located in intergenic regions of Braker and CodingQuarry genes. In addition, ORFs were predicted by translating sequences following all start codons in the genome until a stop codon or the end of the contig was reached. ORFs were predicted from sequences translating to between 50 and 250 aa in length and not predicted from sequence containing any N's. All ORFs encoding proteins were screened for secretion signals followed by RxLR and Crinkler effector motif (as described below) and of those testing positive, those present in intergenic regions were incorporated into gene models.

### Functional annotation of gene models

Draft functional annotations were determined for gene models using InterProScan-5.18–57.0 [52] and through identifying homology between predicted proteins and those contained in the SwissProt database [53] using BLASTP (E-value  $> 1 \times 10^{-100}$ ) [54]. Homology was identified between predicted gene coding sequence and the Pathogen-Host Interactions database (PHI-base; [www.phi-base.org/](http://www.phi-base.org/)) [55] using BLASTX (E-value  $> 1 \times 10^{-30}$ ). Homology was also identified against a set of 50 previously characterised oomycete effectors / avirulence genes using BLASTN (E-value  $> 1 \times 10^{-30}$ ). Functional annotation also identified the Carbohydrate-Active enZyme (CAZyme) encoding genes of *P. cactorum*. This was done using dbCAN [56] and using the CAZyme database classification [57].

Genes encoding putative secreted proteins were identified through prediction of signal peptides using SignalP software versions 2.0, 3.0 and 4.1 [58]. Use of SignalP v2.0, as well as limiting secreted proteins to those with HMM (Hidden Markov Model) scores greater than 0.9 and with cleavage sites between the 10<sup>th</sup> and 40<sup>th</sup> amino acid, was consistent with previous RxLR prediction methodologies [59,60]. Transmembrane proteins and membrane anchored proteins were identified using TMHMM version 2.0 and the GPI-SOM web-server respectively [61,62]. Proteins were considered as ‘putatively secreted’ if they tested positive for a secretion signal using SignalP and lacked a transmembrane domain or membrane anchor signal. Additionally, Phobius software version 1.01 was used to screen proteins for secretion signals missed by SignalP [63]. Proteins containing transmembrane domains or GPI anchored proteins were not excluded from the Crinkler and RxLR effector annotation pipelines discussed below.

### Crinkler effector identification

HMM models for CRN prediction were trained from CRN effectors predicted for *P. infestans*, *P. sojae*, *P. ramorum* and *P. capsici* [9,32]. A HMM model training set of 271 CRNs were selected from 315 described CRNs from *P. infestans*, *P. sojae* and *P. ramorum* [9], with CRNs excluded that lacked characteristic LFLAK or HVLVVP motifs from the LFLAK or DWL domains or contained

ambiguous sites ('X's) in their sequence. Similarly, 65 of 84 described *P. capsici* CRNs were used in the training set [32]. These remaining proteins were considered to represent high confidence CRNs. Alignment of these sequences allowed training of a model to the LFLAK domain (from the conserved 'MV' to 'LFLAK' motifs and a second model to the DWL domain (from the conserved 'WL' to the 'HVLVVVP' motifs). Putative CRNs were identified in predicted proteomes and translated ORFs by HMM searches using both LFLAK and DWL HMM models. Sequences required a HMM score greater than 0 for both models to be considered a putative CRN.

All predicted ORFs from the *P. cactorum* genome were screened using the trained LFLAK and DWL HMM models. Those ORFs with an HMM score greater than 0 for both FLAK and DWL HMM models were retained. As some of these ORFs were found to overlap, redundancy was removed from the dataset by retaining only the ORFs with the greatest LFLAK domain HMM score. The putative CRN ORFs located in intergenic regions of Braker / CodingQuarry gene models were integrated into the final set of gene models.

### RxLR effector identification

Motif and HMM based approaches were used to predict genes encoding RxLR proteins in *P. cactorum* and reference *Phytophthora* spp. genomes. Motif based prediction was based upon previous N-terminal RxLR identification pipelines [29]. Secreted proteins were considered putative RxLRs if an RxLR motif was present up to 100 aa downstream of the signal peptide cleavage point and the protein carried an EER motif within 40 aa downstream of the RxLR position. EER motifs were searched for using the Python regular expression ([ED][ED]+[KR]).

Heuristic based methods for RxLR prediction were used to complement RxLR prediction based upon motif presence. A previously described RxLR HMM model was used to statistically assess secreted proteins for the presence of N-terminal RxLR-like regions [64]. Hits with an HMM score greater than 0 were considered putative RxLR proteins.

All predicted ORFs carrying a secretion signal in the *P. cactorum* genome were screened for RxLR motifs and homology to HMM models. As some predicted ORFs were found to overlap one another, redundancy was removed from the dataset retaining only the ORF with the greatest SignalP HMM score. Those RxLR-containing ORFs located in intergenic regions of Braker / CodingQuarry gene models were integrated into the final set of gene models. All RxLR candidates were searched for presence of C-terminal WY-domains using a previously described HMM model [27].

### Gene orthology analysis

Ortholog identification was performed using OrthoFinder software version 1.1.10 [65] on all *P. cactorum* isolate 10300 predicted proteins and the proteomes of publically available *Phytophthora* species *P. infestans*, *P. parasitica*, *P. capsici* and *P. sojae*. Venn diagrams were plotted using the R package software version 3.5.2 VennDiagram package version 1.6.20 [66]. Further clustering was performed on the combined set of CRN effector proteins from *P. cactorum*, *P. infestans*, *P. parasitica*, *P. capsici* and *P. sojae* using OrthoMCL software version 2.0.9 [67], with the inflation value set to 5 in order to increase resolution within groups.

## Results

### Pathogen genome assembly

*De-novo* genome assembly using ABySS [44] generated a 59.3 Mb assembly in 4,623 contigs, with an N50 value of 56.3 kb (Table 1). Total assembly size was smaller than the other available *Phytophthora* spp. assemblies including *Phytophthora* clade 1c relatives *P. parasitica* and *P.*

**Table 1. Assembly and gene prediction statistics for the *Phytophthora cactorum* genome, with reference to publically available *Phytophthora* spp. genomes from Clades 1, 2 and 7 [43].** Number of core eukaryotic genes (CEGs) identified as complete and present in a single copy are shown for each genome/set of gene models, as determined by BUSCO.

Species	<i>P. cactorum</i>	<i>P. parasitica</i>	<i>P. infestans</i>	<i>P. capsici</i>	<i>P. sojae</i>
Phylogenetic Clade	1a	1b	1c	2	7
Strain	10300	INRA-310	T30-4	LT1534	P6497
Assembly size (Mb)	59.3	82.4	228.5	64	83
Number of contigs	4623	708	4921	917	83
Number of contigs (>1 kb)	2913	708	4598	917	83
Largest contig (kb)	301	4,724	6,928	2,170	13,391
N50 (kb)	56.3	888	1,589	706	7,609
N's per 100 kb	4006	34,613	16,806	12,466	3959
Repeatmasked (Mb)	10.8 (18%)	7.0 (8%)	152.1 (67%)	13.6 (21%)	23.7 (29%)
CEGs in the assembly	274 (90%)	271 (89%)	255 (84%)	269 (89%)	270 (89%)
Predicted genes	23,884	20,822	17,787	19,805	26,584
CEGs in gene models	272 (89%)	271 (89%)	257 (85%)	261 (89%)	262 (86%)

<https://doi.org/10.1371/journal.pone.0202305.t001>

*infestans* assemblies, but was found to contain a similar or greater gene space within the assembly, with BUSCO identifying 283 of 303 core eukaryotic genes (CEGs). Of these CEGs, 274 were present in a single copy within the assembly. The *P. cactorum* genome was found to be repeat-rich, with RepeatModeler and RepeatMasker identifying 18% of the genome as repetitive or low complexity regions. This level of repetitive content was considerably lower than observed in *P. infestans*, but comparable to *P. capsici* that has a similarly sized genome of 64 Mb. Meaningful comparisons of repeat content could not be made between the *P. cactorum* and *P. parasitica* genomes as the scaffolded *P. parasitica* assembly contained a high percentage of N's (Table 1). A total of 23,884 genes encoding 24,189 proteins were predicted from the *P. cactorum* genome with 21,410 genes predicted from the Braker1 pipeline [47], 2,434 additional genes from CodingQuary [51], and a further 40 coding genes from intergenic ORFs identified as putative secreted RxLR or CRN effectors. The number of predicted genes reported in *Phytophthora* spp. shows considerable variation between studies, however *P. cactorum* gene models contained the greatest number of complete single copy CEGs among the assessed *Phytophthora* spp., indicating good representation of gene space within gene models (Table 1).

### Orthology analysis

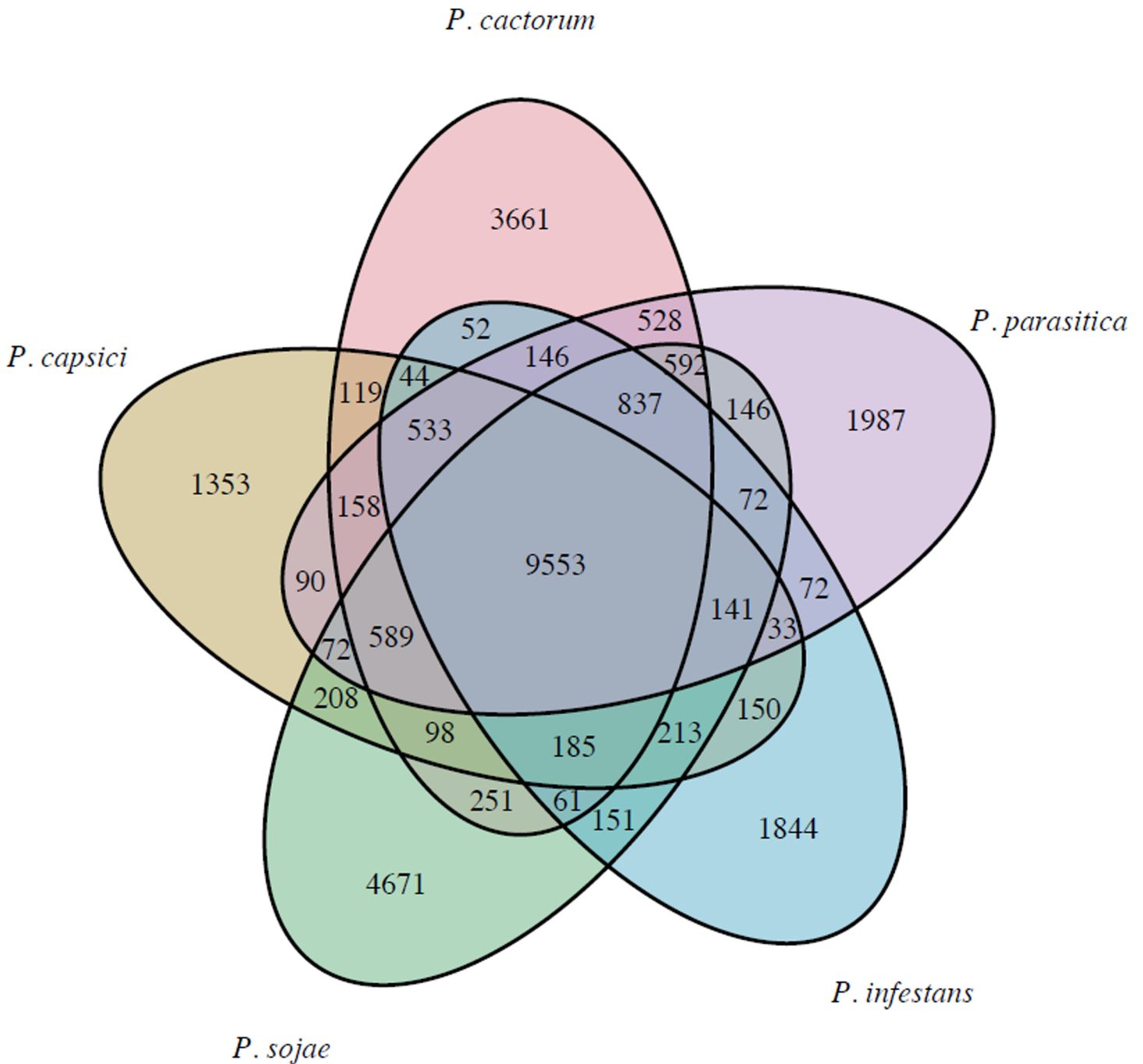
Clustering of predicted proteins from the five *Phytophthora* spp. using OrthoFinder, resulted in 15,162 orthogroups containing 95,739 proteins (87.7% of the total). A total of 20,157 (84%) of predicted *P. cactorum* proteins had identified orthologs in other *Phytophthora* spp. Of these, 9,553 orthogroups contained proteins from all five species, with 6,767 orthogroups consisting of a single protein from each species (Fig 1).

### Intergenic distance

Intergenic distance was determined for each gene, by counting the number of bp to the nearest gene in 5-prime and 3-prime directions. Genes that were on the end of a contig and therefore did not have a neighbouring gene up- or down-stream were discarded from this analysis.

### Functional annotation and secretome prediction

Genomic locations of *P. cactorum* 10300 gene models, their orthology assignment and predicted functional annotations are summarised in S1 Table. A total of 2,234 genes encoded



**Fig 1. Number of shared and unique ortholog groups between *Phytophthora* spp.** Orthogroups determined from clustering 109,187 proteins from *P. cactorum*, *P. parasitica*, *P. infestans*, *P. capsici* and *P. sojae*.

<https://doi.org/10.1371/journal.pone.0202305.g001>

putatively secreted proteins. The number of predicted genes encoding secreted MAMPs, apoplast effectors and cytoplasmic effectors are summarised in Table 2 and discussed below.

### Microbe associated molecular pattern (MAMP) genes

**Sterol-binding proteins.** *Phytophthora* spp. lack the ability to synthesize sterols and are reliant on assimilation from the environment. Secreted sterol binding proteins are known

**Table 2. Total number of predicted effector gene candidates in *Phytophthora cactorum* 10300 and genes associated with triggering plant basal defense (microbe associated molecular patterns, MAMPs).** Numbers of genes shown relate to genes encoding predicted secreted proteins.

Category	Family	Number of proteins
MAMP		
	Sterol binding proteins	47
	Transglutaminase proteins	15
Apoplastic effectors		
	Secreted CAZymes	282
	Protease inhibitors (glucanase)	2
	Phytotoxins	2
	Necrosis inducing proteins	24
	Cutinases	4
	Protease inhibitors (kazal)	14
	Protease inhibitors (cathepsin)	3
Cytoplasmic effectors		
	Crinklers	77
	RxLRs	199

<https://doi.org/10.1371/journal.pone.0202305.t002>

Microbe-Associated Molecular Patterns (MAMPs), triggering host recognition. For this reason, they are also referred to as “elicitors”. A total of 66 genes possessed an elicitin domain (IPR002200), of which 47 were predicted as secreted. These genes showed high levels of local duplication, with 41 of the 66 genes in 11 elicitin gene clusters.

**Transglutaminase proteins.** The *P. sojae* cell wall glycoprotein GP42 is an elicitor of host defence and is functionally characterized as a Ca<sup>2+</sup>-dependant transglutaminase [68]. Recognition of the protein by plant hosts is lost upon mutation of the transglutaminase domain, indicating its importance for recognition. A total of 23 *P. cactorum* genes were predicted to encode transglutanimase domains (IPR032048). These were distributed through 10 orthogroups, with 13 proteins contained in a single orthogroup (OG0000097). Blast searches identified 19 *P. cactorum* genes with homology to *P. sojae* GP42 (S2 Table), each of which was identified by domain searches. Of the 23 proteins carrying transglutaminase domains, 15 were predicted to be secreted.

### Apoplastic effectors

**Carbohydrate active enzymes (CAZymes).** CAZymes play a direct role in pathogenicity, contributing to plant cell wall degradation. A total of 696 transcripts encoding CAZymes were identified in the *P. cactorum* 10300 genome, of which 352 were predicted as carrying an N-terminal signal peptide and 282 were predicted as secreted (removing those with transmembrane and GPI anchors domains). These secreted CAZymes were distributed through glycerinaldehyde hydrolases (GH), carbohydrate binding molecules (CBM), auxiliary activity (AA), carbohydrate esterase (CE) and pectin lyase (PL) and glycosyl transferases (GT) families containing 172, 22, 6, 24, 37 and 21 proteins respectively.

The profile of *P. cactorum* cell wall degrading enzymes was investigated through further study of GH, CBM, AA, CE and PL families (Table 3). Substrate specificity was not further investigated within the GT proteins due to wide polyspecificity (multiple substrates associated with the same GT family) within this group. Cell wall degrading enzymes can be summarized by functions, targeting cellulose, hemicellulose or pectin. Cellulase activity is represented in



**Table 3. Profile of secreted *Phytophthora cactorum* Carbohydrate-Active enZymes (CAZymes) from glycerinaldehyde hydrolase (GH), carbohydrate binding molecules (CBM), auxiliary activity (AA), carbohydrate esterase (CE) and pectin lyase (PL) families, as identified by dbCAN. Numbers are shown for total numbers of N-terminal signal peptide containing proteins, and those considered putative secreted proteins, which lack transmembrane signals of membrane anchors. Target substrates of for each family is shown.**

CAZY family	Substrate	Signal peptide	Secreted proteins
GH17	$\beta$ -1,3-glucans	21	21
GH3	cellulose, hemicellulose (xyloglucans), pectin (RGI), AGPs	19	16
GH28	pectin (HG)	15	15
GH16	hemicellulose (xyloglucans), $\beta$ -1,3-glucans	18	12
GH81	pectin (RGI)	12	12
GH30	cellulose, hemicellulose (xyloglucans), pectin (RGI), AGPs	12	11
GH12	cellulose, hemicellulose (xyloglucans)	12	9
GH72	$\beta$ -1,3-glucans	10	8
GH1	cellulose, hemicellulose (xyloglucans), pectin (RGI)	10	8
GH5	cellulose, hemicellulose (xyloglucans, galactomannans), $\beta$ -1,3-glucans	8	8
GH6	cellulose	7	6
GH78	pectin (RGI)	6	6
GH43	hemicellulose (xylans), pectin, AGP	5	4
GH31	starch, hemicellulose (xyloglucans)	5	4
GH131	$\beta$ -1,3-glucans, hemicellulose ( $\beta$ -1,4-glucans)	5	4
GH7	cellulose	4	4
GH53	pectin (RGI)	4	3
GH32	sucrose	3	3
GH19	N-linked oligosaccharides	3	3
GH10	hemicellulose (xylans)	3	3
GH17, CBM13	$\beta$ -1,3-glucans	3	2
GH18	N-linked oligosaccharides	2	2
GH54	pectin (RGI)	1	1
GH47	N-linked oligosaccharides	1	1
GH38	N-linked oligosaccharides	1	1
GH2	hemicellulose (mannans), glycoproteins (mannans)	1	1
GH16, GT48	hemicellulose (xyloglucans), $\beta$ -1,3-glucans	1	1
GH13	starch	1	1
GH105	pectin (RGI)	1	1
GH31, CBM25	starch	1	1
GH89	N-linked oligosaccharides	2	0
GH114	$\alpha$ -1,4-polygalactosamine	1	0
GH5, CBM43	$\beta$ -1,3-glucans	1	0
CBM63	cellulose	11	9
CBM1	cellulose	10	9
CBM47	fucose binding	2	1
CBM9	hemicellulose (xylans)	1	1
CBM36	xylanase	1	1
CBM32	galactose, PGA and $\beta$ -galactosyl- $\beta$ -1,4-GlcNAc	1	1
CBM38	inulin binding	1	0
AA2	lignin	4	3
AA8	cellulose	3	1
AA10	cellulose	2	1
AA9	cellulose	1	1

(Continued)

Table 3. (Continued)

CAZY family	Substrate	Signal peptide	Secreted proteins
AA7	Glycolate oxidase	2	0
CE8	pectin (HG)	9	8
CE1	hemicellulose	8	4
CE10	non-carbohydrate substrates	6	3
CE13	pectin (HG)	5	3
CE5	hemicellulose	3	3
CE12	pectin (HG, RGI)	2	2
CE3	hemicellulose	1	1
CE4	hemicellulose, N-linked oligosaccharides	1	0
PL3	pectin (HG, RGI)	21	17
PL1	pectin (HG)	19	16
PL4	pectin (RGI)	4	4

HG = homogalacturonan, RGI = rhamnogalacturonan I; GlcNAc = N-acetylglucosamine.

<https://doi.org/10.1371/journal.pone.0202305.t003>

seven GH families, two CBM families and three AA families. Cellulases are well represented in the CBM compliments of *P. parasitica*, *P. ramorum* and *P. sojae*, where CBM1 and CMB63 represented the two largest groups of CBMs. This was also true for *P. cactorum*, where CBM63 and CBM1 proteins represented 81% of the putatively secreted CBM molecules. This is in contrast to fungal necrotrophs which typically possess 1–3 CBM3 proteins [35]. In fungi, CBM1 and CBM63 domains are predominantly accompanied by additional modules [35], however none of the CBM63 or CBM1 proteins in *P. parasitica* are accompanied by other catalytic modules [35]. This was also true of *P. cactorum* CBM63 or CBM1 CAZymes. Hemicellulose targeting secreted CAZymes were represented in 12 GH families, one CBM family and four CBM families. The *P. cactorum* genome encodes large numbers of proteins involved in pectin modification, including GH groups GH28 and GH81 representing the third and fifth most abundant GH groups (15 and 12 proteins), CE8 representing the most abundant CE group (8 proteins) and 37 proteins from PL families PL3, PL1 and PL4. *Phytophthora* spp. are reported to carry expanded pectin targeting CDWE in comparison to fungi [35]. In total, 79 putatively secreted CWDE targeted pectin, which is comparable to the 86 predicted in *P. parasitica*, and in contrast to fungi, which typically have less than 20 PL proteins [35].

Secreted enzymes targeting  $\beta$ -1,3-glucan may function in breakdown of callose, as deposited by the host upon triggering of basal defense.  $\beta$ -1,3-glucans are also found in the pathogen, being present in the oomycete cell wall where they act as MAMPs triggering plant basal defence [69]. Reflecting this, *P. cactorum* carried a large number of genes (31) encoding putatively secreted proteins from five different families targeting  $\beta$ -1,3-glucan. Notably, 21 genes encoded GH17 proteins, which was the most abundant CAZyme family.

**Glucanase inhibitors.** Non-CAZyme proteins are involved in preventing host recognition of *Phytophthora*  $\beta$ -1,3-glucans. Glucanase inhibitor proteins (GIPs) are serine proteases that inhibit degradation of  $\beta$ -1,3/1,6-glucans in the pathogen cell wall and/or the release of defence-eliciting molecules by host endoglucanases [70]. These serine proteases contain a domain that shows homology to the chymotrypsin class of serine proteases, however they lack proteolytic activity and as such belong to a broader class of proteins called serine protease homologs [71]. A total of 34 *P. cactorum* genes were predicted to encode proteins with chymotrypsin domains (IPR001314), with 24 of these predicted as secreted and 28 as homologs of GIP

proteins from *P. infestans* and *P. sojae*. Three of the *P. cactorum* proteins were members of a single orthogroup containing *P. infestans* GIP proteins (PITG\_13636, PITG\_21456), of which two were predicted as secreted and therefore represent high-confidence glucanase inhibitor candidates.

**Phytotoxins.** The PcF toxin family was first described from *P. cactorum* [72], and in line with this BLAST searches identified g2968.t1 as homologous to PcF (NCBI accession: AF354650.1). This gene was a member of an orthogroup with two members from *P. infestans*, one member from *P. parasitica* and two members from *P. capsici*. InterProScan annotation identified two additional phytotoxin candidates (g10773.t1, g16798.t1) carrying the PcF domain (Pfam: PF09461) in addition to g2968.t1. Each of the three identified genes encoded a protein with a N-terminal secretion signal but g16782.t1 was also predicted to encode a transmembrane domain.

**Necrosis inducing proteins.** Necrosis inducing proteins (NLPs) are produced by bacterial, fungal and oomycete plant pathogens [73]. These proteins are associated with the transition from biotrophy to necrotrophy in *Phytophthora* spp. and act by triggering cell death [74]. These proteins may also stimulate immune responses in the host. The repertoire of NLP proteins encoding genes in *P. cactorum* was 43 proteins carrying NLP-like domains (PF05630, IPR008701) of which 24 were predicted as secreted. 30 of these proteins as NPP1 homologs in PHLbase, of which 21 were predicted as secreted. 25 of the 43 genes also showed homology to assembled *P. cactorum* transcripts from previous work. The 43 proteins were distributed through 16 orthogroups, including all 13 members of orthogroup 75 and all 12 members of orthogroup 12. Alignment of all proteins in the 16 NLP orthogroups showed that these proteins represent Type1 NLPs, through conservation of two cytosine sites (alignment positions 624 and 661 in [S1 Data](#)).

**Cutinases.** In addition to the plant cell wall, cutin acts as a barrier to host penetration by plant pathogens. Pathogens often employ methods to circumvent this barrier such as colonisation via stomata or through wounds. *P. cactorum* is considered to infect via the roots of strawberry, however may cause above-ground symptoms such as strawberry fruit infection, known as leather rot. In total, seven genes were annotated as cutinase genes (PF01083), and four of these putative cutinases were predicted as secreted. Interestingly, three of the four secreted cutinases and a non-secreted cutinase (g10526, g10527, g10528, g10530) were clustered in a 5 Kb region of the genome. Two of these genes belonged to the same orthogroup, which showed an expansion of genes in *P. sojae* (14 genes), but similar numbers in the other *Phytophthora* spp. (3–4 genes). The other two *P. cactorum* genes were present in single-gene orthogroups unique to *P. cactorum*. Closer investigation revealed that one of these two genes was truncated, and the other incomplete due to being located on the end of the contig.

**Protease inhibitors.** Plant hosts secrete proteases into the apoplastic space to degrade pathogen-secreted effectors. As such, oomycetes are known to secrete protease inhibitors to counteract these defenses [36]. A total of 22 genes encoding Kazal-type protease inhibitors (IPR002350) were identified in *P. cactorum* gene models, with 14 of these predicted as secreted. It was noted that 12 of the 22 genes were located within 8 Kb of another Kazal-domain encoding gene, in clusters of two or three genes. Despite this, the 22 genes represented 18 different orthogroups, indicating historical duplication and divergence between these effector genes. A further four genes encoding proteins with cathepsin propeptide inhibitor domains (IPR013201) were identified, three of which were predicted as secreted. All were located on different contigs and were members of distinct orthogroups. A number of secreted cystatin-like cysteine protease inhibitors have been identified from *P. infestans* (EPIC1-EPIC4), including EPIC2B which has been shown to inhibit the tomato defence response through interaction with an apoplastic papain-like cysteine protease [75]. Three *P. cactorum* genes were predicted to encode secreted cystatin-like

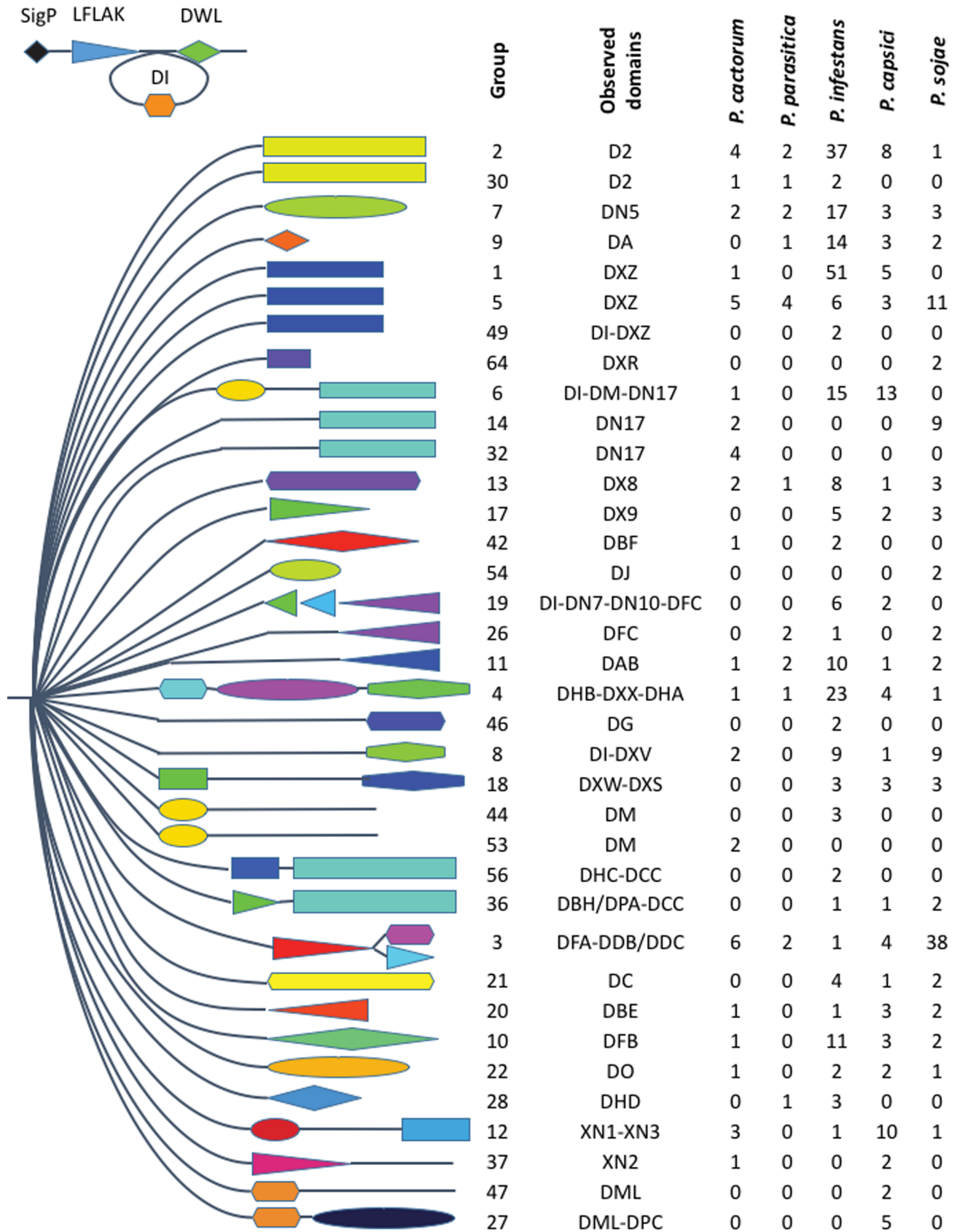
cysteine protease inhibitors, containing cystatin (IPR000010, IPR027214) or cystatin protease inhibitor (IPR018073, IPR020381) domains. These genes were in three orthogroups, each containing a single gene from *P. cactorum*. Blast searches identified the three genes as homologs of *EPIC1*, *EPIC3* and *EPIC4*.

## Cytoplasmic effectors

**Crinkler annotation.** A novel method of CRN prediction was developed based upon identification of the characteristic LFLAK and DWL domains. Trained Hmm models are provided in [S2 Data](#) and [S3 Data](#). Application of the LFLAK DWL models to *P. infestans* and *P. capsici* was used to validate the LFLAK-DWL approach. In total, 265 *P. infestans* and 175 *P. capsici* proteins were predicted encoding putative CRNs. Of the 194 proteins previously identified as CRNs in *P. infestans* [9], 35 were not identified by the LFLAK-DWL approach, each lacking the 'HVLVVVP' motif from the DWL domain. Similar results were observed for results from *P. capsici*, where 71 of the 84 previously identified CRNs were identified by the LFLAK-DWL approach [32], and the remaining 13 were found to contain ambiguous sites ('X's). Application of the LFLAK-DWL to reference gene models and ORFs allowed identification of 265 CRNs in *P. infestans*, 35 in *P. parasitica*, 114 in *P. capsici* and 159 in *P. sojae*, with 4, 98, 32 and 89 candidates identified from translated ORFs, respectively (Fasta sequences available in [S4 Data](#)).

Application of the developed LFLAK-DWL approach to *P. cactorum* identified a total of 77 putative CRN effector genes, with three of these identified from ORF gene models. Inspection of the *P. cactorum* CRN gene models showed that 17 (22%) were incomplete, lacking stop codons due to being located on the ends of contigs. This may reflect the modular structure and duplication of CRNs leading to difficulty in genome assembly of these regions. CRNs are known to be secreted from the host cell but often lack predictable secretion signals, with *e.g.* 58% of identified *P. capsici* CRNs lacking secretion signals [32]. We found similar results with 56% of *P. cactorum* CRNs lacking a signal peptide as predicted by SignalP 2, 3, 4 and Phobius software. Phobius was more sensitive than SignalP 2, 3 and 4, identifying signal peptides in 32 of the 77 CRNs as secreted, whereas the SignalP approaches predicted a combined total of 22 as secreted, with two that were not detected by Phobius.

The modular structure of CRNs and the diversity of CRN domains within *Phytophthora* spp. was further investigated using an orthology analysis on the total set of 650 predicted CRNs between the five studied species. Clustering using orthoMCL resulted in 73 groups of CRN proteins, with groups observed to separate by C-terminal domain ([Fig 2](#)). All of the 39 previously described C-terminal domains were identified within the clustered proteins, as well as the variable DI domain within the N-terminal region [9,32]. *P. cactorum* CRNs were present in groups representing 21 of these domains, whereas 14, 31, 31 and 27 domains were represented in groups containing *P. parasitica*, *P. infestans*, *P. capsici* and *P. sojae* CRNs. *P. infestans* showed signs of gene expansion in some groups including those encoding DXZ domains (59 *P. infestans* proteins vs 4–13 from other species), D2 domains (39 *P. infestans* proteins vs 1–8 from other species), DHB-DXX-DHA domains (23 *P. infestans* proteins vs 1–4 from other species) proteins. Similar expansion was not observed in *P. cactorum* CRN genes, with most populous groups representing DXZ, DN17 and DFA-DDB/DDC domains. Many proteins in *P. infestans* expanded orthogroups were identical to one another, indicating that CRN proteins are subject to frequent duplication, and as such the total numbers of CRNs observed in a genome is likely to be highly influenced by the quality of the genome assembly. An additional 137 predicted CRN proteins in 37 orthogroups did not contain any recognizable CRN domains.



**Fig 2. Clustering of *Phytophthora* spp. crinklers separates the proteins by their C-terminal domain.** All crinklers possess a conserved LFLAK and DWL domain, with some also possessing a DI domain in the N-terminal region. Crinklers proteins were observed to cluster by C-terminal domain as described in Haas (2009) and Stam (2013). The cluster (group) of proteins is shown along with observed domains and the number of *P. cactorum*, *P. parasitica*, *P. infestans*, *P. capsici* and *P. sojae* genes contained within each group.

<https://doi.org/10.1371/journal.pone.0202305.g002>

**RxLR identification.** A combined approach of regular expression searches for RxLR-EER motifs, as well as searches using HMM models identified 199 putative RxLR effectors in the *P. cactorum* assembly, with 162 of these predicted from predicted gene models and a further 37 from ORFs. Searches for WY domains found 92 WY-domain containing RxLRs. Functional annotation was largely absent for these proteins, but InterProScan annotations were present for ten proteins and a further five were predicted to be CAZymes (Table 4). Many of these domains have been associated with virulence in *Phytophthora* or other organisms [35,76–80]. This included three RxLRs with Nudix-hydrolase annotations, a domain present in Avr3b. Avr3b from *P. sojae* is expressed at early stages of infection and delivered into the host cell where it matures itself through recruitment of GGmCYP1, leading to suppression of effector triggered immunity [77,81]. Genes in ortholog groups containing *PiAvr3b* and other characterised RxLRs were identified (Table 5). *P. cactorum* carried genes in orthogroups containing *P. infestans* *Avr1*, *Avr3b*, *Avr4*, *Avr-blb1* and *Avr-Smira2*. *Avr1* is understood to manipulate basal defence through interaction with a plant exocyst subunit and thereby disturbing vesicle trafficking [82]. Two genes from *P. cactorum* were in the same orthogroup as *P. infestans* *Avr-blb1*, however one was truncated. Truncation has been observed in ~10% of *P. sojae* and *P. ramorum* RxLRs [83]. Furthermore, truncation leading to loss of function in *Avr4* has been shown to prevent host recognition, determining a race structure in *P. infestans* [14]. *Avr-blb1* is understood to interact with a lectin receptor kinase associated with the plasma membrane, leading to destabilising of the cell wall-plasma membrane to promote infection [76]. A total of 35 *P. cactorum* RxLR candidates were members of orthogroups containing a single gene from both *P. cactorum* and *P. infestans*. Similar orthology assignments could be made for 33 *P. cactorum* RxLR candidates and *P. sojae* genes. Characterisation of these core RxLRs will aid understanding of the fundamental infection strategy conserved between *Phytophthora* spp.

**Table 4. Functional annotations of *Phytophthora cactorum* RxLR candidates.** Orthogroup assignment shows conservation of these genes throughout *Phytophthora* spp. Numbers of genes in each orthogroup are shown for *P. cactorum* (Pcac), *P. parasitica* (Ppar), *P. infestans* (Pinf), *P. capsici* (Pcap) and *P. sojae* (Psoj).

RxLR gene ID	Orthogroup	Orthogroup contents	Notable annotations
g553.t1	OG0004128	Pcac(1):Pinf(1):Ppar(1):Pcap(1):Psoj(1)	Leucine-rich repeat domain (IPR032675)
g1729.t1	OG0004656	Pcac(1):Pinf(1):Ppar(1):Pcap(1):Psoj(1)	Conserved regions 1–4 of stealth proteins (PF17101, PF11380, PF17102 & PF17103)
g2445.t1	OG0004967	Pcac(1):Pinf(1):Ppar(1):Pcap(1):Psoj(1)	Ryanodine receptor domain (IPR003032)
g2934.t1	OG0001997	Pcac(2):Pinf(3):Ppar(2):Pcap(0):Psoj(0)	CAZY:GT44
g4805.t1	OG0005907	Pcac(1):Pinf(1):Ppar(1):Pcap(1):Psoj(1)	Concanavalin A-like lectin/glucanase domain (IPR013320)
g5243.t1	OG0011620	Pcac(1):Pinf(0):Ppar(1):Pcap(1):Psoj(1)	SMP-30/Gluconolactonase/LRE-like region (PF08450)
g7310.t1	OG0011769	Pcac(1):Pinf(1):Ppar(1):Pcap(0):Psoj(1)	RanBP2-type Zinc finger domain (IPR001876)
g8318.t1	OG0000314	Pcac(6):Pinf(5):Ppar(5):Pcap(5):Psoj(4)	NUDIX hydrolase domains (IPR000086)
g10092.t1	OG0000314	Pcac(6):Pinf(5):Ppar(5):Pcap(5):Psoj(4)	NUDIX hydrolase domains (IPR000086)
g12307.t1	OG0000363	Pcac(5):Pinf(3):Ppar(7):Pcap(3):Psoj(5)	Intradiol ring-cleavage dioxygenase domain (IPR000627)
g13307.t1	OG0000351	Pcac(5):Pinf(3):Ppar(10):Pcap(2):Psoj(3)	CAZY:GT54
g13922.t1	OG0000571	Pcac(3):Pinf(6):Ppar(3):Pcap(4):Psoj(1)	CAZY:CE2
g14748.t1	OG0016955	Pcac(1):Pinf(0):Ppar(0):Pcap(0):Psoj(0)	Cytochrome P450 domain (IPR001128)
g16698.t1	OG0000532	Pcac(3):Pinf(5):Ppar(5):Pcap(3):Psoj(2)	CAZY:GT44
g19791.t1	OG0012635	Pcac(1):Pinf(1):Ppar(1):Pcap(0):Psoj(1)	Lipid-binding start domain (IPR023393)
g23189.t1	OG0018641	Pcac(1):Pinf(0):Ppar(0):Pcap(0):Psoj(0)	CAZY:GT2, GT41

<https://doi.org/10.1371/journal.pone.0202305.t004>

**Table 5. *Phytophthora cactorum* genes in orthogroups shared with characterized *P. infestans* RxLR candidates.** Orthogroup assignment shows conservation of these genes throughout *Phytophthora* spp. Numbers of genes in each orthogroup are shown for *P. cactorum* (Pcac), *P. parasitica* (Ppar), *P. infestans* (Pinf), *P. capsici* (Pcap) and *P. sojae* (Psoj).

<i>P. cactorum</i> gene ID	Contig	<i>P. infestans</i> Avr gene	<i>P. infestans</i> gene ID	Orthogroup	Orthogroup contents	Notes
g15126.t1	contig_485	<i>Avr1</i>	PITG_16663	OG0000777	Pcac(2):Pinf(2):Ppar(4):Pcap(4):Psoj(2)	TLLR at RxLR motif location
g16706.t1	contig_608	<i>Avr1</i>	PITG_16663	OG0000777	Pcac(2):Pinf(2):Ppar(4):Pcap(4):Psoj(2)	
g5545.t1	contig_94	<i>Avr3b</i>	PITG_15732	OG0013112	Pcac(1):Pinf(1):Ppar(1):Pcap(0):Psoj(0)	NUDIX hydrolase domain (IPR000086)
g4951.t1	contig_80	<i>Avr4</i>	PITG_07387	OG0011587	Pcac(1):Pinf(1):Ppar(2):Pcap(0):Psoj(0)	
g6635.t1	contig_121	<i>Avrblb1</i>	PITG_21388	OG0001713	Pcac(2):Pinf(2):Ppar(4):Pcap(0):Psoj(0)	Truncated protein
g6663.t1	contig_121	<i>Avrblb1</i>	PITG_21388	OG0001713	Pcac(2):Pinf(2):Ppar(4):Pcap(0):Psoj(0)	
g15879.t1	contig_543	<i>AvrSmira2</i>	PITG_07558	OG0000427	Pcac(2):Pinf(4):Ppar(5):Pcap(3):Psoj(7)	
g18867.t1	contig_844	<i>AvrSmira2</i>	PITG_07558	OG0000427	Pcac(2):Pinf(4):Ppar(5):Pcap(3):Psoj(7)	

<https://doi.org/10.1371/journal.pone.0202305.t005>

Thirteen RxLR candidates lacked a recognisable EER motif and were not identified by the RxLR HMM model, but were identified by the presence of secretion signal, RxLR motif and WY domain. BLAST searches identified two of these genes as homologs to *P. infestans Avr-smira2* and a further four of these genes were identified as homologs to *P. sojae PSR2* and two as homologs to *Avh5*. Homologs to these characterised RxLR genes highlight the importance of using multiple sources of evidence in RxLR identification.

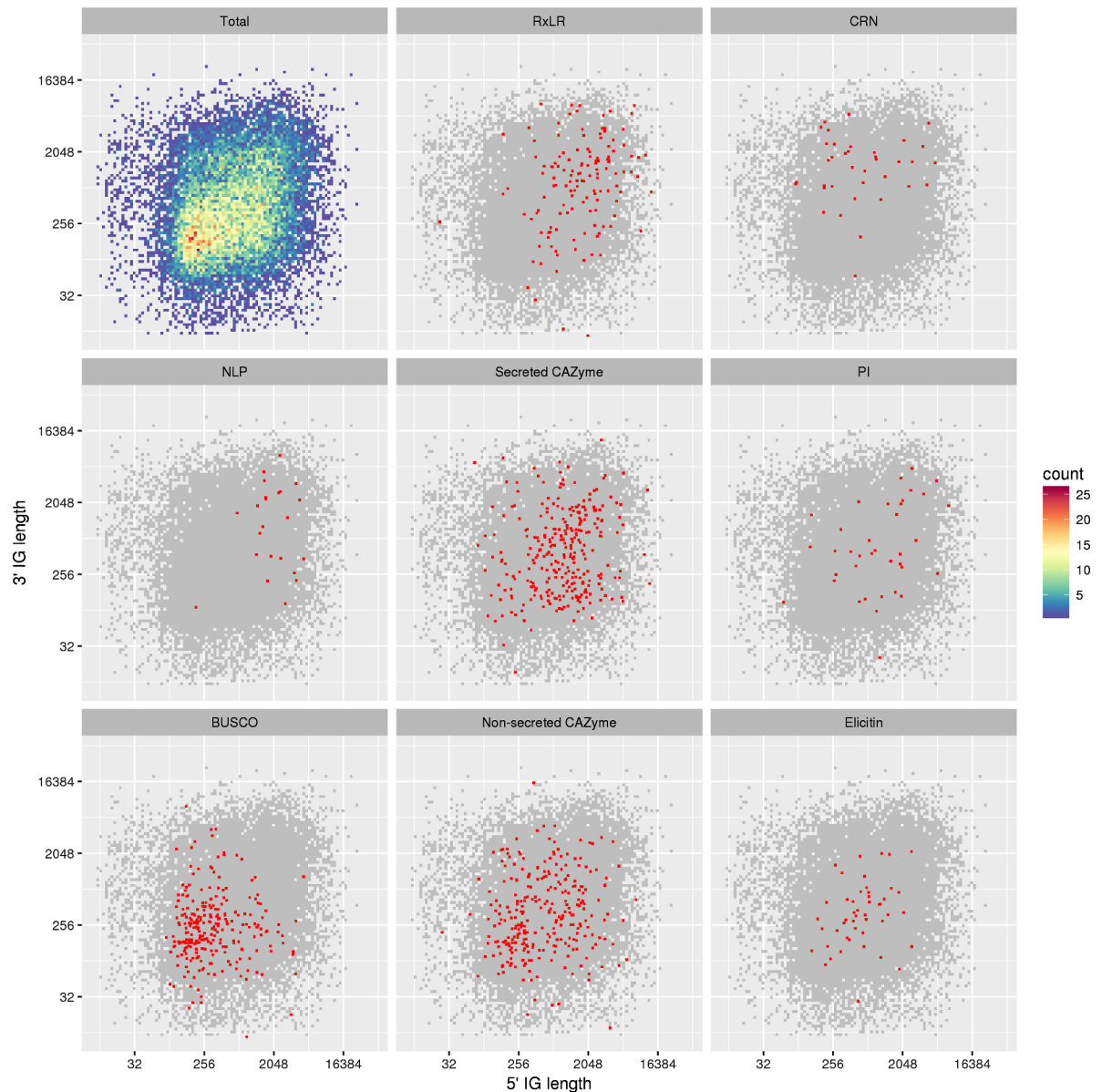
### Genomic distribution of *P. cactorum* effectors

Rapidly evolving RxLR and CRN genes are predominantly located in gene-sparse regions, with greater intergenic distances (IGDs) than core eukaryotic genes [9]. The 5' and 3' flanking distance between each *P. cactorum* gene and its neighbours were taken as measurements of local gene density (Fig 3), following exclusion of 5041 genes (21%) that neighbored a contig break (Table 6). Effector genes were located in gene sparse regions of the *P. cactorum* genome, with RxLR genes having greater mean 5' and 3' IGDs than observed for non-RxLR genes ( $p < 0.001$  and  $p < 0.001$ , respectively with 10,000 permutations). CRN genes were found to have mean 3' IGDs greater than that observed for non-CRN genes ( $p = 0.0148$ , with 10,000 permutations), but this was not the case for 5' regions. The larger in IGD in the 3' but not 5' region of CRN genes compared to the 5' region was further investigated by looking at functional annotations of the 5' neighbouring genes to CRNs. Fifteen of the 34 5' neighbours of CRN genes were found to have functional annotations, but no clear trend in gene function could be determined. However, not all effector candidates showed these patterns, with no significant difference observed in intergenic distance between protease inhibitors and neighboring genes ( $p > 0.05$ ). Secreted *P. cactorum* CAZymes proteins were found to have significantly greater 5' IG distance. Non-effector candidate elicitors had IGDs with no difference in distribution to all genes ( $p > 0.05$ ). Interestingly, putative non-secreted CAZymes were observed to have significantly shorter 5' and 3' IG distances than the total gene set ( $p = < 0.001$  and  $p = < 0.001$ , respectively with 10,000 permutations). This indicates that the forces driving genomic arrangement of regions containing RxLR and CRN cytoplasmic effector candidates and apoplasmic CAZyme effector candidates are distinct to those of other effector families in *P. cactorum*.

## Discussion

### A new genomic resource to study strawberry crown rot

*P. cactorum* is a persistent pathogen of strawberry and an economically significant pathogen of apple [41]. Genomic resources are available for these hosts [84–87], and recent work has



**Fig 3. Intergenic distance of cytoplasmic and apoplastic effectors as well as non-effector candidates.** Intergenic distance (5' and 3') of all *P. cactorum* 10300 genes is displayed in a density plot (Total) with scale bar indicating gene density within the plot. Additional plots highlight subsets of effector candidates within the distribution including RxLR and crinkler cytoplasmic effector candidates, secreted CAZymes, protease inhibitors and necrosis inducing protein (NLP) apoplastic effector candidates. Distribution of non-effector candidates is shown for conserved eukaryotic genes (BUSCO), non-secreted carbohydrate-active enzymes (CAZymes) and elicitors.

<https://doi.org/10.1371/journal.pone.0202305.g003>

identified resistance-associated QTL for cultivated strawberry [19]. Despite this, genomic resources for the pathogen are limited to identification of ESTs expressed during infection [88] and transcript expression during oospore germination characterized [49,89]. We report the sequencing, annotation and assembly of the *P. cactorum* genome, totalling 59 Mb, with a total of 23,884 predicted transcripts. The assembly was fragmented, in 4623 contigs, with 2913 over 1 kb. However, BUSCO statistics were indicative of a highly-complete assembly and detection of 89% of CEGs as present in a single copy within predicted gene models was greater than that observed from other *Phytophthora* spp. Assembly fragmentation was attributed to



**Table 6. Number of genes neighboring the start or end of 4,623 *Phytophthora cactorum* contigs by effector category.**

	Total genes	Neighboring contig breaks	% neighboring contig breaks
All genes	23884	5041	21.1
RxLRs	199	61	30.7
CRNs	76	39	51.3
NLPs	24	7	29.2
Protease inhibitors (all)	22	8	36.4
Secreted CAZymes	282	57	20.2
Non-secreted CAZY	410	61	14.9
Elicitins	47	10	21.3
BUSCO genes	272	16	5.9

The occurrence of genes neighboring contig breaks was not evenly distributed between gene categories ( $X^2 = 104.23$ ,  $df = 8$ ,  $p < 0.01$ ).

<https://doi.org/10.1371/journal.pone.0202305.t006>

the high repeat content (18%) observed in the assembly. The level of repetitive content was similar to that observed in the similarly sized genome of *P. sojae* but did not show the same levels of genome expansion as Clade 1 species *P. parasitica* or *P. infestans*. The sequenced and annotated *P. cactorum* genome is an important genomic resource that will aid functional study of effector gene candidates, as well as providing a resource to study the genomic basis of host specificity, which has been reported in the pathogen [90–95].

### Genomic characterisation of a broad profile of MAMPs and effectors

*Phytophthora* pathogens utilise a diverse range of secreted apoplastic and cytoplasmic effectors to aid infection. This work characterised the *P. cactorum* genome, identifying both apoplastic and cytoplasmic effector candidates as well as non-effectors that are typical of MAMP elicitors of host defence. This study unveiled the diversity of effectors in the *P. cactorum* genome, supplementing those effectors identified during development and cyst germination [49,88] with those that may be specifically expressed during infection and the transition to necrotrophy. This study identified considerably greater numbers of CRN, elicitors, GH, PL and RxLR candidates than previously identified in the *P. cactorum* transcriptome [49]. Equal or greater numbers of genes encoding NLPs, protease inhibitors, cutinases and PcF domain-carrying proteins were identified, however some of the candidates were discarded due to possession of a transmembrane domains or a GPI anchor.

This study reports a novel method for CRN prediction. The two-model LFLAK-DWL approach ensures identification is based upon the characteristic N-terminal domains of CRNs and not upon the variable C-terminal functional domains or upon regular-expression searches for conserved motifs, which may not be flexible enough to allow for sequence variation. This provides new opportunities for identification of new functional CRN domains and will advance research in this poorly understood effector family.

### Identification of homologs to well characterised avirulence genes

Establishing orthology between predicted proteomes is an important tool for translation of functional research from model *Phytophthora* species into *P. cactorum*. A total of 20,157 (84%) of predicted *P. cactorum* proteins had identified orthologs in other *Phytophthora* spp. Proteins in shared ortholog groups between *P. infestans* and *P. cactorum* allowed identification of two *Avr1* homologs, one *Avr3b*, one *Avr4*, two *Avrblb1* homologs (of which one was truncated)

and two homologs of *AvrSmira2* (Table 5). These characterised avirulence genes represent key targets for further functional study.

### Evidence for a two-speed genome

Effector genes have previously been characterised as showing uneven distributions throughout *Phytophthora* genomes, with measurements of intergenic distance showing that effector genes are located in gene-sparse regions of the *P. infestans* genome [9]. This has led to the concept of a two-speed genome in these organisms, where different regions of the genome are subject to different evolutionary pressures [96]. *P. cactorum* RxLR, CRN and secreted CAZyme effector candidates showed increased IG distance over non-effector genes, supporting the concept of a two-speed genome in *P. cactorum*. Fragmentation of *P. cactorum* assembly meant that 21% of genes were excluded from this analysis, due to being located on the end of a contig. Unsurprisingly, functional groups of genes within this group were not evenly represented on contig ends with 30% of RxLR and 50% crinkler genes located on contig ends in contrast to 6% of BUSCO conserved eukaryotic genes. A high frequency of contig breaks was observed in the 3' region of CRN genes and may have biased these distances to be shorter than if measurements were taken from a more contiguous assembly. These analyses should be repeated when improved assemblies become available. Furthermore, the low occurrence of conserved eukaryotic genes neighbouring contig breaks highlights that although these genes are comparatively useful in assessing assembly quality, their lack of an even distribution throughout difficult-to-assemble regions means that these genes do not accurately reflect the true "gene-space" in the assembly.

### Outcomes for breeding durable disease resistance

A broad complement of effectors and *Avr* genes are described in our characterisation of the *P. cactorum* genome. Qualitative resistance to *Phytophthora* pathogens is frequently determined by recognition of an RxLR in a gene-for gene dependant manner [23]. However, recognition of the *P. infestans* RxLR effector AVRSmira2 in field conditions is associated with quantitative resistance in potato [97]. Quantitative resistance to *Phytophthora* diseases has also been associated with basal defence [24,25]. Accordingly, this study characterises a broad range of effector genes and provides candidates to investigate the basis of quantitative strawberry resistance to *P. cactorum* [19–22]. RxLR effectors are still priority candidates disease related pathogen genes for functional study of strawberry resistance to *P. cactorum*, particularly homologs of *AvrSmira2* characterised avirulence genes.

### Supporting information

**S1 Table. Functional annotation of *P. cactorum* predicted proteins.** Data contains information on location, sequence, secretion status, identification as an RxLR, crinkler or CAZyme, orthology information (including orthogroup, number of proteins present in the orthogroup by species and orthogroup contained proteins), blast homology information (PHIbase, Swis-sprot and characterized oomycete avr genes) and identified InterProScan annotations. (XLSX)

**S2 Table. *P. cactorum* 10300 genes with homology to known *Phytophthora* effector gene candidates.** The orthogroup is shown for the query gene, with numbers of genes in each orthogroup shown for *P. cactorum* (Pcac), *P. parasitica* (Ppar), *P. infestans* (Pinf), *P. capsici* (Pcap) and *P. sojae* (Psoj), as well as functional annotation of each gene. Results showing best tBLASTx hits of all *P. cactorum* genes to a custom database with an E-value <  $1 \times 10^{-30}$ . (XLSX)

**S1 Data. Alignment of proteins from the 16 orthogroups representing necrosis inducing proteins (NLP).** Conservation of cytosine sites at alignment positions 624 and 661 identifies these proteins as Type1 NLPs.

(FASTA)

**S2 Data. Hmm models used for identification of Crinklers in *Phytophthora spp.* proteins.**

Hmm model for identification of the Crinkler LFLAK.

(HMM)

**S3 Data. Hmm models used for identification of Crinklers in *Phytophthora spp.* proteins.**

Hmm model for the Crinkler DWL domain.

(HMM)

**S4 Data. Fasta sequences of predicted Crinkler proteins.** Proteins sequences are included from *P. cactorum* (Pcac), *P. parasitica* (Ppar), *P. infestans* (Pinf), *P. capsici* (Pcap) and *P. sojae* (Psoj) genomes.

(FASTA)

## Acknowledgments

AA, CFN, LAL and RJH were funded by BBSRC grant BB/K017071/1. MB, EL and sequencing costs were financed by a strategic NIBIO project (basic funding). All authors thank Sophien Kamoun for valuable support throughout the project.

## Author Contributions

**Conceptualization:** Liliana M. Cano, Richard J. Harrison, May B. Brurberg.

**Formal analysis:** Andrew D. Armitage, Erik Lysøe, Charlotte F. Nellist, Liliana M. Cano.

**Funding acquisition:** Richard J. Harrison, May B. Brurberg.

**Investigation:** Liliana M. Cano, Richard J. Harrison.

**Methodology:** Andrew D. Armitage.

**Project administration:** Richard J. Harrison, May B. Brurberg.

**Resources:** Laura A. Lewis, Richard J. Harrison, May B. Brurberg.

**Software:** Andrew D. Armitage.

**Supervision:** Richard J. Harrison, May B. Brurberg.

**Validation:** Andrew D. Armitage.

**Visualization:** Andrew D. Armitage.

**Writing – original draft:** Andrew D. Armitage, Liliana M. Cano, Richard J. Harrison.

**Writing – review & editing:** Charlotte F. Nellist, Richard J. Harrison, May B. Brurberg.

## References

1. Kamoun S (2003) Molecular genetics of pathogenic oomycetes. *Eukaryotic Cell* 2: 191–199. <https://doi.org/10.1128/EC.2.2.191-199.2003> PMID: 12684368
2. Thines M, Kamoun S (2010) Oomycete-plant coevolution: recent advances and future prospects. *Curr Opin Plant Biol* 13: 427–433. <https://doi.org/10.1016/j.pbi.2010.04.001> PMID: 20447858

3. Kamoun S, Furzer O, Jones JDG, Judelson HS, Ali GS, Dalio RJD, et al. (2015) The Top 10 oomycete pathogens in molecular plant pathology. *Mol Plant Pathol* 16: 413–434. <https://doi.org/10.1111/mpp.12190> PMID: 25178392
4. Eikemo H, Brurberg MB, Davik J (2010) Resistance to *Phytophthora cactorum* in Diploid *Fragaria* Species. *HortScience*.
5. Stensvand A, Herrero ML, Talgø V (1999) Crown rot caused by *Phytophthora cactorum* in Norwegian strawberry production. *EPP Bulletin* 29: 155–158. <https://doi.org/10.1111/j.1365-2338.1999.tb00809.x>
6. Fennimore SA, Duniway JM, Browne GT, Martin FN, Ajwa HA, Westerdahl BB, et al. (2008) Methyl bromide alternatives evaluated for California strawberry nurseries. *Calif Agric (Berkeley)* 62: 62–67. <https://doi.org/10.3733/ca.v062n02p62>
7. Cal AD, Martinez-Treceño A, Lopez-Aranda JM, Melgarejo P (2004) Chemical alternatives to methyl bromide in spanish strawberry nurseries. *Plant Dis* 88: 210–214. <https://doi.org/10.1094/PDIS.2004.88.2.210>
8. Franceschetti M, Maqbool A, Jiménez-Dalmaroni MJ, Pennington HG, Kamoun S, Banfield MJ (2017) Effectors of filamentous plant pathogens: commonalities amid diversity. *Microbiol Mol Biol Rev* 81. <https://doi.org/10.1128/MMBR.00066-16> PMID: 28356329
9. Haas BJ, Kamoun S, Zody MC, Jiang RHY, Handsaker RE, Cano LM, et al. (2009) Genome sequence and analysis of the Irish potato famine pathogen *Phytophthora infestans*. *Nature* 461: 393–398. <https://doi.org/10.1038/nature08358> PMID: 19741609
10. Tyler BM, Tripathy S, Zhang X, Dehal P, Jiang RHY, Aerts A, et al. (2006) *Phytophthora* genome sequences uncover evolutionary origins and mechanisms of pathogenesis. *Science* 313: 1261–1266. <https://doi.org/10.1126/science.1128796> PMID: 16946064
11. Lamour KH, Mudge J, Gobena D, Hurtado-Gonzales OP, Schmutz J, Kuo A, et al. (2012) Genome sequencing and mapping reveal loss of heterozygosity as a mechanism for rapid adaptation in the vegetable pathogen *Phytophthora capsici*. *Mol Plant Microbe Interact* 25: 1350–1360. <https://doi.org/10.1094/MPMI-02-12-0028-R> PMID: 22712506
12. Fry WE, Birch PRJ, Judelson HS, Grünwald NJ, Danies G, Everts KL, et al. (2015) Five Reasons to Consider *Phytophthora infestans* a Reemerging Pathogen. *Phytopathology* 105: 966–981. <https://doi.org/10.1094/PHYTO-01-15-0005-FI> PMID: 25760519
13. Armstrong MR, Whisson SC, Pritchard L, Bos JIB, Venter E, Avrova AO, et al. (2005) An ancestral oomycete locus contains late blight avirulence gene *Avr3a*, encoding a protein that is recognized in the host cytoplasm. *Proc Natl Acad Sci USA* 102: 7766–7771. <https://doi.org/10.1073/pnas.0500113102> PMID: 15894622
14. van Poppel PMJA, Guo J, van de Vondervoort PJI, Jung MWM, Birch PRJ, Whisson SC, et al. (2008) The *Phytophthora infestans* avirulence gene *Avr4* encodes an RXLR-dEER effector. *Mol Plant Microbe Interact* 21: 1460–1470. <https://doi.org/10.1094/MPMI-21-11-1460> PMID: 18842095
15. Champouret N, Bouwmeester K, Rietman H, van der Lee T, Maliepaard C, Heupink A, et al. (2009) *Phytophthora infestans* isolates lacking class I *ipiO* variants are virulent on *Rpi-blb1* potato. *Mol Plant Microbe Interact* 22: 1535–1545. <https://doi.org/10.1094/MPMI-22-12-1535> PMID: 19888819
16. Oh S-K, Young C, Lee M, Oliva R, Bozkurt TO, Cano LM, et al. (2009) In planta expression screens of *Phytophthora infestans* RXLR effectors reveal diverse phenotypes, including activation of the *Solanum bulbocastanum* disease resistance protein *Rpi-blb2*. *Plant Cell* 21: 2928–2947. <https://doi.org/10.1105/tpc.109.068247> PMID: 19794118
17. Gilroy EM, Breen S, Whisson SC, Squires J, Hein I, Kaczmarek M, et al. (2011) Presence/absence, differential expression and sequence polymorphisms between *PiAVR2* and *PiAVR2*-like in *Phytophthora infestans* determine virulence on R2 plants. *New Phytol* 191: 763–776. <https://doi.org/10.1111/j.1469-8137.2011.03736.x> PMID: 21539575
18. Sugimoto T, Kato M, Yoshida S, Matsumoto I, Kobayashi T, Kaga A, et al. (2012) Pathogenic diversity of *Phytophthora sojae* and breeding strategies to develop *Phytophthora*-resistant soybeans. *Breed Sci* 61: 511–522. <https://doi.org/10.1270/jsbbs.61.511> PMID: 23136490
19. Nellist CF, Vickerstaff R, Sobczyk MK, Marina-Montes C, Brain P, Wilson FM, et al. (2018) Quantitative Trait Loci Controlling *Phytophthora cactorum* Resistance in the Cultivated Octoploid Strawberry (*Fragaria x ananassa*). *BioRxiv*. <https://doi.org/10.1101/249573>
20. Denoyes-Rothan B, Lerceteau-Köhler E, Guérin G, Bosseur S, Bariac J, Martin E, et al. (2004) Qtl analysis for resistance to *Colletotrichum acutatum* and *Phytophthora cactorum* in octoploid strawberry (*Fragaria x ananassa*). *Acta Hort*: 147–152. <https://doi.org/10.17660/ActaHortic.2004.663.19>
21. Shaw DV, Hansen J, Browne GT, Shaw SM (2008) Components of genetic variation for resistance of strawberry to *Phytophthora cactorum* estimated using segregating seedling populations and their parent genotypes. *Plant Pathology* 57: 210–215. <https://doi.org/10.1111/j.1365-3059.2007.01773.x>

22. Shaw DV, Hansen J, Browne GT (2006) Genotypic Variation for Resistance to *Phytophthora cactorum* in a California Strawberry Breeding Population. *Journal of the American Society for Horticultural Science*.
23. Anderson RG, Deb D, Fedkenheuer K, McDowell JM (2015) Recent progress in RXLR effector research. *Mol Plant Microbe Interact* 28: 1063–1072. <https://doi.org/10.1094/MPMI-01-15-0022-CR> PMID: 26125490
24. Vega-Sánchez ME, Redinbaugh MG, Costanzo S, Dorrance AE (2005) Spatial and temporal expression analysis of defense-related genes in soybean cultivars with different levels of partial resistance to *Phytophthora sojae*. *Physiological and Molecular Plant Pathology* 66: 175–182. <https://doi.org/10.1016/j.pmpp.2005.07.001>
25. Thomas R, Fang X, Ranathunge K, Anderson TR, Peterson CA, Bernards MA (2007) Soybean root suberin: anatomical distribution, chemical composition, and relationship to partial resistance to *Phytophthora sojae*. *Plant Physiol* 144: 299–311. <https://doi.org/10.1104/pp.106.091090> PMID: 17494920
26. Wawra S, Trusch F, Matena A, Apostolakis K, Linne U, Zhukov I, et al. (2017) The RxLR Motif of the Host Targeting Effector AVR3a of *Phytophthora infestans* Is Cleaved before Secretion. *Plant Cell* 29: 1184–1195. <https://doi.org/10.1105/tpc.16.00552> PMID: 28522546
27. Win J, Krasileva KV, Kamoun S, Shirasu K, Staskawicz BJ, Banfield MJ (2012) Sequence divergent RXLR effectors share a structural fold conserved across plant pathogenic oomycete species. *PLoS Pathog* 8: e1002400. <https://doi.org/10.1371/journal.ppat.1002400> PMID: 22253591
28. Kourelis J, van der Hoorn RAL (2018) Defended to the Nines: 25 years of Resistance Gene Cloning Identifies Nine Mechanisms for R Protein Function. *Plant Cell* 30: 285–299. <https://doi.org/10.1105/tpc.17.00579> PMID: 29382771
29. Torto TA, Li S, Styer A, Huitema E, Testa A, Gow NA, et al. (2003) EST mining and functional expression assays identify extracellular effector proteins from the plant pathogen *Phytophthora*. *Genome Res* 13: 1675–1685. <https://doi.org/10.1101/gr.910003> PMID: 12840044
30. Jupe J, Stam R, Howden AJM, Morris JA, Zhang R, Hedley PE, et al. (2013) *Phytophthora capsici*-tomato interaction features dramatic shifts in gene expression associated with a hemi-biotrophic lifestyle. *Genome Biol* 14: R63. <https://doi.org/10.1186/gb-2013-14-6-r63> PMID: 23799990
31. Stam R, Howden AJM, Delgado-Cerezo M, M M Amaro TM, Motion GB, Pham J, et al. (2013) Characterization of cell death inducing *Phytophthora capsici* CRN effectors suggests diverse activities in the host nucleus. *Front Plant Sci* 4: 387. <https://doi.org/10.3389/fpls.2013.00387> PMID: 24155749
32. Stam R, Jupe J, Howden AJM, Morris JA, Boevink PC, Hedley PE, et al. (2013) Identification and Characterisation CRN Effectors in *Phytophthora capsici* Shows Modularity and Functional Diversity. *PLoS ONE* 8: e59517. <https://doi.org/10.1371/journal.pone.0059517> PMID: 23536880
33. Rajput NA, Zhang M, Shen D, Liu T, Zhang Q, Ru Y, et al. (2015) Overexpression of a *Phytophthora* Cytoplasmic CRN Effector Confers Resistance to Disease, Salinity and Drought in *Nicotiana benthamiana*. *Plant Cell Physiol* 56: 2423–2435. <https://doi.org/10.1093/pcp/pcv164> PMID: 26546319
34. Amaro TMMM, Thilliez GJA, Motion GB, Huitema E (2017) A perspective on CRN proteins in the genomics age: evolution, classification, delivery and function revisited. *Front Plant Sci* 8: 99. <https://doi.org/10.3389/fpls.2017.00099> PMID: 28217133
35. Blackman LM, Cullerne DP, Hardham AR (2014) Bioinformatic characterisation of genes encoding cell wall degrading enzymes in the *Phytophthora parasitica* genome. *BMC Genomics* 15: 785. <https://doi.org/10.1186/1471-2164-15-785> PMID: 25214042
36. Tian M, Benedetti B, Kamoun S (2005) A Second Kazal-like protease inhibitor from *Phytophthora infestans* inhibits and interacts with the apoplastic pathogenesis-related protease P69B of tomato. *Plant Physiol* 138: 1785–1793. <https://doi.org/10.1104/pp.105.061226> PMID: 15980196
37. Orsomando G, Brunetti L, Pucci K, Ruggeri B, Ruggieri S (2011) Comparative structural and functional characterization of putative protein effectors belonging to the PcF toxin family from *Phytophthora* spp. *Protein Sci* 20: 2047–2059. <https://doi.org/10.1002/pro.742> PMID: 21936011
38. Kamoun S, van West P, de Jong AJ, de Groot KE, Vleeshouwers V, Govers F (1997) A gene encoding a protein elicitor of *Phytophthora infestans* is down-regulated during infection of potato. *Mol Plant Microbe Interact* 10: 13–20. <https://doi.org/10.1094/MPMI.1997.10.1.13> PMID: 9002268
39. Gottlieb D, Knaus RJ, Wod SG (1978) Differences in the sterol synthesizing pathways of sterol-producing and non-sterol-producing fungi [Pythiaceae]. *Phytopathology*.
40. Wood SG, Gottlieb D (1978) Evidence from cell-free systems for differences in the sterol biosynthetic pathway of *Rhizoctonia solani* and *Phytophthora cinnamomi*. *Biochem J* 170: 355–363. PMID: 637850
41. Erwin DC, Ribeiro OK (1996) *Phytophthora* diseases worldwide. *Phytophthora diseases worldwide*.
42. Grenville-Briggs LJ, Kushwaha SK, Cleary MR, Witzell J, Savenkov EI, Whisson SC, et al. (2017) Draft genome of the oomycete pathogen *Phytophthora cactorum* strain LV007 isolated from European beech (*Fagus sylvatica*). *Genom Data* 12: 155–156. <https://doi.org/10.1016/j.gdata.2017.05.010> PMID: 28560165

43. Martin FN, Blair JE, Coffey MD (2014) A combined mitochondrial and nuclear multilocus phylogeny of the genus *Phytophthora*. *Fungal Genet Biol* 66: 19–32. <https://doi.org/10.1016/j.fgb.2014.02.006> PMID: 24603058
44. Simpson JT, Wong K, Jackman SD, Schein JE, Jones SJM, Birol I (2009) ABySS: a parallel assembler for short read sequence data. *Genome Res* 19: 1117–1123. <https://doi.org/10.1101/gr.089532.108> PMID: 19251739
45. Gurevich A, Saveliev V, Vyahhi N, Tesler G (2013) QUASt: quality assessment tool for genome assemblies. *Bioinformatics* 29: 1072–1075. <https://doi.org/10.1093/bioinformatics/btt086> PMID: 23422339
46. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM (2015) BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31: 3210–3212. <https://doi.org/10.1093/bioinformatics/btv351> PMID: 26059717
47. Hoff KJ, Lange S, Lomsadze A, Borodovsky M, Stanke M (2016) BRAKER1: Unsupervised RNA-Seq-Based Genome Annotation with GeneMark-ET and AUGUSTUS. *Bioinformatics* 32: 767–769. <https://doi.org/10.1093/bioinformatics/btv661> PMID: 26559507
48. Stanke M, Morgenstern B (2005) AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic Acids Res* 33: W465–7. <https://doi.org/10.1093/nar/gki458> PMID: 15980513
49. Chen XR, Zhang BY, Xing YP, Li QY, Li YP, Tong YH, et al. (2014) Transcriptomic analysis of the phytopathogenic oomycete *Phytophthora cactorum* provides insights into infection-related effectors. *BMC Genomics* 15: 980. <https://doi.org/10.1186/1471-2164-15-980> PMID: 25406848
50. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29: 15–21. <https://doi.org/10.1093/bioinformatics/bts635> PMID: 23104886
51. Testa AC, Hane JK, Ellwood SR, Oliver RP (2015) CodingQuarry: highly accurate hidden Markov model gene prediction in fungal genomes using RNA-seq transcripts. *BMC Genomics* 16: 170. <https://doi.org/10.1186/s12864-015-1344-4> PMID: 25887563
52. Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, et al. (2014) InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30: 1236–1240. <https://doi.org/10.1093/bioinformatics/btu031> PMID: 24451626
53. Bairoch A, Apweiler R (2000) The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Res* 28: 45–48. <https://doi.org/10.1093/nar/28.1.45> PMID: 10592178
54. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215: 403–410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2) PMID: 2231712
55. Winnenburg R, Baldwin TK, Urban M, Rawlings C, Köhler J, Hammond-Kosack KE (2006) PHI-base: a new database for pathogen host interactions. *Nucleic Acids Res* 34: D459–64. <https://doi.org/10.1093/nar/gkj047> PMID: 16381911
56. Huang L, Zhang H, Wu P, Entwistle S, Li X, Yohe T, et al. (2018) dbCAN-seq: a database of carbohydrate-active enzyme (CAZyme) sequence and annotation. *Nucleic Acids Res* 46: D516–D521. <https://doi.org/10.1093/nar/gkx894> PMID: 30053267
57. Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B (2009) The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Res* 37: D233–8. <https://doi.org/10.1093/nar/gkn663> PMID: 18838391
58. Käll L, Krogh A, Sonnhammer ELL (2004) A combined transmembrane topology and signal peptide prediction method. *J Mol Biol* 338: 1027–1036. <https://doi.org/10.1016/j.jmb.2004.03.016> PMID: 15111065
59. Bhattacharjee S, Hiller NL, Liolios K, Win J, Kanneganti T-D, Young C (2006) The malarial host-targeting signal is conserved in the Irish potato famine pathogen. *PLoS Pathog* 2: e50. <https://doi.org/10.1371/journal.ppat.0020050> PMID: 16733545
60. Win J, Morgan W, Bos J, Krasileva KV, Cano LM, Chaparro-Garcia A, et al. (2007) Adaptive evolution has targeted the C-terminal domain of the RXLR effectors of plant pathogenic oomycetes. *Plant Cell* 19: 2349–2369. <https://doi.org/10.1105/tpc.107.051037> PMID: 17675403
61. Krogh A, Larsson B, von Heijne G, Sonnhammer ELL (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol* 305: 567–580. <https://doi.org/10.1006/jmbi.2000.4315> PMID: 11152613
62. Fankhauser N, Mäser P (2005) Identification of GPI anchor attachment signals by a Kohonen self-organizing map. *Bioinformatics* 21: 1846–1852. <https://doi.org/10.1093/bioinformatics/bti299> PMID: 15691858
63. Käll L, Krogh A, Sonnhammer ELL (2007) Advantages of combined transmembrane topology and signal peptide prediction—the Phobius web server. *Nucleic Acids Res* 35: W429–32. <https://doi.org/10.1093/nar/gkm256> PMID: 17483518

64. Whisson SC, Boevink PC, Moleleki L, Avrova AO, Morales JG, Gilroy EM, et al. (2007) A translocation signal for delivery of oomycete effector proteins into host plant cells. *Nature* 450: 115–118. <https://doi.org/10.1038/nature06203> PMID: 17914356
65. Emms DM, Kelly S (2015) OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol* 16: 157. <https://doi.org/10.1186/s13059-015-0721-2> PMID: 26243257
66. Chen H, Boutros PC (2011) VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. *BMC Bioinformatics* 12: 35. <https://doi.org/10.1186/1471-2105-12-35> PMID: 21269502
67. Li L, Stoeckert CJ, Roos DS (2003) OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res* 13: 2178–2189. <https://doi.org/10.1101/gr.1224503> PMID: 12952885
68. Brunner F, Rosahl S, Lee J, Rudd JJ, Geiler C, Kauppinen S, et al. (2002) Pep-13, a plant defense-inducing pathogen-associated pattern from *Phytophthora* transglutaminases. *EMBO J* 21: 6681–6688. <https://doi.org/10.1093/emboj/cdf667> PMID: 12485989
69. Robinson SM, Bostock RM (2015)  $\beta$ -glucans and eicosapolyenoic acids as MAMPs in plant-oomycete interactions: past and present. *Front Plant Sci* 5: 797. <https://doi.org/10.3389/fpls.2014.00797> PMID: 25628639
70. Kamoun S (2006) A catalogue of the effector secretome of plant pathogenic oomycetes. *Annu Rev Phytopathol* 44: 41–60. <https://doi.org/10.1146/annurev.phyto.44.070505.143436> PMID: 16448329
71. Damasceno CMB, Bishop JG, Ripoll DR, Win J, Kamoun S, Rose JK (2008) Structure of the glucanase inhibitor protein (GIP) family from *phytophthora* species suggests coevolution with plant endo-beta-1,3-glucanases. *Mol Plant Microbe Interact* 21: 820–830. <https://doi.org/10.1094/MPMI-21-6-0820> PMID: 18624645
72. Orsomando G, Lorenzi M, Raffaelli N, Dalla Rizza M, Mezzetti B, Ruggieri S (2001) Phytotoxic protein PcF, purification, characterization, and cDNA sequencing of a novel hydroxyproline-containing factor secreted by the strawberry pathogen *Phytophthora cactorum*. *J Biol Chem* 276: 21578–21584. <https://doi.org/10.1074/jbc.M101377200> PMID: 11262411
73. Gijzen M, Nürnberger T (2006) Nep1-like proteins from plant pathogens: recruitment and diversification of the NPP1 domain across taxa. *Phytochemistry* 67: 1800–1807. <https://doi.org/10.1016/j.phytochem.2005.12.008> PMID: 16430931
74. Feng BZ, Zhu XP, Fu L, Lv RF, Storey D, Tooley P, et al. (2014) Characterization of necrosis-inducing NLP proteins in *Phytophthora capsici*. *BMC Plant Biol* 14: 126. <https://doi.org/10.1186/1471-2229-14-126> PMID: 24886309
75. Tian M, Win J, Song J, van der Hoorn R, van der Knaap E, Kamoun S (2007) A *Phytophthora infestans* cystatin-like protein targets a novel tomato papain-like apoplastic protease. *Plant Physiol* 143: 364–377. <https://doi.org/10.1104/pp.106.090050> PMID: 17085509
76. Bouwmeester K, de Sain M, Weide R, Gouget A, Klamer S, Canut H, et al. (2011) The lectin receptor kinase LecRK-I.9 is a novel *Phytophthora* resistance component and a potential host target for a RXLR effector. *PLoS Pathog* 7: e1001327. <https://doi.org/10.1371/journal.ppat.1001327> PMID: 21483488
77. Kong G, Zhao Y, Jing M, Huang J, Yang J, Xia Y, et al. (2015) The activation of *phytophthora* effector *avr3b* by plant cyclophilin is required for the nudix hydrolase activity of *avr3b*. *PLoS Pathog* 11: e1005139. <https://doi.org/10.1371/journal.ppat.1005139> PMID: 26317500
78. Dong S, Wang Y (2016) Nudix effectors: A common weapon in the arsenal of plant pathogens. *PLoS Pathog* 12: e1005704. <https://doi.org/10.1371/journal.ppat.1005704> PMID: 27513453
79. Sperisen P, Schmid CD, Bucher P, Zilian O (2005) Stealth proteins: in silico identification of a novel protein family rendering bacterial pathogens invisible to host immune defense. *PLoS Comput Biol* 1: e63. <https://doi.org/10.1371/journal.pcbi.0010063> PMID: 16299590
80. Xu RQ, Blanvillain S, Feng JX, Jiang BL, Li XZ, Wei HY, et al. (2008) *AvrAC*(Xcc8004), a type III effector with a leucine-rich repeat domain from *Xanthomonas campestris* pathovar *campestris* confers avirulence in vascular tissues of *Arabidopsis thaliana* ecotype Col-0. *J Bacteriol* 190: 343–355. <https://doi.org/10.1128/JB.00978-07> PMID: 17951377
81. Dong S, Yin W, Kong G, Yang X, Qutob D, Chen Q, et al. (2011) *Phytophthora sojae* avirulence effector *Avr3b* is a secreted NADH and ADP-ribose pyrophosphorylase that modulates plant immunity. *PLoS Pathog* 7: e1002353. <https://doi.org/10.1371/journal.ppat.1002353> PMID: 22102810
82. Du Y, Mpina MH, Birch PRJ, Bouwmeester K, Govers F (2015) *Phytophthora infestans* RXLR Effector AVR1 Interacts with Exocyst Component Sec5 to Manipulate Plant Immunity. *Plant Physiol* 169: 1975–1990. <https://doi.org/10.1104/pp.15.01169> PMID: 26336092
83. Jiang RHY, Tripathy S, Govers F, Tyler BM (2008) RXLR effector reservoir in two *Phytophthora* species is dominated by a single rapidly evolving superfamily with more than 700 members. *Proc Natl Acad Sci USA* 105: 4874–4879. <https://doi.org/10.1073/pnas.0709303105> PMID: 18344324

84. Hirakawa H, Shirasawa K, Kosugi S, Tashiro K, Nakayama S, Yamada M, et al. (2014) Dissection of the octoploid strawberry genome by deep sequencing of the genomes of *Fragaria* species. *DNA Res* 21: 169–181. <https://doi.org/10.1093/dnares/dst049> PMID: 24282021
85. Davik J, Sargent DJ, Brurberg MB, Lien S, Kent M, Alsheikh M (2015) A ddRAD Based Linkage Map of the Cultivated Strawberry, *Fragaria x ananassa*. *PLoS ONE* 10: e0137746. <https://doi.org/10.1371/journal.pone.0137746> PMID: 26398886
86. Daccord N, Celton J-M, Linsmith G, Becker C, Choisne N, Schijlen E, et al. (2017) High-quality de novo assembly of the apple genome and methylome dynamics of early fruit development. *Nat Genet* 49: 1099–1106. <https://doi.org/10.1038/ng.3886> PMID: 28581499
87. Antanaviciute L, Fernández-Fernández F, Jansen J, Banchi E, Evans KM, Viola R, et al. (2012) Development of a dense SNP-based linkage map of an apple rootstock progeny using the *Malus Infinium* whole genome genotyping array. *BMC Genomics* 13: 203. <https://doi.org/10.1186/1471-2164-13-203> PMID: 22631220
88. Chen X, Klemsdal SS, Brurberg MB (2011) Identification and analysis of *Phytophthora cactorum* genes up-regulated during cyst germination and strawberry infection. *Curr Genet* 57: 297–315. <https://doi.org/10.1007/s00294-011-0348-0> PMID: 21698431
89. Chen XR, Huang SX, Zhang Y, Sheng GL, Zhang BY, Li QY, et al. (2017) Transcription profiling and identification of infection-related genes in *Phytophthora cactorum*. *Mol Genet Genomics* 293: 541–555. <https://doi.org/10.1007/s00438-017-1400-7> PMID: 29218408
90. Hantula J, Lilja A, Parikka P (1997) Genetic variation and host specificity of *Phytophthora cactorum* isolated in Europe. *Mycol Res* 101: 565–572. <https://doi.org/10.1017/S0953756296002900>
91. Hantula J, Lilja A, Nuorteva H, Parikka P, Werres S (2000) Pathogenicity, morphology and genetic variation of *Phytophthora cactorum* from strawberry, apple, rhododendron, and silver birch. *Mycol Res* 104: 1062–1068. <https://doi.org/10.1017/S0953756200002999>
92. Lilja A, Karjalainen R, Parikka P, Kammiovirta K, Nuorteva H (1998) Pathogenicity and genetic variation of *Phytophthora cactorum* from silver birch and strawberry. *European Journal of Plant Pathology* 104: 529–535.
93. Thomidis T (2001) Testing variability in pathogenicity of *Phytophthora cactorum*, *P. citrophthora* and *P. syringae* to apple, pear, peach, cherry and plum rootstocks. *Phytoparasitica* 29: 47–49. <https://doi.org/10.1007/BF02981813>
94. Thomidis T (2003) Variability in pathogenicity among Greek isolates of *Phytophthora cactorum* to four peach rootstocks. *Aust J Exp Agric* 43: 99. <https://doi.org/10.1071/EA01203>
95. Bhat RG, Colowit PM, Tai TH, Aradhya MK, Browne GT (2006) Genetic and Pathogenic Variation in *Phytophthora cactorum* Affecting Fruit and Nut Crops in California. *Plant Dis* 90: 161–169. <https://doi.org/10.1094/PD-90-0161>
96. Dong S, Raffaele S, Kamoun S (2015) The two-speed genomes of filamentous pathogens: waltz with plants. *Curr Opin Genet Dev* 35: 57–65. <https://doi.org/10.1016/j.gde.2015.09.001> PMID: 26451981
97. Rietman H, Bijsterbosch G, Cano LM, Lee H-R, Vossen JH, Jacobsen E, et al. (2012) Qualitative and quantitative late blight resistance in the potato cultivar Sarpö Mira is determined by the perception of five distinct RXLR effectors. *Mol Plant Microbe Interact* 25: 910–919. <https://doi.org/10.1094/MPMI-01-12-0010-R> PMID: 22414442