

Effector Repertoire of *Phytophthora betacei*: in Search of Possible Virulence Factors Responsible for its Host Specificity

Paola Rojas^a, David Urbina^a, David Ayala^{a,b}, Natalia Guayazan^a, Maria Fernanda Mideros^a, Adriana Bernal^c, Martha Cardenas^a and Silvia Restrepo ^{a*}.

^a Laboratorio de Micología y Fitopatología, Facultad de Ingeniería, Universidad de los Andes, Carrera 1 No 18A-12, Bogotá.

^bLaboratorio de Biología Computacional y Ecología Microbiana (BCEM), Universidad de los Andes, Carrera 1 No 18A-10, Bogotá.

^cLaboratorio de Interacciones Moleculares de Microorganismos en Agricultura – LIMMA, Universidad de los Andes, Carrera 1 No 18A-10, Bogotá.

* Corresponding author, Silvia Restrepo; E-mail: srestrep@uniandes.edu.co

Abstract

Phytophthora betacei is an oomycete plant pathogen closely related to *Phytophthora infestans*. It infects tree tomato (*Solanum betaceum*) in northern South America, but is, under natural conditions, unable to infect potatoes or tomatoes, the main hosts of its sister species *P. infestans*. We characterized, and compared the effector repertoires of *P. betacei* and other *Phytophthora* species. To this end, we used *in silico* approaches to predict and describe the repertoire of secreted proteins in *Phytophthora* species and determine unique and core effectors. *P. betacei* has the largest proteome and secretome of all *Phytophthora* species evaluated. We identified between 450 and 1933 candidate effector genes in *P. ramorum*, *P. sojae*, *P. cactorum*, *P. parasitica*, *P. palmivora*, *P. infestans*, and *P. betacei* genomes. The *P. betacei* predicted secretome contains 5653 proteins, 1126 of which are apoplastic effectors and 807 cytoplasmic effectors. Genes encoding cytoplasmic effectors include 791 genes with an RxLR domain (the largest number known so far in a *Phytophthora* species) and 16 with a Crinkler (CRN) domain. We detected homologs of previously described avirulence gene (Avr) present in *Phytophthora* spp, such as Avr1, Avr3b, Avr4, and Avrblb1, suggesting a high level of effector gene conservation among *Phytophthora* species. Nonetheless, fewer CRN effectors were obtained in *P. betacei* compared to all other *Phytophthora* species analyzed. The comparison between *P. infestans* and *P. betacei* effector profiles shows unique features in *P. betacei* that might be involved in pathogenesis and host preference. Indeed, 402 unique predicted effector genes were detected in *P. betacei*, corresponding to 197 apoplastic effector genes, 203 RxLR cytoplasmic effector genes, and 2 effector genes with CRN domain. This is the first characterization of the effector profile of *P. betacei* and the broadest comparison of predicted effector repertoires in the genus *Phytophthora* following a standardized prediction pipeline. The resultant *P. betacei* putative effector repertoire provides a reasonable set of proteins whose experimental validation could lead to understand the specific virulence factors responsible for the host specificity of this species.

Keywords Oomycetes, *Phytophthora betacei*, effector genes, comparative genomics, host specificity

1. Introduction

The oomycete genus *Phytophthora* is known for its role as a plant pathogen that infects a wide range of plants of economic importance (Hardham, 2001; Kamoun, 2007). Nowadays there are 142 described species and more than 40 provisionally recognized species as a result of the available genomic technologies and the use of different concepts to define a species within the genus (Grünwald and Flier, 2005; Fry, 2008; Haas et al., 2009; Yang et al., 2017). Notoriously in this group stands *Phytophthora infestans*, a pathogen causing the late blight disease of potato (*Solanum tuberosum*) (Fry, 2008). This pathogen is able to infect other members of the Solanaceae family, including tomato (*Solanum lycopersicum*) and tree tomato (*Solanum betaceum*) (Mideros et al., 2018). Recently, *P. betacei* was proposed as a new species within the clade 1c using phylogenetic, population genetics and morphological approaches (Mideros et al., 2018). However, *P. betacei* has not been reported infecting potatoes or tomatoes, the main hosts of its sister species *P. infestans* (Mideros et al., 2018). These observations suggest that this species presents host specificity for tree tomato (*Solanum betaceum*) (Mideros et al., 2018). A detailed investigation of the infection cycle of *P. betacei* and *P. infestans* allowed to clearly identify their biotrophic and necrotrophic stages and to notice that *P. betacei* shows a typical hemibiotrophic infection (Guayazán et al., 2017).

Plant pathogens introduce effector proteins inside host plant cells to promote infection (Evangelisti et al., 2017). Effector proteins target different sites in host plant tissues. Some effectors act in the extracellular space, where they interfere with apoplastic plant proteins involved in plant defense; other effectors such as RxLR and Crinkler (CRN) families, translocate inside host cells (Kamoun, 2006, 2007). RxLR effectors are characterized by the presence of a secretion signal peptide followed by a conserved N-terminal domain defined by the RxLR (Arg-Xaa-Leu-Arg) consensus sequence. The RxLR domain is required for the translocation inside plant cells (Whisson et al., 2007) but it is dispensable for the biochemical activity of the effectors when expressed directly inside host cells (Schoebitz et al., 2013). Thus, RxLR effectors are usually short proteins with low similarity in their C terminal. They are localized into diverse subcellular compartments where they interact with plant target proteins playing an important role in the infection process (Vleeshouwers et al., 2008; Oh et al., 2009). Besides that, the LFLAK motif is present in CRN effectors, named after a crinkling and necrosis phenotype caused by some CRN proteins when expressed in plants (Amaro et al., 2017). Critically, expressed mature CRN proteins retained cell death-inducing activity, suggesting functions targeting cytoplasmic host factors, a hypothesis that was confirmed when the translocation activity of CRN, carrying an LFLAK motif at the N-terminus, was demonstrated (Oh et al., 2009; Schornack et al., 2009, 2010; Stam et al., 2013b).

Effector proteins modulate the immune response of the host and enable the infection process (Kamoun, 2007). Thus, in some cases, they may determine the pathogenicity in a susceptible host (Haldar et al., 2006; Ntoukakis et al., 2009). In this sense, *Phytophthora infestans* exhibits a high evolutionary potential and its genome shows a discontinuous distribution of gene density. Effector genes, such as members of the RXLR and CRN families, are located in expanded, repeat-rich and gene-sparse regions of the genome, corresponding to highly plastic genomic regions, enriched in transposable elements (TEs)

(Haas *et al.*, 2009). Raffaele *et al.*, (2010) supported this hypothesis by identifying new previously-overlooked genes involved in virulence of *P. infestans* using a comparative genomics and transcriptomics approach. Then, they provided genomic context to each effector by quantitatively delimiting the gene-dense and gene sparse regions in the *P. infestans* genome. The effector repertoire of *P. betacei* remains unexplored, and there is no information about the factors responsible for its host specificity. Thus, it could be possible that *P. betacei* presents an effector repertoire that allows it to infect only *S. betaceum* and preclude it from infecting the other main hosts of its sister species *P. infestans*.

The objective of this study was to predict and characterize the *P. betacei* repertoire of effector proteins and its genomic context. We compared this repertoire with that of *P. infestans* and other *Phytophthora* species, in order to shed light on potential candidates that might define host range in this clade of the genus *Phytophthora*. This is the first contribution addressed to characterize the effector profile in this new species, therefore, it would be important to describe new possible virulence factors probably responsible for the host specificity of *P. betacei*.

2. Methods

2.1 Proteome and secretome predictions

The genome of *P. betacei* (strain P8084; accession number: PRJNA608953) as well as the genomes and annotation information of *P. ramorum* (strain: Pr102; accession number: GCA_000149735.1), *P. infestans* (strain: Refseq T30-4; Accession number: GCA_000142945.1), *P. cactorum* (Strain: 10300; accession number: GCA_003287315.1), *P. sojae* (strain: V3; accession number: GCA_000149755.2), *P. parasitica* (strain: INRA310; accession number: GCA_000247585.2), and the transcriptome of *P. palmivora* (strain: LILI; accession number: PRJNA503573) available at the NCBI database were downloaded and used as input to obtain the proteome of each species. Unfortunately, the genome of *P. andina* could not be included in the analysis due to its genome assembly poor quality. For *P. betacei*, the genome sequence, is the result of a hybrid assembly between Illumina and PacBio data. The annotation information of *P. betacei* was carried out using MAKER2 (Hold and Yandell, 2011) with default settings and using the annotation information of *P. infestans* T30-4 as input (data not published, available upon request). Briefly, the proteomes were predicted using the gffread function with the ‘coding only’ and ‘only print mRNAs with a full valid CDS’ options, this function is available at the cufflinks (v. 2.2.1) package (Trapnell *et al.*, 2012). Next, to obtain the secretome for the six *Phytophthora* species included in this analysis, we based our prediction strategy on the methodology proposed by Evangelisti *et al.*, 2017 and implemented in the SecretSanta pipeline developed by Gogleva *et al.* (2018) that takes predicted proteomes as input files. This pipeline corresponds to an interface in R (R Core Team, 2018) that uses different tools to allow the prediction of extracellular proteins that are secreted in a classical way. To this end, Gogleva’s pipeline uses (i) SignalP (v. 2.0) (Nielsen and Krogh, 1998), SignalP (v. 3.0) (Bendtsen *et al.*, 2004) and the most recent version SignalP (v. 4.1) (Petersen *et al.*, 2011) to predict signal peptides and cleavage sites with thresholds specific for oomycetes

sequences; (ii) TMHMM (v. 2.0) (Krogh et al., 2001) to discard proteins with predicted transmembrane domains; (iii) TargetP (v. 1.1) (Emanuelsson et al., 2000) to select proteins that do not target plastids or mitochondria. Finally, (iv) all proteins with terminal ‘KDEL’ or ‘HDEL’ motifs were removed, because these motifs are known to be ER-retention signals (Evangelisti et al., 2017). It is worth mentioning that we used the transcriptome of *P. palmivora* as a positive control to assure the correct performance of Gogleva’s pipeline. The whole pipeline is depicted in Supplementary Figure 1.

To avoid SignalP output bias, the first resulting set of predicted proteins having a signal peptide was retained. Then, a retrieval of partial proteins was performed, using the function `m_slicer` which generates sequences with alternative translation start sites based on the assumption that translation start sites might be misclassified in the proteome, which in turn would result in signal peptides also being misclassified. The `m_slicer` output was used as an input for secretome prediction pipeline (SecretSanta), as described above, we call this set of proteins, rescued proteins (Supplementary Figure 1).

Finally, we removed duplicate sequences using `cd-hit` (v. 4.6.8) (Li and Godzik, 2006) with a similarity cutoff of 100%. This process resulted in a non-redundant dataset of putative secreted proteins that was denominated as the secretome.

2.2 Functional annotation of secreted proteins

To annotate the predicted secretome, different approaches were used. Initially we performed a `blastp` (Altschul et al., 1997) search against the GenBank non-redundant (NR) database with an e-value cutoff of $\leq 10^{-6}$. In addition, draft functional annotations were assigned for proteins using InterProScan (v. 5.18–57.0) (Jones et al., 2014) with the default parameters. The search was performed using databases of functional domains, as implemented previously (Evangelisti et al., 2017), such as PANTHER, Pfam, Coils, Gene3D, Superfamily, Smart, Pirsf and Prints (Zdobnov and Apweiler, 2001). Based on the InterProScan output, we performed an analysis of GO (gene ontology) terms annotations. Graphs only show GO terms with a frequency above the 75 quantile of the GO terms count for the three ontology classes (biological process, molecular function and cellular component). Apoplastic effectors were predicted using ApoplastP (Sperschneider et al., 2018) using default parameters. For the prediction of candidate effectors having RxLR and CRN motifs, the `effectR` package was used (Tabima and Grünwald, 2019). Once the effector types were determined, we compared the effector profile obtained for *P. betacei* with the effectors reported for the other six *Phytophthora* species. obtaining a catalog of core effectors and the unique putative effectors present only in the *P. betacei* genome.

To validate our secretome predicted proteins, we collected the reported information on the effector profile of the six evaluated species and generated three control databases: (i) database with 1585 effector proteins predicted *in silico*, where 650 proteins are of CRN type (Armitage et al., 2018) and 935 are of RxLR type (Haas et al., 2009). (ii) database with 36 biologically validated proteins, where 16 are CRN-type effectors (Stam et al., 2013a) and 20 correspond to RxLR-type effectors (Wang et al., 2011). (iii). database with *P. betacei* transcripts (Accession number: XXX). Then, we screened our secretome predicted proteins against the databases (i) and (ii) using `blastp` (v. 2.6.2) with an e-value

threshold of $\leq 10^{-6}$. Only hits with an identity percentage greater than 90% and a coverage above 80% were considered.

2.3 Orthology analysis

Orthologs identification across all the tested *Phytophthora* species proteomes was performed using OrthoMCL (v. 2.0.9) (Li et al., 2003), with default parameters. Next, we selected only those orthologous groups associated with the secretome of each analyzed species. Additionally, upset plots showing the orthologous groups were generated using UpSetPlot library (v. 0.3.0.post3) (Lex et al., 2014). This analysis was also performed to identify orthologous groups of the RxLR and CRN effector proteins from *P. infestans* and *P. betacei*. We obtained a catalog of core effectors and the unique putative effectors present only in the *P. betacei* genome. We defined the core and unique effectors based on the orthologous groups obtained. The unique effectors correspond to singletons (proteins that did not cluster with any other protein from the other species compared). The core effectors correspond to proteins in an orthologous group where proteins from all the species compared are present.

2.4 Genomic context

Based on the hypothesis regarding the possible importance of gene distribution to identify new effector candidates, we performed a genomic context analysis using the methodology of Raffaele et al., (2010). We designed a custom script in which we implemented the following steps: (i) First, we identified the single-copy core orthologs using the results obtained in the orthology analysis. (ii) Second, to quantitatively evaluate the parameters that allowed segregation between Gene Dense Regions (GDRs) and Gene Sparse Regions (GSRs) in our data, we evaluated different values for the length cutoff (L). L was used to classify the lengths of flanking intergenic regions (FIRs), measured for each gene, as dense, sparse, or in between. For L values between 100bp and 6Kbp, we computed the percentage of core ortholog genes falling in GDRs and GSRs of the total genes located in both of these categories. Of the total core ortholog genes, we calculated the percentage of those genes classified as dense. (iii) We computed the segregation rate defined as the difference between the percentage of core ortholog genes in GDRs and the percentage of core ortholog genes in GSRs. (iv) To select the FIRs cutoff L that best classifies the data, we selected the one that maximizes the segregation rate and in which the percentage of core ortholog genes classified as either gene dense or in the boundaries of gene regions (in-between).

3. Results

3.1 Secretome prediction

We used an *in silico* secretome prediction pipeline to identify the putative secretome encoded by *P. betacei* (Figure 1). Subsequently, this secretome was compared to those obtained for the other analyzed *Phytophthora* species. The pipeline was implemented to predict signal peptides and cellular localization and to exclude proteins with internal transmembrane domains or an endoplasmic reticulum (ER) retention signal. For the seven species analyzed, secretomes ranged between 10.0% and 13.9% of the total proteome. *Phytophthora betacei* secretome was the largest one with a total of 5653 secreted proteins predicted, corresponding to 13.9% of its proteome (Figure 2; Supplementary Table S1); *P. palmivora* secretome was represented by a 10.5% of its total proteome; for *P. infestans* and *P. parasitica*, the secretome represented 11.4% and 10.8% respectively (Figure 2). Protein size distribution analysis of the *P. betacei* secretome revealed that all its proteins (5653) contained between 1999 and 27 amino acids (aa). In the case of *P. infestans* secreted proteins contained between 1562 and 49 aa (Supplementary Table S2). Regarding the seven species, *P. betacei*'s largest predicted protein was above the average of the largest predicted proteins of all the evaluated species. Similarly, this species has the shortest proteins in comparison with the six species evaluated.

3.2 Functional annotation of the secretome

For the functional annotation of *Phytophthora* secretome, we detected the protein domains using Pfam (v. 32.0) and Superfamily (v. 1.73) HMM model databases and Gene Ontology (GO) terms mapping using InterProScan. A total of 22226 secreted proteins were obtained for the seven *Phytophthora* species evaluated. Of these secreted proteins we found that 13704 proteins had at least one predicted domain, and of these, 7234 proteins were associated to at least one GO term (Figure 3). The GO term categorization of the secreted proteins revealed that *P. betacei* had a greater number of proteins in all GO categories than the other analyzed *Phytophthora* species. The “molecular function” category for *P. betacei* presented a greater number of entries when compared to the other six species, particularly to *P. infestans* which showed a lower number of proteins in each GO category with respect to all the evaluated species (Figure 3). As it was the case for the proportion of secreted proteins relative to the whole proteome (Figure 2A), the proportions of proteins in each GO categories were similar for *P. palmivora* and *P. betacei* (Figure 3).

We found 52 “biological process” ontologies, 12 “cellular component” ontologies and 84 “molecular function” ontologies shaping the secretome profile of *P. betacei*. Proteolysis (GO:0006508) and carbohydrate metabolic processes (GO:0005975) showed the highest count among biological processes ontologies in the *P. betacei* secretome compared to the rest of the proteomes (Figure 4A; Supplementary Table S3). Other related biological processes are related to Interference with host cell signaling pathways (protein phosphorylation (GO:0006468)) and protein degradation for pathogenesis (GO:0009405). For the cellular component category, organelle and membrane terms were dominant (Figure 4B). The terms that were direct descendants (child terms) of Extracellular region (GO:0005576) were related to apoplast (GO:0048046). Regarding “molecular function” terms, genes related to binding (GO:0005515; GO:0005524) and catalytic activity (GO:0003824) were highly represented. Remaining terms indicate that *P. betacei* uses plant cell wall

degradation activity (cellulase activity (GO:0008810), polygalacturonase activity (GO:0004650), pectate lyase activity (GO:0030570), pectinesterase activity (GO:0030599), and hydrolase activity (GO:0004553)). Another important molecular function ontology is the endopeptidase inhibitor (GO:0004867), corresponded to Kazal-like serine protease inhibitors (Figure 4C).

3.3 The secretome from *P. betacei* revealed both Conserved and Unique Effector Domains

In order to identify the effector proteins in the *P. betacei* secretome, we performed an effector domain search for apoplastic and cytoplasmic effectors. Out of the 5653 secreted proteins found in *P. betacei*, 1126 were found to be apoplastic effectors (Figure 5), corresponding to 19.9% of the whole secretome. Comparing with the other *Phytophthora* species evaluated, apoplastic effectors in *Phytophthora infestans* represented 26.8% of its secretome, and *P. ramorum* showed the higher proportion with 36.9% of the secretome represented by apoplastic effectors, while *P. cactorum* showed 27.2% of these effectors in its secretome.

For *P. betacei*, 751 (59.6 %) of the predicted apoplastic effectors had at least one motif detected by InterProScan search. Of these proteins, 231 (31%) correspond to biological process, 93 (12.%) to cellular component and 427 (54.7%) to molecular function (Supplementary Table 4). Among them, the most representatives are extracellular region (GO:0005576) for cellular component, carbohydrate metabolic process (GO:0005975) for biological process and protein binding (GO:005515) for molecular function (Figure 6A). In addition, we found GO terms associated to infection processes that are highly represented in the set of the predicted apoplastic effectors, such as, pathogenesis and proteolysis. Regarding the blast annotation strategies, we obtained that 538 (47%) out of the 1126 putative apoplastic effectors were annotated when mapped against non-redundant (nr) database. Of these, pectin lyase, protease inhibitors, cutinases, Phytotoxic protein (PcF) and necrosis peptide-like proteins (NLPs) domain-carrying proteins previously reported in different *Phytophthora* species were identified. Additionally, we detected elicitors, which can mediate sterol obtention during pathogenesis as well as recognition by some plant cells. On the other hand, we found a high amount of predicted proteins that match with hypothetical proteins related to *Phytophthora* species. It would be interesting to determine the function of these proteins in the future.

As mentioned before, cytoplasmic effectors such as RxLR and CRNs are proteins consisting of an N-terminal carrying the conserved domains. For *P. betacei*, we obtained 807 predicted cytoplasmic effectors (12.4% of the secretome) using the effectR package. Of these, 791 correspond to RxLR and 16 to the CRN type (Figure 7). In the alignments of the N-terminal regions of RxLR and CRN effectors, both RxLR_EER and LFLAK_HVL motifs were found (Figure 8A and Figure 8B, respectively). *Phytophthora infestans* showed the highest proportion of cytoplasmic effectors in its secretome, with 32.0%, while *P. ramorum* showed the lowest one with 10.5% (Supplementary Table 1). As expected the majority of N-terminal sequences on the predicted proteins showed a high similarity

(Figure 8A) as evidenced in multiple sequence alignment of such effectors. The C-terminal regions, on the other hand, are highly diverse as they are thought to specify effector biochemical functions (Figure 8A and 8B).

The putative RxLR effectors were compared among the seven *Phytophthora* species in order to obtain the RxLR core effectors (*P. palmivora* was not included because the predicted effectors were obtained from a transcriptome - see methods). We obtained 10 RxLR core effectors (Figure 9A) in which the first conserved regions in proteins were found to be situated at the N-terminus, featuring a highly conserved RxLR motif. In addition, we observed that the most frequent amino acids at the x position of the effector motif, were serine, leucine, arginine and lysine (Figure 8A).

With respect to the GO terms annotation of the cytoplasmic effectors, in *P. betacei*, 25 (3.2 %) of the predicted RxLR effectors had at least one detected motif in the InterProScan search (Supplementary Table 4). Of these proteins, six effectors (24%) were related to biological processes, two (8%) to cellular components and 17 (68 %) to molecular functions. The most representative GO terms included were hydrolase activity (GO:0016787) for molecular function, nucleoside metabolic process (GO:0009116) and protein glycosylation (GO:0006486) for biological process and two Go terms associated to cellular component, that are, ribosome (GO:0005840) and integral component of membrane (GO:0016021) (Figure 6B). In addition, the domains Zinc finger FLYWCH-type and PexRD2 WYL were registered as the most frequent ones (Figure 8C). Regarding the blast annotation strategies, we obtained that 322 (40.7 %) out of the 791 RxLR effectors were annotated when mapped against non-redundant (nr) database. Of these, in the functional characterization of these RxLR core effector, homologs of genes such as Avr1, Avr3b, Avr4, Avrblb1, Avrblb2 and Avrvt1 were detected, indicating a high level of conservation of RxLR effector genes in *Phytophthora* species (Supplementary Table 4). For both the apoplastic and cytoplasmic effectors predicted for *P. betacei* we found a high amount of hypothetical proteins related to *Phytophthora* species.

On the other hand, we found just one CRN effector shared among all the six species compared (Figure 9B). This CRN core effector featured an Ubiquitin-like (Ubl) domain that is thought to be responsible for secretion and translocation into the host cell. The majority of *Phytophthora* CRN C-terminal regions contained the depicted domain structure (NTPase + HTH + REase) (Figure 8D). Domains as Ubiquitin-like (Ubl_SNRNP25), HAD hydrolase, TIGR01548 family and Elicitin also were found (Supplementary Table 4).

To define core and unique genes between *P. betacei* and *P. infestans*, apoplastic and cytoplasmic effectors in both species were clustered into gene families based on the orthology analysis. In total, 349 apoplastic secreted proteins were predicted for *P. betacei*. Of them, 152 are shared with *P. infestans*, and 197 were considered unique, representing 56% of the predicted apoplastic effectors of *P. betacei* (Figure 9C). Comparison between cytoplasmic effectors of *P. betacei* and *P. infestans*, showed that 62.5% (203) corresponded to *P. betacei* unique RxLR type effectors and 37.5% (122) corresponded to effectors shared between both species (Figure 9D). In the case of CRN type effectors, of 16 predicted CRN effectors in *P. betacei*, two were categorized as unique to *P. betacei* representing the 12.5% of the CRN effectors predicted in our species of interest and seven were shared with *P.*

infestans (Figure 9E). When performing blastp searches of the effectors categorized as unique of *P. betacei*, most of them show alignments with very low identity percentages and query coverage with hypothetical proteins from other *Phytophthora* species (Supplementary Table 4).

From a total of 205 unique predicted cytoplasmic effector proteins (203 RxLR, 2CRN), 156 mapped on the *P. betacei* transcriptome (accession # XX), of which two of them corresponded to CRN type effectors. A total of 47 out of the 156 predicted effectors expressed in *P. betacei* had at least one annotation based on InterProScan results. The other predicted effectors corresponded to unknown proteins or hypothetical proteins that had not been characterized or linked to known genes.

3.4 Genomic distribution of *P. betacei* effectors

We identified 1600 core orthologs for all the sampled species. We established a FIRs cutoff length $L = 1.3\text{Kbp}$ based on the criteria established (see methods). At this cut-off, 7.6% of the core orthologs genes were assigned to GSRs, leaving 58.8% of the core orthologs genes assigned to GDRs and a remaining 33.6% of the single-copy core orthologs genes assigned to In-Between regions. At the established cutoff, 6.2% of all the genes in GDRs were core orthologs, 3.1% of all the genes in In-Between regions were core orthologs and, 1.0% of the genes in GSRs were core orthologs (Figure 10A). The GDRs contain 14924 genes corresponding to 29.8% of *P. betacei* genes. The GSRs contain 11564 genes representing 26.6% of *P. betacei* genes. The 16856 genes assigned to the left-superior and right-inferior quadrants correspond to 38.8% of *P. betacei* genes (Figure 10B). The rest of the genes (4.8% of *P. betacei*) were excluded either because they lacked one FIR (located at one border of the scaffold) or they overlapped with other genes. GSRs contain 34.5% of *P. betacei* genes identified in this study as possible effectors. Also 65.5% of the *P. betacei* genes identified as unique RxLR fall in GSRs. This represents an enrichment consistent with previous analysis for other *Phytophthora* species. In particular, genes identified as unique RxLR present a 2.5-fold enrichment compared with the percentage of total genes assigned to GSRs regions. All the genes classified as unique CRN effectors of the species were assigned to In-Between regions, and no CRN effectors were found in GDRs (Figure 10 C). We selected 114 apoplastic and 74 RxLR *P. betacei* unique effectors located in GSR regions to be considered as possible virulence factors (Supplementary Table 6).

4. Discussion

In this study, we have generated, identified and carefully selected the *P. betacei* secretome to predict the effector profile of this species. We have also identified its main similarities and differences with several *Phytophthora* species and in particular, its sister species *P. infestans*. To accomplish our main objective, we compared different species of the *Phytophthora* genus and generated the catalog of unique and core effectors for each of them. The unique effectors in *P. betacei* will allow us in the near future to experimentally assess if they are involved in host specificity. It is interesting to highlight that, while there

are important differences in the number of total proteins among species, the number of secreted proteins showed less differences.

Our results showed that *P. betacei* is the species with the largest proteome and the highest proportion of secretome with respect to its proteome. This species also presents a large number of unique apoplastic and cytoplasmic effectors. These effectors are highly diverse in their sequences and most of them are localized in gene sparse regions (GSR) in the genome. These characteristics suggest that they might be considered as strong candidate virulence factors. Previously, gene distribution (Raffaele et al., 2010; Haas et al., 2013) and gene diversity (Lee et al., 2006; Win et al., 2007; Schornack et al., 2009) were proposed as two main characteristics of effector proteins in *P. infestans*, the sister species of *P. betacei*. In *P. infestans* most known effector genes were found to be in GSRs and presented C-Terminal polymorphic regions, as we found for *P. betacei* in this study (Figures 8A and Figure 8B).

Phytophthora species utilize a diverse range of secreted apoplastic and cytoplasmic effectors to support the infection mechanisms. Effector prediction from genomes or transcriptomes of *Phytophthora* in previous studies have revealed variable numbers of effectors in different species (Jiang et al., 2008; Haas et al., 2009; Evangelisti et al., 2017; Armitage et al., 2018). Our results agree with previous annotations of diverse *Phytophthora* species tested (Supplementary, Table S5). We identified a considerably greater number of apoplastic effector candidates (1126) in *P. betacei* than in the other *Phytophthora* species evaluated. Equal or greater numbers of genes encoding elicitors, NLPs, protease inhibitors, cutinases and PcF domain-carrying proteins were identified, the main apoplastic effectors involved in the virulence in *P. infestans*, *P. sojae* and *P. cactorum* (Gough et al., 2001; Tian and Kamoun, 2005; Raffaele et al., 2010; Resjö et al., 2017)

Cytoplasmic effectors are defined as those that function specifically within host cells. The past decade of genome sequencing has allowed the identification of hundreds of candidate cytoplasmic effectors in *Phytophthora* genomes. The secretome of *P. betacei* revealed 791 candidate RxLR while for *P. infestans* more than 550 effector genes were predicted (Haas et al., 2009). Proteins in shared ortholog groups between *P. betacei* and *P. infestans* allowed identification of RxLR effector genes fulfilling the typical features of oomycete Avr genes as they encode for modular proteins that contain an N-terminal signal peptide followed by an RxLR motif and a C-terminal effector domain, and occur in gene-sparse regions of the *P. infestans* genome, these effector genes represent key targets for further functional studies.

The CRN protein family is an understudied class of oomycete effectors, difficult to identify and classify. Thus, CRN proteins have hampered functional studies in different oomycetes. Here, we applied a pipeline of CRN prediction, manual verification and mapping studies in catalogs of effectors already reported for the *Phytophthora* species evaluated. This approach improved CRN identification and prediction accuracy for both full length genes and pseudogenes, compared to previously published results (Haas et al., 2009; Zhang et al., 2019). We identified 16 CRN effectors in *P. betacei* and when compared to *P. infestans*, two unique effectors for *P. betacei* were found. Existing descriptions of CRN domain composition and structure, allowed us to identify conserved

domains in all *Phytophthora* species analyzed, but also to identify unique ones. One of the two unique *P. betacei* CRN effector (P8084_finalAssembly_32667) has a domain that was previously described in *P. capsici*, as a novel domain called DPA that may have specific roles but that had not been found so far in other oomycetes (Haas et al., 2009; Stam et al., 2013a). This effector could be proposed as a novel type of effector for the clade, and its relevance in pathogenicity needs to be validated. Hass et al., 2009 determined that CRN proteins are modular with domains that execute distinct functions, with a highly conserved N-terminal domain of around 130 amino acids and containing both an LFLAK motif and diversified DWL domains. In our study, the highly conserved HVLVXXP motif marks the end of the N-terminal region as it is considered a recombination hotspot where C-terminal regions, carrying effector functions are linked up (Figure 5A) (Haas et al., 2009). In our annotation the unique effectors presented a conserved domain corresponding to a ubiquitin-like (Ubl) domain with a beta-grasp Ubl fold, a common structure involved in protein-protein interactions (Amaro et al., 2017).

Effector genes have previously been characterized as showing uneven distributions throughout *Phytophthora* genomes, with measurements of flanking intergenic regions (FIRs) showing that effector genes are located in gene-sparse regions of the *P. infestans* genome (Haas et al., 2009) and the same scenario was observed for *P. betacei*. 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 (Dong et al., 2015). The predicted apoplastic effectors, RxLR and CRN, from *P. betacei* showed increased FIRs distance compared with non-effector genes, supporting that genomic variation in *P. betacei* shows uneven "two-speed" evolutionary rates with the presence of GSRs and GDRs, as previous studies indicated in several *Phytophthora* genomes (Raffaele and Kamoun, 2012; Dong et al., 2015). Unique *P. betacei* effectors that were found to be located in GSRs might be helping the pathogen to colonize and might therefore help explain the higher fitness of this species on tree tomato when compared to its sister species *P. infestans* (Mideros et al., 2018). On the other hand, the unique effectors candidates might also be recognized by potato and tomato and hence could be involved in the lack of ability of *P. betacei* to infect these hosts. Alternatively, the effectorome of *P. betacei* might not be sufficiently effective against crucial targets in tomato and potato cells. Nonetheless functional assays must be performed to validate these hypotheses.

In summary, in the present study we predicted and annotated the putative effector profile of *P. betacei*. We found seven characteristics representative of the *P. betacei* secretome when compared to other *Phytophthora* species: (i). The *P. betacei* proteome has a higher number of proteins. (ii). The proportion of the secretome compared to the whole proteome is higher (13.9%) (iii). Secreted proteins are smaller. (iv). This novel species presents a higher number of RxLR type effectors. (v). The number of cytoplasmic CRN type effectors is lower (vi). One of the CRN proteins show a domain so far only reported for *P. capsici* (vii). It has 203 cytoplasmic RxLR effectors and 2 unique CRN effectors, and they are polymorphic in their C-terminal region. These characteristics suggest that *P. betacei*, produces a greater number of effectors. Unique effectors that are found in the GSR zones of the *P. betacei* genome are proposed as virulence factors candidates involved in its host specificity, 114 apoplastic and 74 RxLR effectors. These effectors might make it more efficient at colonizing tree tomato when compared to closely related *Phytophthora* species.

These effectors might also explain the lack of ability of this species to colonize potato and tomato through recognition by R genes in these plants, which might in turn serve as resistance sources for tree tomato.

5. Acknowledgments

We want to acknowledge the High-Performance Computing Service at Universidad de los Andes, Bogotá-Colombia for contributing to the research development providing us with HPC resources.<http://hpc.uniandes.edu.co>. Finally, we thank LAMFU and Computational Biology and Microbial Ecology (BCEM) groups at the Universidad de los Andes, Bogotá-Colombia.

6. Author Contributions Statement

PR, DU, SR, AB designed the analyses strategies; PR, DU: Performed the analyses. DA, NG, MFM, contributed to the analyses. PR, DA, AB, MC, SR: wrote the manuscript. All authors contributed to manuscript revision, read and approved the submitted version.

7 Conflict of Interest Statement

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

8 References

- Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., et al. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389–402.
- Amaro, T. M. M. M., Thilliez, G. J. A., Motion, G. B., and Huitema, E. (2017). A Perspective on CRN Proteins in the Genomics Age: Evolution, Classification, Delivery and Function Revisited. *Front. Plant Sci.* 8, 99.
- Armitage, A. D., Lysøe, E., Nellist, C. F., Lewis, L. A., Cano, L. M., Harrison, R. J., et al. (2018). Bioinformatic characterisation of the effector repertoire of the strawberry pathogen *Phytophthora cactorum*. *PLoS One* 13, e0202305.
- Bendtsen, J. D., Nielsen, H., Von Heijne, G., and Brunak, S. (2004). Improved prediction of signal peptides: SignalP 3.0. *J. Mol. Biol.* 340, 783–795.
- Dong, S., Raffaele, S., and Kamoun, S. (2015). The two-speed genomes of filamentous pathogens: Waltz with plants. *Curr. Opin. Genet. Dev.* 35, 57–65.
- Emanuelsson, O., Nielsen, H., Brunak, S., and von Heijne, G. (2000). Predicting Subcellular Localization of Proteins Based on their N-terminal Amino Acid Sequence. *J. Mol. Biol.* 300, 1005–1016.
- Evangelisti, E., Gogleva, A., Hainaux, T., Doumane, M., Tulin, F., Quan, C., et al. (2017). Time-resolved dual transcriptomics reveal early induced *Nicotiana benthamiana* root genes and conserved infection-promoting *Phytophthora palmivora* effectors. *BMC Biol.* 15, 1–24.
- Fry, W. (2008). *Phytophthora infestans*: the plant (and R gene) destroyer. *Mol. Plant Pathol.* 9, 385–402.
- Gogleva, A., Drost, H.-G., and Schornack, S. (2018). SecretSanta: flexible pipelines for functional secretome prediction. *Bioinformatics* 34, 2295–2296. doi:10.1093/bioinformatics/bty088.
- Gough, J., Karplus, K., Hughey, R., and Chothia, C. (2001). Assignment of homology to genome sequences using a library of hidden Markov models that represent all proteins of known structure. *J. Mol. Biol.* 313, 903–19.
- Grünwald, N. J., and Flier, W. G. (2005). The Biology of *Phytophthora infestans* at Its Center of Origin. *Annu. Rev. Phytopathol.* 43, 171–190.
- Guayazan, N., Mideros, M. F., Gogleva, A., Schornack, S., and Restrepo, S. (2017). Characterization of the infection cycle of *Phytophthora betacei* during disease development on tree tomato (*Solanum betaceum*). [dissertation/master's thesis]. [Bogotá]: Universidad de Los Andes.

- Haas, B. J., Kamoun, S., Zody, M. C., Jiang, R. H. Y., Handsaker, R. E., Cano, L. M., et al. (2009). Genome sequence and analysis of the Irish potato famine pathogen *Phytophthora infestans*. *Nature* 461, 393–398.
- Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., et al. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* 8, 1494–1512.
- Haldar, K., Kamoun, S., Hiller, N. L., Bhattacharje, S., and van Ooij, C. (2006). Common infection strategies of pathogenic eukaryotes. *Nat. Rev. Microbiol.* 4, 922–931.
- Hardham, A. R. (2001). The cell biology behind *Phytophthora* pathogenicity. *Australas. Plant Pathol.* 30, 91.
- Holt, C., Yandell, M. (2011). MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC bioinformatics*, 12, 491.
- Jiang, R. H. Y., Tripathy, S., Govers, F., and Tyler, B. M. (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. U. S. A.* 105, 4874–4879.
- Jones, P., Binns, D., Chang, H. Y., Fraser, M., Li, W., McAnulla, C., et al. (2014). InterProScan 5: Genome-scale protein function classification. *Bioinformatics* 30, 1236–1240.
- Kamoun, S. (2006). A Catalogue of the Effector Secretome of Plant Pathogenic Oomycetes. *Annu. Rev. Phytopathol.* 44, 41–60.
- Kamoun, S. (2007). Groovy times: filamentous pathogen effectors revealed. *Curr. Opin. Plant Biol.* 10, 358–365.
- Krogh A, Larsson B, von Heijne G, and Sonnhammer EL. 2001. Predicting transmembrane protein topology with a hidden markov model: application to complete genomes. *Journal of Molecular Biology* 305, 567–580..
- Lee, S.-J., Kelley, B. S., Damasceno, C. M. B., St John, B., Kim, B.-S., Kim, B.-D., et al. (2006). A functional screen to characterize the secretomes of eukaryotic pathogens and their hosts in planta. *Mol. Plant. Microbe. Interact.* 19, 1368–77.
- Lex, A., Gehlenborg, N., Strobel, H., Vuillemot, R., and Pfister, H. (2014). UpSet: Visualization of intersecting sets. *IEEE Trans. Vis. Comput. Graph.* 20, 1983–1992.
- Li, L., Stoeckert, C. J., and Roos, D. S. (2003). OrthoMCL: Identification of Ortholog Groups for Eukaryotic Genomes. *Genome Res.* 13, 2178–2189.

- Li, W., and Godzik, A. (2006). Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22, 1658–1659.
- Mideros, M. F., Turissini, D. A., Guayazán, N., Ibarra-Avila, H., Danies, G., Cárdenas, M., et al. (2018). *Phytophthora betacei*, a new species within *Phytophthora* clade 1c causing late blight on *Solanum betaceum* in Colombia. *Persoonia - Mol. Phylogeny Evol. Fungi*.
- Nielsen, H., and Krogh, A. (1998). Prediction of signal peptides and signal anchors by a hidden Markov model. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* 6, 122–130.
- Ntoukakis, V., Mucyn, T. S., Gimenez-Ibanez, S., Chapman, H. C., Gutierrez, J. R., Balmuth, A. L., et al. (2009). Host Inhibition of a Bacterial Virulence Effector Triggers Immunity to Infection. *Science*. 324, 784–787.
- Oh, S.-K., Young, C., Lee, M., Oliva, R., Bozkurt, T. O., Cano, L. M., 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 online* 21, 2928–2947. doi:10.1105/tpc.109.068247.
- Petersen, T. N., Brunak, S., Von Heijne, G., and Nielsen, H. (2011). SignalP 4.0: Discriminating signal peptides from transmembrane regions. *Nat. Methods* 8, 785–786.
- Raffaele, S., and Kamoun, S. (2012). Genome evolution in filamentous plant pathogens: why bigger can be better. *Nat. Rev. Microbiol.* 10, 417–430.
- Raffaele, S., Win, J., Cano, L. M., and Kamoun, S. (2010). Analyses of genome architecture and gene expression reveal novel candidate virulence factors in the secretome of *Phytophthora infestans*. *BMC Genomics* 11, 637.
- R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Resjö, S., Brus, M., Ali, A., Meijer, H. J., Sandin, M., Govers, F., and Andreasson, E. (2017). Proteomic analysis of *Phytophthora infestans* reveals the importance of cell wall proteins in pathogenicity. *Mol. Cell. Proteomics*, 16, 1958-1971.
- Schoebitz, M., Ceballos, C., and Ciampi, L. (2013). Effect of immobilized phosphate solubilizing bacteria on wheat growth and phosphate uptake. 13, 1–10.
- Schorneck, S., Huitema, E., Cano, L. M., Bozkurt, T. O., Oliva, R., Van Damme, M., et al. (2009). Ten things to know about oomycete effectors. *Mol. Plant Pathol.* 10, 795–803.
- Schorneck, S., Van Damme, M., Bozkurt, T. O., Cano, L. M., Smoker, M., Thines, M., et al. (2010). Ancient class of translocated oomycete effectors targets the host nucleus. *Proc. Natl. Acad. Sci. U. S. A.* 107, 17421–17426.

- Sperschneider, J., Dodds, P. N., Singh, K. B., and Taylor, J. M. (2018). ApoplastP: prediction of effectors and plant proteins in the apoplast using machine learning. *New Phytol.* 217, 1764–1778.
- Stam, R., Jupe, J., Howden, A. J. M., Morris, J. A., Boevink, P. C., Hedley, P. E., et al. (2013a). Identification and Characterization CRN Effectors in *Phytophthora capsici* Shows Modularity and Functional Diversity. *PLoS One* 8, e59517.
- Stam, R., Howden, A. J. M., Delgado Cerezo, M., Amaro, T. M. M. M., Motion, G. B., Pham, J and Huitema, E. (2013b). Characterization of cell death inducing *Phytophthora capsici* CRN effectors suggests diverse activities in the host nucleus. *Frontiers in plant science*, 4, 387.
- Tabima, J. F., and Grünwald, N. J. (2019). effectR: An Expandable R Package to Predict Candidate RxLR and CRN Effectors in Oomycetes Using Motif Searches. *Mol. Plant. Microbe. Interact.* 32, 1067–1076.
- Tian, M., and Kamoun, S. (2005). A two disulfide bridge Kazal domain from *Phytophthora* exhibits stable inhibitory activity against serine proteases of the subtilisin family. *BMC biochemistry*, 6, 15.
- Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D. R., et al. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.* 7, 562–578.
- Vleeshouwers, V. G. A. A., Rietman, H., Krenek, P., Champouret, N., Young, C., Oh, S.-K., et al. (2008). Effector Genomics Accelerates Discovery and Functional Profiling of Potato Disease Resistance and *Phytophthora infestans* Avirulence Genes. *PLoS One* 3, e2875.
- Wang, Q., Han, C., Ferreira, A. O., Yu, X., Ye, W., Tripathy, S., et al. (2011). Transcriptional programming and functional interactions within the *Phytophthora sojae* RXLR effector repertoire. *Plant Cell* 23, 2064–2086.
- Whisson, S. C., Boevink, P. C., Moleleki, L., Avrova, A. O., Morales, J. G., Gilroy, E. M., et al. (2007). A translocation signal for delivery of oomycete effector proteins into host plant cells. *Nature* 450, 115–118.
- Win, J., Morgan, W., Bos, J., Krasileva, K. V., Cano, L. M., 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.
- Yang, X., Tyler, B. M., and Hong, C. (2017). An expanded phylogeny for the genus *Phytophthora*. *IMA fungus*, 8, 355.

Zdobnov, E. M., and Apweiler, R. (2001). InterProScan--an integration platform for the signature-recognition methods in InterPro. *Bioinformatics* 17, 847–8.

Zhang, X., Liu, B., Zou, F., Shen, D., Yin, Z., Wang, R., et al. (2019). Whole Genome Re-sequencing Reveals Natural Variation and Adaptive Evolution of *Phytophthora sojae*. *Front. Microbiol.* 10.

9 Figure

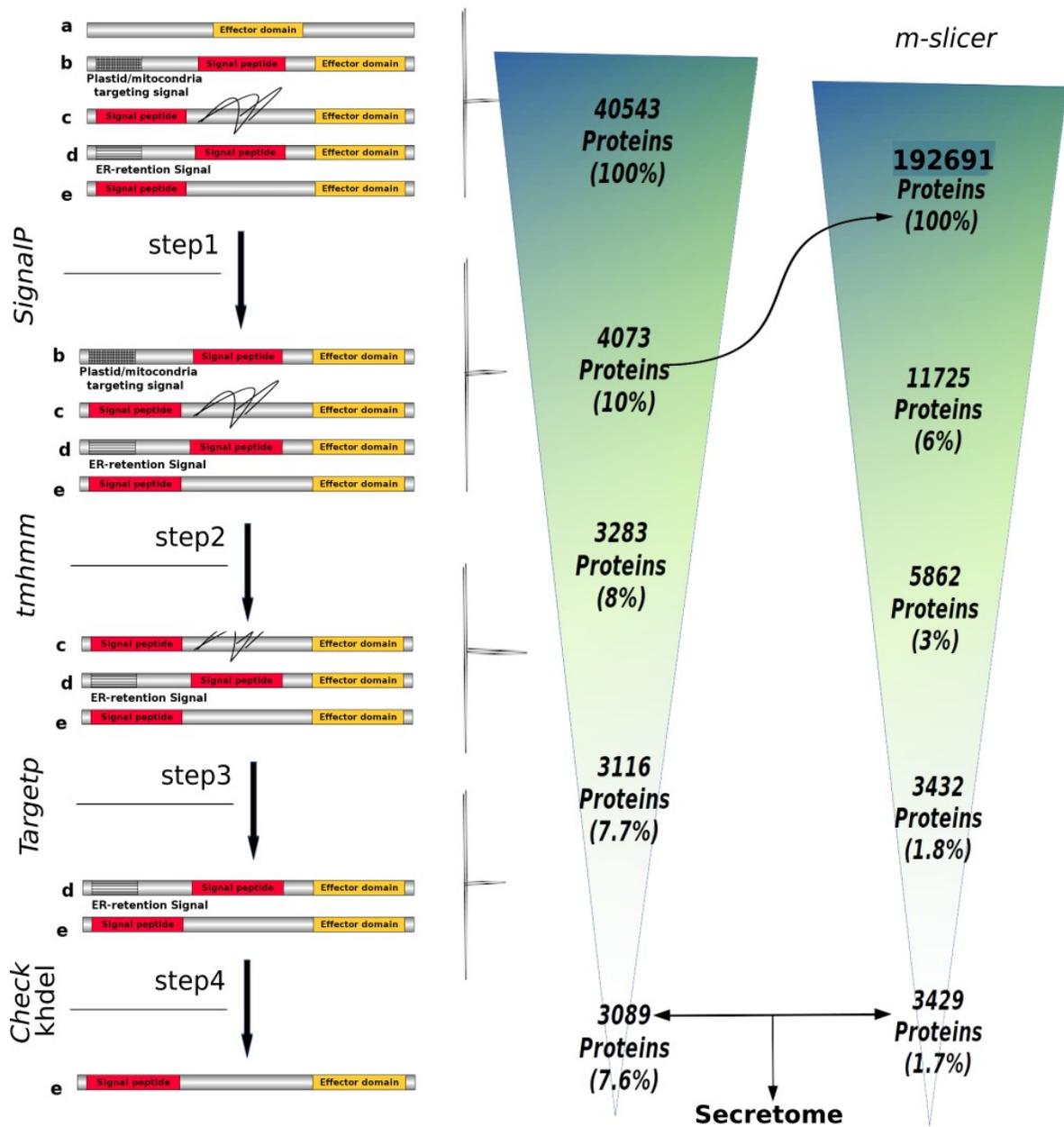


Figure 1. Number of proteins obtained throughout the SecretSanta pipeline used for the prediction of the *Phytophthora betacei* secretome.

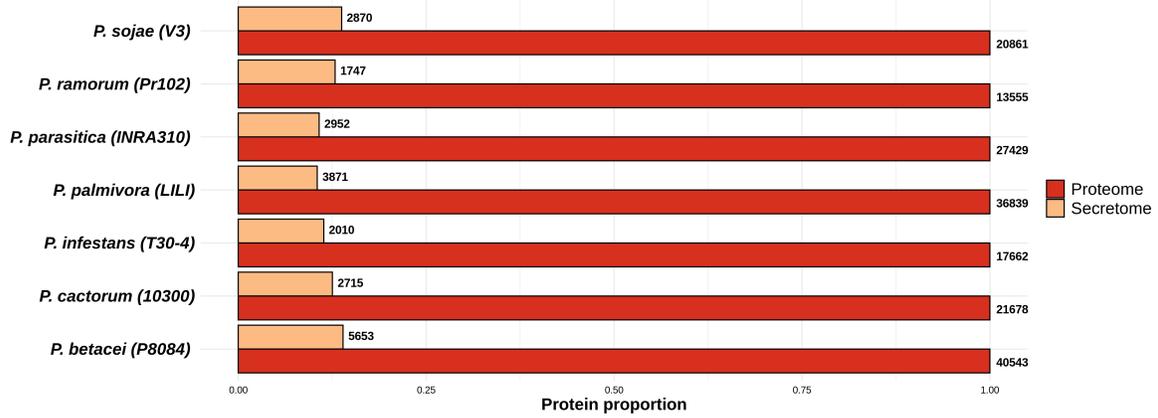


Figure 2. Proportion of the predicted secretomes relative to the proteomes of the *Phytophthora* species evaluated.

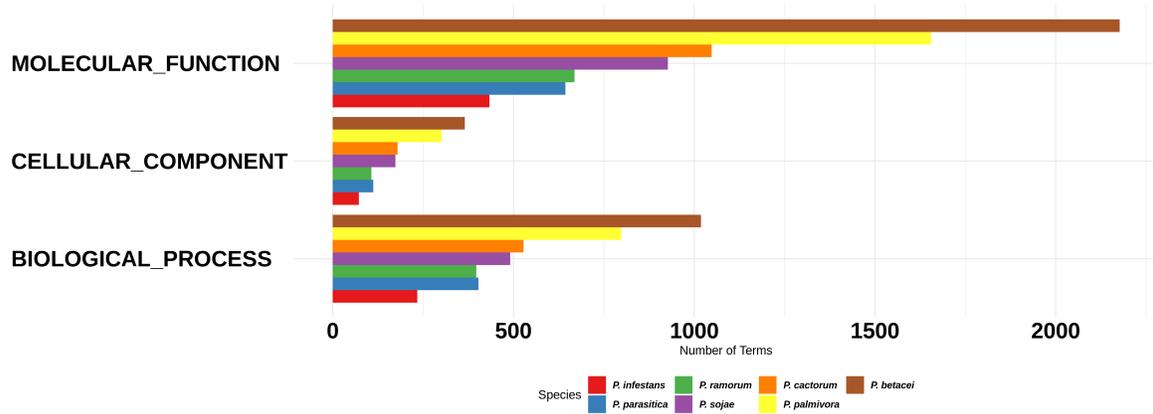


Figure 3. Gene Ontology categories for the *Phytophthora* species secretome.

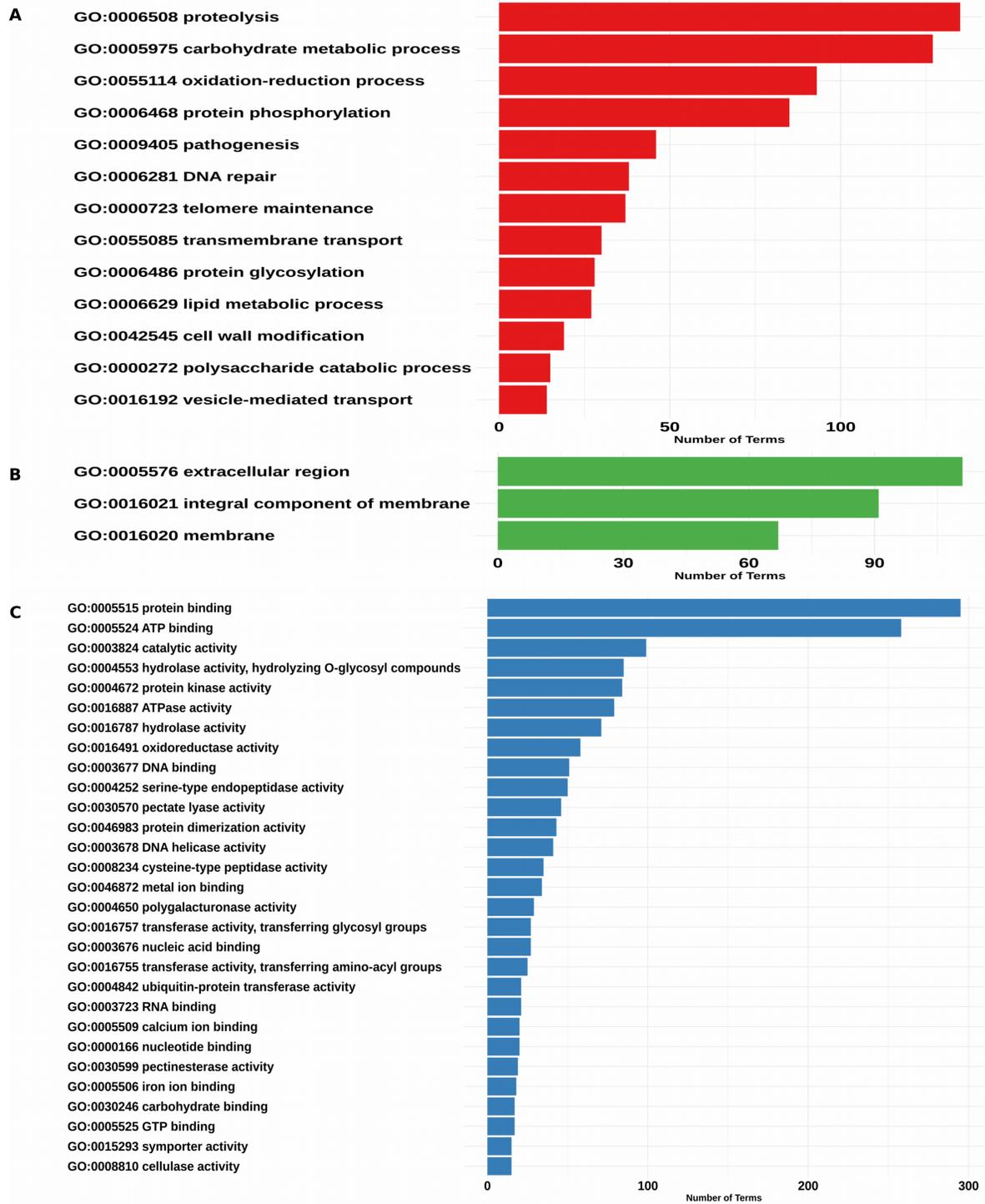


Figure 4. Assignment of GO terms in the *Phytthora betacei* secretome.

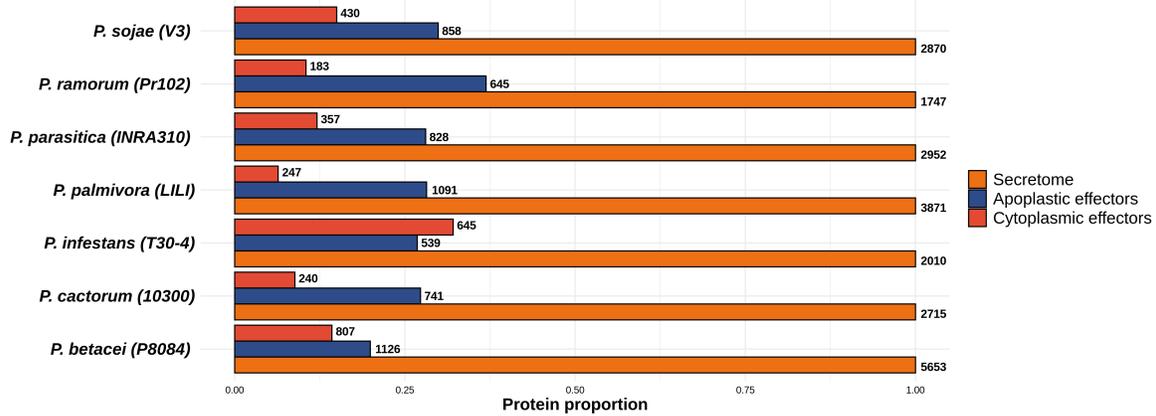


Figure 5. Proportion of predicted effectors relative to the predicted secretome of the *Phytophthora* species evaluated.

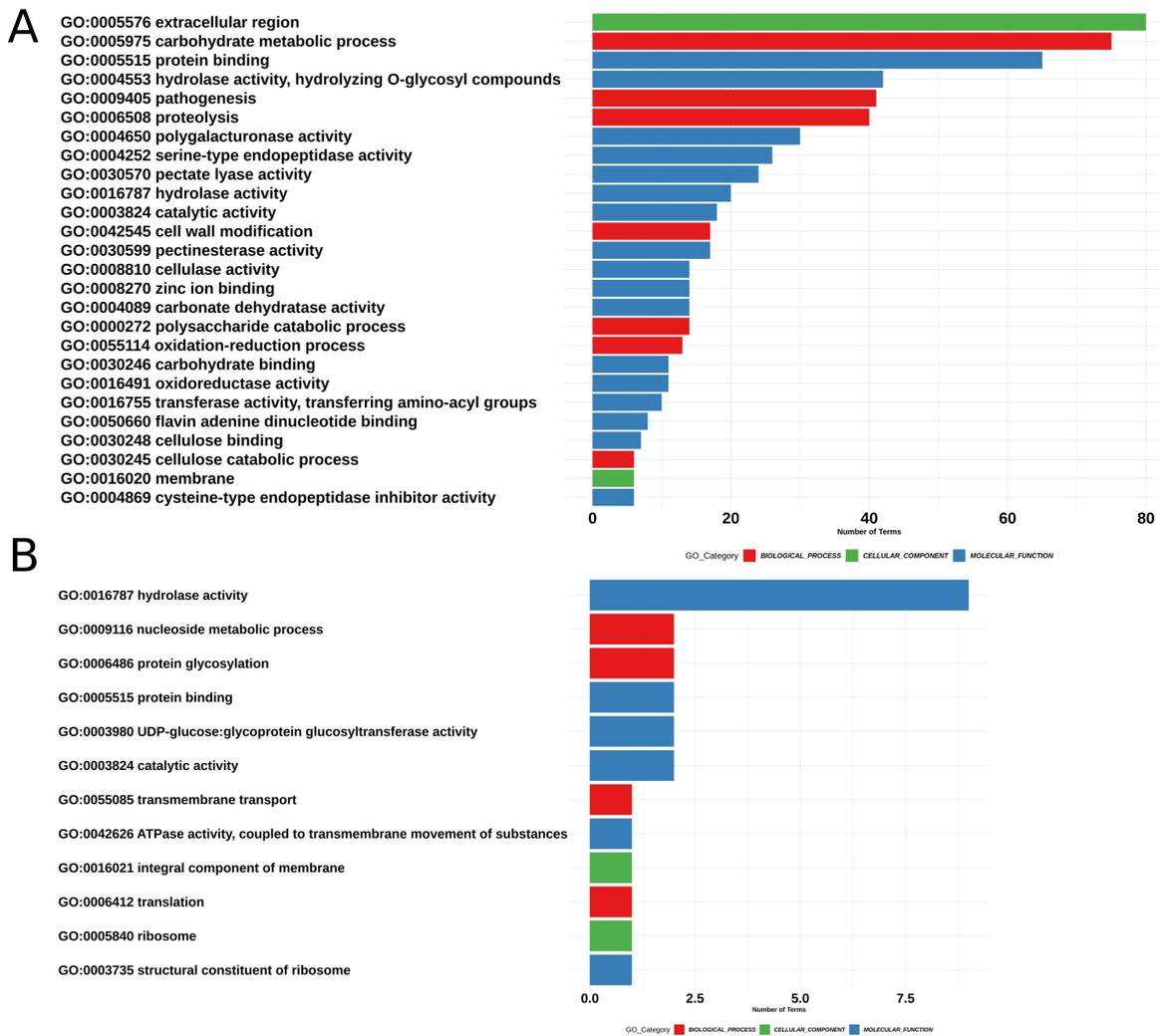


Figure 6. Assignment of GO terms in the *Phytophthora betacei* effectors.

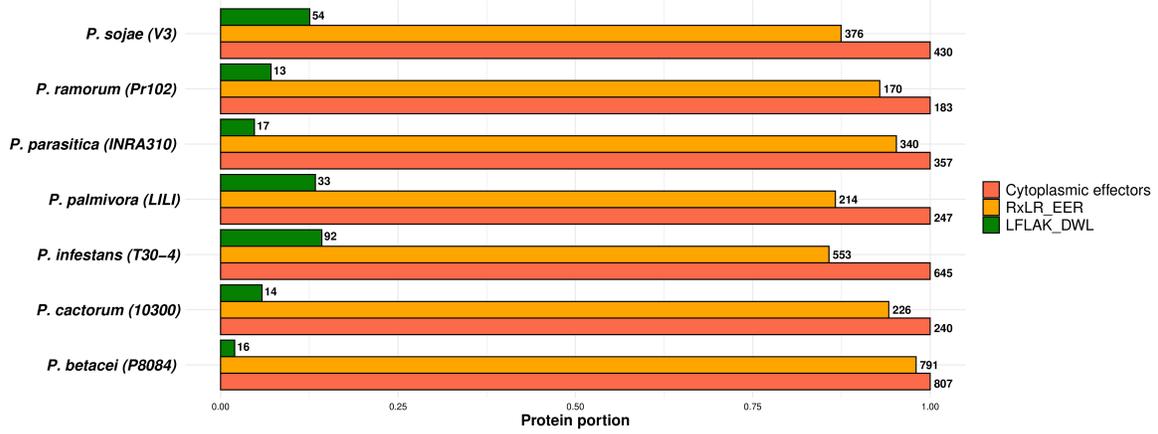


Figure 7. Proportion of effectors with RxLR_EER and LXLFLAK_DWL (CRNs) motifs relative to the total amount of predicted cytoplasmic proteins.

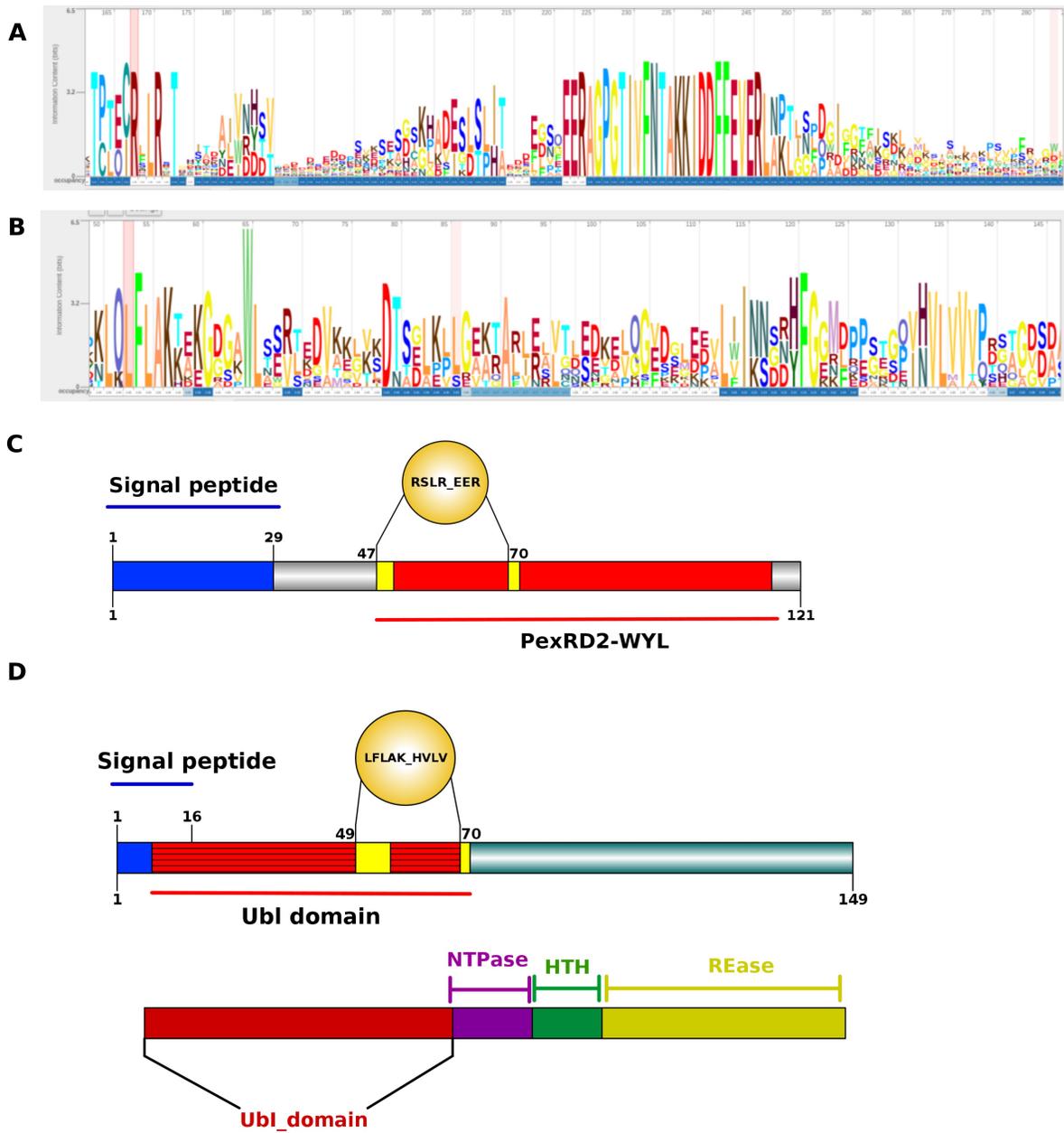


Figure 8. Structure analysis of *Phytophthora betacei* cytoplasmic effectors.

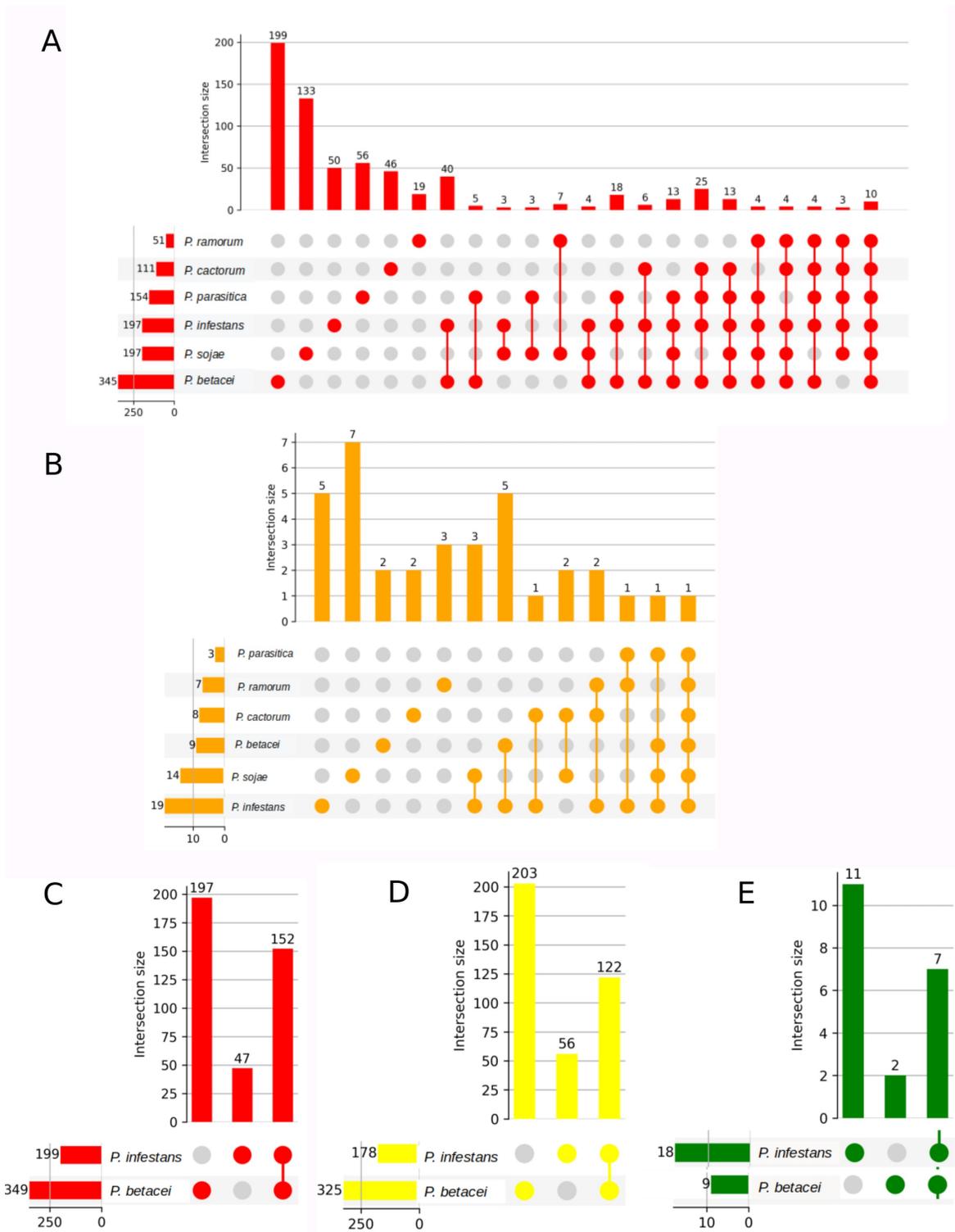


Figure 9. Number of unique ortholog effectors in *Phytophthora* species.

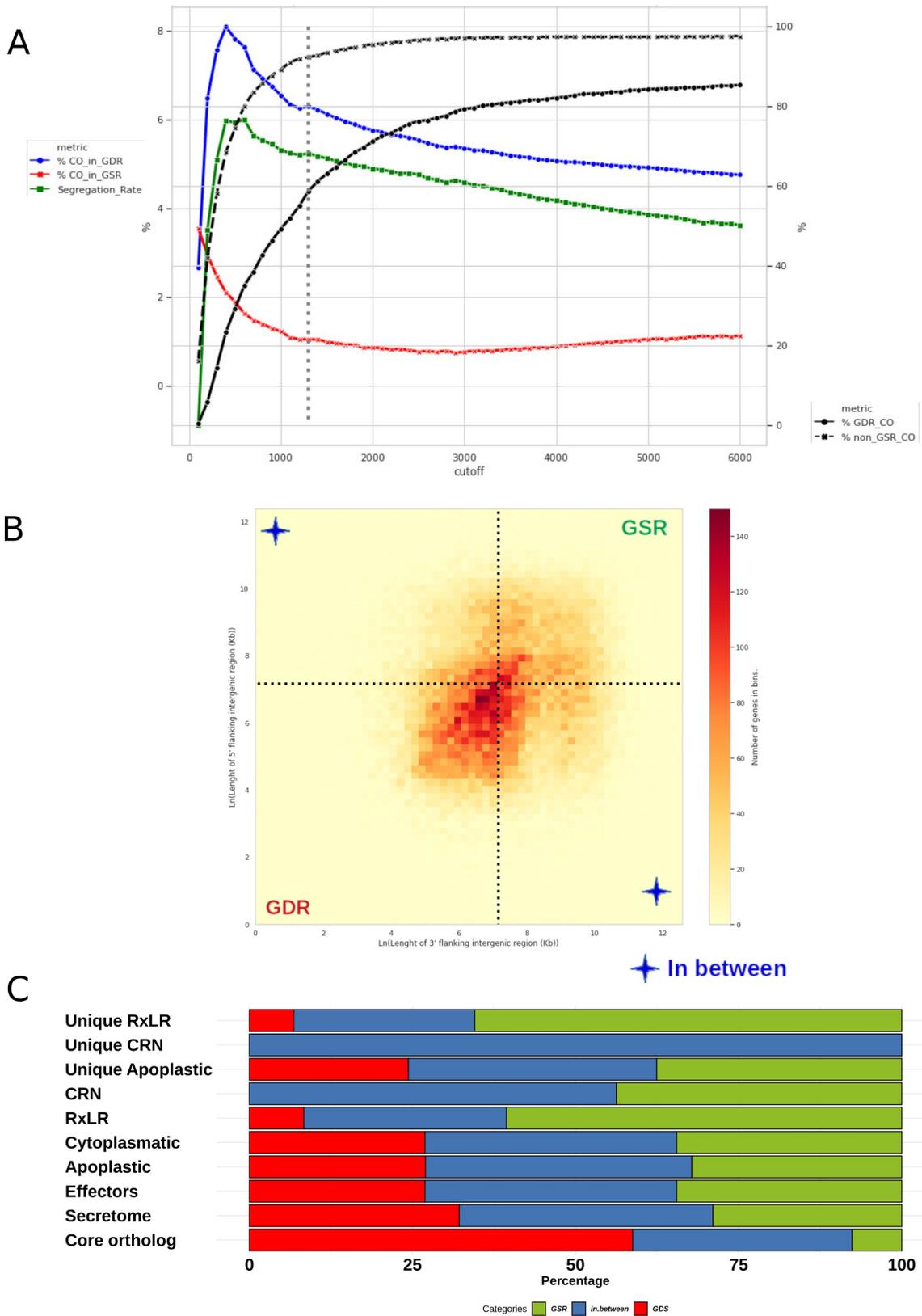


Figure 10. Distribution of *Phytophthora betacei* genes according to the length of their FIRs

10 Figure legends

Figure 1. Number of proteins obtained throughout the SecretSanta pipeline used for the prediction of the *Phytophthora betacei* secretome.

The secretome prediction pipeline is depicted on the left. The proteins include: (a) proteins without signal peptide, (b) proteins with mitochondria/plastid targeting signals, (c) proteins with transmembrane domains (black lines) located after signal peptide, (d) proteins with short ER lumen retention signals and (e) proteins with a signal peptide and an effector domain; ensuring the absence of motifs and domains preventing the protein from being secreted or those targeting it to specific organelles. Step 1: identification of short signal peptides at the N-terminal end of a protein using different version of SignalP. In steps 2, 3 and 4, proteins with motifs and domains preventing the protein from being secreted or those that target it to specific organelles are filtered, using TMHMM, TargetP and KHDEL. The number of proteins retained in each step is shown on the right. The m_slicer function that generates sequences with alternative translation start sites, had as input the first resulting set of predicted proteins having a signal peptide (Rightwards arrow).

Figure 2. Proportion of the predicted secretomes relative to the proteomes of the *Phytophthora* species evaluated.

Proportion of secreted proteins relative to the total proteome for all *Phytophthora* species evaluated. Numbers at the end of each bar indicate the number of proteins predicted in the secretome and proteome of each species. Data were normalized to the size of proteome.

Figure 3. Gene Ontology categories for the *Phytophthora* species secretome.

Number of proteins annotated with GO molecular function, GO cellular component and GO biological process in the secretome. *P. betacei* shows a greater number of proteins associated in each GO category, unlike *P. infestans*, which has a fewer number of associated proteins compared to the other species evaluated.

Figure 4. Assignment of GO terms in the *Phytophthora betacei* secretome.

Number of proteins annotated in each GO category. **(A)** Proteins annotated with GO biological process **(B)** Proteins annotated with GO cellular component **(C)** Proteins annotated with GO molecular function. We only show those GO terms that are present in the upper 75% of the annotated effector proteins.

Figure 5. Proportion of predicted effectors relative to the predicted secretome of the *Phytophthora* species evaluated.

Proportion of effectors in the secretome. Total amount of proteins predicted as apoplasmic effectors and cytoplasmic effectors. Numbers at the end of each bar indicate the number of proteins predicted for the effectors' categories and for the secretome. Data were normalized to the size of secretome.

Figure 6. Assignment of GO terms in the *Phytophthora betacei* effectors.

Number of proteins associated with each GO Term. The corresponding GO category is shown: Biological process (Red), cellular component (Green) and molecular function (Blue). **(A)** GO terms in putative apoplastic effectors **(B)** GO terms in putative RxLR effectors. We only show those GO terms that are present in the upper 75% of the annotated effector proteins.

Figure 7. Proportion of effectors with RxLR_EER and LXLFLAK_DWL (CRNs) motifs relative to the total amount of predicted cytoplasmic proteins.

Numbers at the end of each bar indicate the number of proteins predicted for each effector category (RxLR and LFLAK). Data were normalized to the number of cytoplasmic effectors.

Figure 8. Structure analysis of *Phytophthora betacei* cytoplasmic effectors.

(A) Architecture of a RxLR_EER putative effector **(B)** Architecture of LFLAK_HVVP CRN putative effector. The bigger the letter, the more conserved the amino acid site. Numbers in the sequence logo are referring to the corresponding positions in the alignment. The C-terminal region is polymorphic. The consensus sequence pattern was calculated using Skyline (V.2.0). **(C)** N-terminal domain consisting of a Signal Peptide (1-29 aa) and a recognized domain in the unique putative proteins with RxLR motif in *P. betacei*. **(D)** N-terminal domain consisting of a Signal Peptide (1-16 aa) and a Ubiquitin like (Ubl) domain recognized in *P. betacei* and *P. infestans* that is thought to be responsible for secretion and translocation into the host cell. At the bottom of the figure, we show the C-terminal domains with feature distinct domain types as NTPase + HTH + REase.

Figure 9. Number of unique ortholog effectors in *Phytophthora* species.

Number of effectors from *P. betacei* which are orthologous to those of *P. infestans*. On the bottom left side horizontal bar chart, we show the cluster size of each species. The first columns show singletons effectors present only in each species. Next column shows the number of ortholog effectors shared between the species. **(A)** Orthogroups determined by clustering RxLR effectors in species evaluated without *P. palmivora*. **(B)** Orthogroups determined by clustering CRN effectors in species evaluated without *P. palmivora*. **(C)** Orthogroups determined by clustering apoplastic effectors between *P. betacei* and *P. infestans*. **(D)** Orthogroups determined by clustering RxLR effectors between *P. betacei* and *P. infestans*. **(E)** Orthogroups determined by clustering CRN effectors between *P. betacei* and *P. infestans*.

Figure 10. Distribution of *Phytophthora betacei* genes according to the length of their FIRs

Delimitation and effector content of *P. betacei* gene sparse regions (GSRs) **(A)** Simulation of core ortholog gene segregation. Genes with both flanking intergenic regions (FIRs)

longer than a value 'L' were considered to be present in a gene-sparse region (GSR), whereas genes with both FIRs below L were considered to belong to a gene-dense region (GDR). The percentage of core orthologs in GDRs (%CO_in_GDR) (blue line), the percentage of core orthologs in GSRs (%CO_in_GSR) (red line) and segregation rate (green line) were calculated for a specific cutoff L value. The core ortholog genes segregation rate was defined as the difference between %CO_in_GDR (blue line) and %CO_in_GSR (red line). The highest core ortholog genes segregation rate in which the percentage of core ortholog genes classified as sparse was less than 10% was obtained for L = 1.3Kb (dotted lines) **(B)** Distribution of *P. betacei* genes according to the length of their FIRs. All *P. betacei* predicted genes were sorted into 2-variable bins according to their 3'FIR (Y-axis) and 5'FIR(X-axis). The 1.3Kbp limit for GSRs genes (dotted lines) delimits three groups of genes: genes found in GDRs, GSRs, and In-between regions. **(C)** Distribution of gene groups into the GDRs and GSRs of *P. betacei*. The proportion of secretome, effectors, RXLR core effector genes, CRN core effector genes and the unique RxLR/ CRN that occur in GSRs (green), GDRs (red) and in between (blue) is shown.

11 Supplementary Material

Supplementary Figure 1. Overview of *Phytophthora* species proteome analysis workflow with SecretSanta library developed by Gogleva et al., (2017). The most recent Signalp 4.1 version, was executed by specifying sensitive = TRUE to predict certain classes of secreted oomycete.

Supplementary Table 1. Record of the net values and proportions obtained during the prediction and annotation of the secretome of the different species from *Pythophthora* evaluated.

Supplementary Table 2. Size of proteins secreted for each species of *Phytophthora* evaluated.

Supplementary Table 3. Functional characterization of GO found in effector RxLR and CRN in species evaluated.

Supplementary Table 4. Annotation generated for apoplastic, RxLR and CRN effectors using Blastp, Interproscan and GO categorization for core and unique effectors.

Supplementary Table 5. Number of cytoplasmic effects obtained throughout the study and cytoplasmic effectors reported to the *Pythophthora* species evaluated. Here “trCDS” corresponds to the predicted effectors type RxLR and CRN of each species that were predicted using effectR package, and “published” corresponds to the predicted effectors type RxLR and CRN of each species that were predicted and published on NCBI.

Supplementary Table 6. Apoplastic and cytoplasmic effectors candidates as possible virulence factors responsible for *P. betacei* host specificity

12 Data Availability Statement

The datasets generated for this study can be found in the repository effectors_predict. (https://github.com/ProjasE/effectors_predict).