PORTLAND PRESS

Research Article

In-depth secretome analysis of *Puccinia striiformis* f. sp. *tritici* in infected wheat uncovers effector functions

Ahmet Caglar Ozketen[1], Ayse Andac-Ozketen[1], Bayantes Dagvadorj[1,2], Burak Demiralay[1] and Mahinur S. Akkaya[3]

[1] Middle East Technical University, Biotechnology Program, Ankara 06800, Turkey; [2] Division of Plant Sciences, Research School of Biology, The Australian National University, Canberra, Australian Capital Territory 2601, Australia; [3] School of Bioengineering, Dalian University of Technology, No. 2 Linggong Road, Dalian, Liaoning 116023, China

Correspondence: Mahinur S. Akkaya (msa@dlut.edu.cn) or Ahmet Caglar Ozketen (a.caglarozketen@gmail.com)

OPEN ACCESS

The importance of wheat yellow rust disease, caused by *Puccinia striiformis* f. sp. *tritici* (*Pst*), has increased substantially due to the emergence of aggressive new *Pst* races in the last couple of decades. In an era of escalating human populations and climate change, it is vital to understand the infection mechanism of *Pst* in order to develop better strategies to combat wheat yellow disease. The present study focuses on the identification of small secreted proteins (SSPs) and candidate-secreted effector proteins (CSEPs) that are used by the pathogen to support infection and control disease development. We generated *de novo* assembled transcriptomes of *Pst* collected from wheat fields in central Anatolia. We inoculated both susceptible and resistant seedlings with *Pst* and analyzed haustoria formation. At 10 days post-inoculation (dpi), we analyzed the transcriptomes and identified 10550 Differentially Expressed Unigenes (DEGs), of which 6072 were *Pst*-mapped. Among those *Pst*-related genes, 227 were predicted as PstSSPs. *In silico* characterization was performed using an approach combining the transcriptomic data and data mining results to provide a reliable list to narrow down the ever-expanding repertoire of predicted effectorome. The comprehensive analysis detected 14 Differentially Expressed Small-Secreted Proteins (DESSPs) that overlapped with the genes in available literature data to serve as the best CSEPs for experimental validation. One of the CSEPs was cloned and studied to test the reliability of the presented data. Biological assays show that the randomly selected CSEP, Unigene17495 (PSTG_10917), localizes in the chloroplast and is able to suppress cell death induced by INF1 in a *Nicotiana benthamiana* heterologous expression system.

Introduction

Puccinia striiformis f. sp. *tritici* (*Pst*) is an obligate biotrophic fungus, the causative agent of wheat stripe (yellow) rust disease, a disease that can drastically reduce the production of wheat worldwide. The use of fungicides is a suitable solution for combating *Pst*, but using resistant wheat strains to control Pst genetically would be better for the environment and more cost-effective. In order to achieve host resistance against rapid evolving virulent strains of *Pst*, understanding the molecular basis of the disease is crucial.

Pst infects the leaf of the wheat plant by penetrating to the interior through the stomata and later differentiating into various developmental entities before finally entering into the mesophyll cells in which it can generate a feeding structure called a haustorium. The main interface for pathogen–host interactions is in the haustorium that expands between the plant cell wall and the cell membrane. This interface can activate primary host defenses when pathogen-associated molecular patterns (PAMPs) are detected by pattern recognition receptors (PRRs), resulting in PAMP-triggered immunity (PTI) [1]. Virulence is

achieved by *Pst* through the deployment of secreted proteins, termed as effectors, into the apoplastic fluid or inside the host [2]. The effectors are useful weapons in the fungal arsenal that allow *Pst* (i) to overcome PTI, (ii) to establish a suitable environment, and (iii) to help proliferation [3]. Particular effectors are recognized by cytoplasmic receptor proteins in the host cell, triggering a defense response called effector-triggered immunity (ETI). Compatible and incompatible interactions depend on whether the effectors can achieve virulence by evading Resistance (R) proteins, while the R proteins evolve to detect the effectors. Therefore, evolutionary pressure imparts a driving force on the race between host resistance factors and pathogen virulence factors [4].

Investigating the secreted effectorome of the fungus guides us toward a better comprehension of the nature of pathogenic virulence and determinants that can lead to avirulence. The combination of next-generation sequencing (NGS) and data mining approaches represent a great opportunity to determine candidate effectors secreted by the fungus. Several studies attempting to construct a secretome of either *Pst* or its closest relatives (stem rust pathogen *Puccinia graminis* f. sp. *tritici* (*Pgt*) and leaf rust pathogen *Puccinia triticina* (*Ptt*)) have been published. Yin and colleagues reported a set of secreted proteins generated from expressed sequence tag (EST) sequences of a cDNA library generated from *Pst* haustoria [5]. A repertoire of small-secreted proteins (SSPs) was reported using a pipeline for effector mining in the genome sequence information of *Pgt* and *Melampsora larici populina* (the pathogen that causes poplar leaf rust) [6]. The isolate PST-130 was sequenced and assembled to discover *Pst* genes and identify candidate effectors [7]. The genome sequences of *Pgt* (CRL 75-36-700-3), *Ptt* (BBDD), PST-78 (2K-041), PST-1 (3-5-79), PST-127 (08-220), PST-CYR-32 (09-001) were released by the Broad Institute (http://www.broadinstitute.org/). Cantu and colleagues used comprehensive genome analysis to identify candidate secreted effectors in two United Kingdom isolates (PST-87/7 and PST-08/21) and two United States isolates (PST-21 and PST-43) [8]. They also released data of infection-expressed genes (6 and 14 dpi) and haustoria-expressed genes (7 dpi) for PST-87/7. Another report was published using transcriptome sequencing on *Pst* isolate 104E137A to elucidate haustoria-secreted proteins (HSPs) and expose the best candidate effectors [9]. A pioneer correlation analysis between genomes of seven existing races and seven new races of *Pst* was pursued to predict avirulence (Avr) determinants among SSPs [10]. In this study, 7 new races of *Pst* were sequenced with NGS and then combined with 7 older published genomes for a total of 14 races of *Pst* that were then subjected to correlation analysis to point out Avr candidates [10].

Data mining methodology is a promising strategy to identify candidate-secreted effector proteins (CSEPs) in genome and transcriptome data, as the incorporation of newly generated data continually improves the prediction algorithms using deep machine learning. Likewise, our understanding of the infection and disease progression of the pathogen is steadily expanding. Therefore, new information (candidate effectors) can be generated by exhausting presently available sequence data. A combined effort of using both comparative transcriptomics and data mining to narrow down the ever-expanding sets of CSEPs will simplify and focus on the work being done to discover the function of each CSEP experimentally. Only a few reports have focused on integrating both *in silico* prediction and transcriptome sequencing data on particular *Pst* races at the defined time intervals of haustoria formation and infection [8,9].

In the present study, our objective is to identify differentially expressed gene sequences (DEGs) of *Pst* during the interaction with both susceptible and resistant host genotypes. *YR10* gene is one of CC-NB-LRR type resistance (R) genes [11,12]. We acquired *de novo* transcriptome sequencing data from both susceptible (AvocetS) and resistant (AvocetYR10) wheat cultivars infected with *Pst* for 10 days in order to discover differentially expressed secretome repertoires composed of small secreted proteins at a unique time point of the infection process after haustoria formation. Detected PstDEGs were then subjected to *in silico* prediction analysis to identify differentially expressed small secreted proteins (PstDESSPs). Then, the cataloged secretome candidates were analyzed in-depth via annotation and prediction programs to sort them into classified functions. Furthermore, a comparison of published *Pst* expression data [8,9] indicated the presence of 14 overlapping mutual CSEPs as the most promising subset for experimental validation. To test the reliability of our work, cDNA of one of the CSEP genes (Unigene17495, homologous to PSTG10917) was generated from *Pst*-infected wheat leaves (10 dpi) and cloned into *Agrobacterium tumefaciens*-compatible expression vectors. Heterologous expression of the CSEP in *Nicotiana benthamiana* revealed it was localized to the chloroplast in plant cells. Furthermore, we observed the biological function of Unigene17495 to be suppression of the INF1-mediated cell death response to *N. benthamiana*.

Materials and methods
Plant materials, growth, and pathogen inoculation

Near-isogenic wheat lines of Avocet-S and Avocet-YR10 were used to study both compatible and incompatible interactions upon *Pst* inoculation. *Pst* isolates were collected from the fields of Anatolia, Turkey, and obtained from

the 'General Directorate of Agricultural Research and Policies (TAGEM) of The Ministry of Food, Agriculture, and Livestock'. Avocet-S cultivars are susceptible to *Pst*TR (which has virulence toward Yr2, 6, 7, 8, 9, 25, A, EP, Vic, Mich), whereas Avocet-Yr10 is resistant. Seeds were planted in soil and grown in a growth chamber under ideal conditions (relative humidity 60%, 22°C, 16 h light, and 8 h dark). The *Pst* inoculations were performed when the seedlings reached the two-leaf stage by spraying freshly collected urediniospores directly on to the leaves. Before the inoculation, urediniospores (100 mg) were germinated for 5 min at 42°C and treated with mineral oil (1 ml) to enhance their attachment to the leaf blade. For the control samples (mock inoculations), only mineral oil was used. After air-drying for 20 min, the plants were kept in the dark at 10°C overnight in the presence of continuous mist in order for infection to take place. After overnight infection, the standard growth conditions were resumed. The wheat leaves were collected at 10 days post-inoculation (10 dpi) when the disease symptoms were visible on the susceptible cultivar.

RNA extraction, *de novo* transcriptome sequencing, assembly, and annotation

The flow chart in Figure 1 describes the overall protocol followed for collecting infected leaf samples using homogenization with liquid nitrogen and a sterile mortar and pestle. For 0.1 g leaf sample, 0.02 g poly(vinylpolypyrrolidone) (PVPP) was added during mortar and pestle grinding. For each condition, three biological replicates were homogenized together. Total RNA isolation was carried out using QIAzol Lysis Reagent (Qiagen) by following the manufacturer's protocol. A NanoDrop was used for the spectrophotometric quantification of the total RNA samples. The samples were shipped to Beijing Genomics Institute (BGI) in absolute ethanol for sequencing. BGI measured the RNA quality of the samples on an Agilent 2100 Bioanalyzer, and the samples with a RIN value above 8 were allowed to proceed to mRNA enrichment and cDNA synthesis using random hexamers. Single nucleotide adenine was added to the purified and end-repaired short cDNA fragments. After adapter ligation, size selection was used to obtain templates for PCR amplification. The Illumina HiSeq™ 2000 platform was used to perform *de novo* transcriptome sequencing. BGI-Shenzhen (Shenzhen, China) conducted all the protocol steps from mRNA enrichment to sequencing. All quality controls were evaluated during sequencing using the Agilent 2100 Bioanalyzer and ABI StepOnePlus Real-Time PCR System for quantification and qualification of the sample libraries.

Before assembly, clean reads were obtained from raw reads using internal BGI software (filter_fq) that follows three main parameters: (i) elimination of reads with adaptors, (ii) removal of reads with unidentified nucleotides larger than 5%, (iii) elimination of low-quality reads (if the percentage of the low-quality value ($Q \leq 10$) reads are higher than 20%, they are removed). For *de novo* transcriptome assembly, clean reads of each library were assembled into contigs and unigenes by the Trinity program (release-20130225, http://trinityrnaseq.sourceforge.net/) [13].

Unigenes were aligned using the Blastx program (e-value < 0.00001) and the following protein databases: NR (NCBI, non-redundant database) [14], Swiss-Prot (EMBL protein database) [15], KEGG (Kyoto Encyclopedia of Genes and Genomes) [16], and COG (Clusters of Orthologous Groups) [17]. The direction of the unigene sequences was determined using the best alignment scores. NR, Swiss-Prot, KEGG, and COG were sourced to predict the coding region (CDS) of unigenes in the subsequent order of priority if a conflict in sequence orientation arose. ESTScan software was performed on unigenes with no successive alignments [18]. The results for NR, NT, Swiss-Prot, KEGG, and COG database alignments were also used for unigene function annotation. To obtain gene ontology (GO) annotation from NR annotations, the Blast2GO program (v2.5.0) was performed (e-value < 0.00001) on fungi databases (taxa, 4751) [19]. Web Gene Ontology Annotation Plot (WEGO) software was used for the classification of GO annotations [20]. The KEGG database was employed to investigate the metabolic pathway analysis of unigenes [16].

Analysis of diferential expressions

DEGs were identified using the FPKM method also known as RPKM method [21]. In this method, FPKM stands for fragments per kb per million reads, and RPKM stands for reads per kb per million reads. We used the following calculation formula:

$$FPKM = \frac{10^6 C}{NL/10^3}$$

FPKM signifies the expression of Unigene A, C is the number of fragments that uniquely aligned to Unigene A, N is the total number of fragments that uniquely aligned to all unigenes, and L is the base number in the CDS of Unigene A. As we obtained FPKM values for each unigene, we calculated FPKM ratios between two samples at a time. We defined DEGs to have an FPKM ratio ≥ 2 and a false discovery rate (FDR) ≤ 0.001. We executed GO functional analysis (*P*-value ≤ 0.05), including GO functional enrichment and functional classification on DEGs. Furthermore,

Experimental and Sampling Summary

Mock inoculated AvocetS leaves (ASC)

Pathogen inoculated AvocetS leaves (ASI)

Mock inoculated AvocetYR10 leaves (AYC)

Pathogen inoculated AvocetYR10 leaves (AYI)

Sequencing and Assembly Flow

Total RNA and quality control

Illumina sequencing

De Novo assembly

Expression Difference Analysis (FDR<0.001, |log₂Ratio|≥1)

Mapping Pathogen Genes

Homology search against Puccinia and Wheat

Removal of Unigenes matched to Wheat

Removal of Unigenes matched neither Puccinia nor Wheat

In Depth Secretome Analysis

Final Set of Puccinia matched Unigenes (PstDEGs)

Secretome Prediction (SP:Yes, TM:No, <300a.a)

Blast2GO Annotation

Filtering for Pathogen Genes

Differentially expressed small secreted proteins of Pst (PstDESSPs)

Characterization

Lipases (LED)

Conserved Domain Analysis

Proteases (MEROPS)

Cysteine Rich Effector Analysis (C ≥ 4)

CASzymes (dbCAN)

Effectorome Prediction (EffectorP 2.0)

Subcellular Localization

Comparison against Published Reports

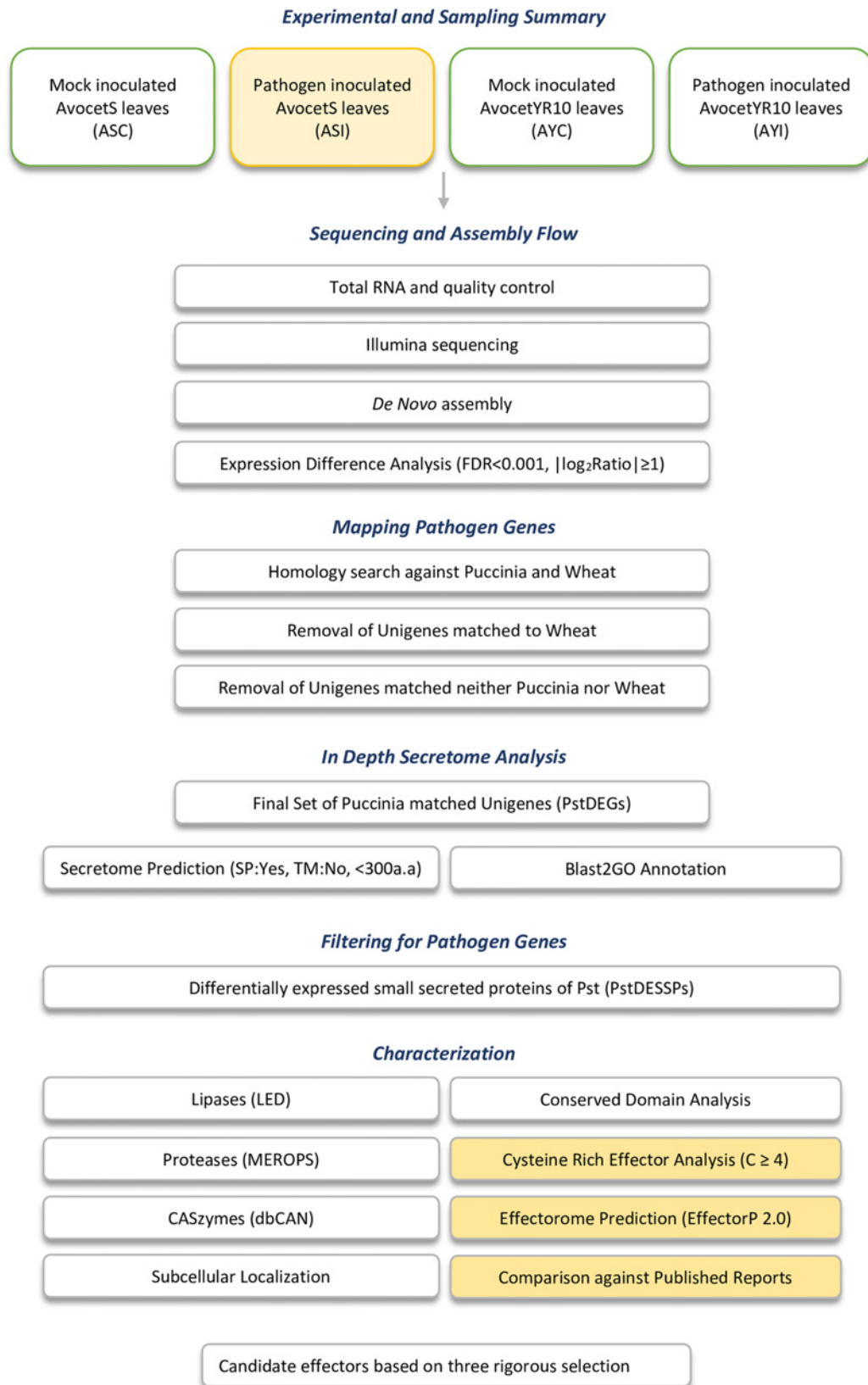Candidate effectors based on three rigorous selection

Figure 1. Flowchart to summarize sampling, sequencing, and *in silico* analysis
Pipeline for the steps performed to achieve *de novo* assembly, expression analysis, and mapping of *Pst* genes. FDR: False discovery rate.

we carried out the KEGG pathway analysis (Q-value ≤ 0.05) for pathway enrichment to further understand biological functions [22].

Identification of pathogen-associated DEGs

We identified our DEGs through transcriptome data obtained from pathogen-inoculated wheat leaves. The project generated two types of data: one is a *de novo* assembled transcriptome of *Pst*. The other one is DEGs mapped to wheat D genome [23], which is advantageous to assess both compatible and incompatible interaction of wheat with Pst. We separated DEGs mapped to wheat (the data not shown), the results will be discussed elsewhere to analyzed host-associated changes against *Pst*. We proceeded with Avocet-Pst/Yr10-Pst combination to obtain *de novo* assembled transcriptome of *Pst* and focused on pathogen-mapped DEGs. However, we needed to eliminate any possible contaminating sequences from the host. To accomplish this, we constructed a local database (named 'The Pucciniale database') by merging protein data of *Puccinia striiformis* f. sp. *tritici* (isolate 2K41-Yr9, race PST-78), *Puccinia graminis* f. sp. *tritici* (strain CRL 75-36-700-3, race SCCL), and *Puccinia triticina* (isolate 1-1, race 1-bbbd). All protein data were obtained from the Broad Institute Puccinia website (http://www.broadinstitute.org/). We also merged protein data of the genome of *Triticum aestivum* (IWGCv.1, downloaded from 'ensemble.plant.org' website) [24] to 'The Pucciniale database'. We performed a BlastX analysis (E-value cutoff ≤ 10^{-5} or $1e^{-5}$) on DEGs using our local database of Pucciniale and wheat [25]. If the hit matched with the wheat proteome or produced a significant hit with the Pucciniale proteome, we discarded the sequences following the decision rule. The best hits within the Pucciniale proteome were evaluated as PstDEGs. The sequences with no significant hits against neither the pathogen nor host were assessed as unknown novel DEGs.

Construction and characterization of secretome data

We predicted the ORFs and protein sequences from the PstDEG data using the ORFPredictor web tool (http://bioinformatics.ysu.edu/tools/OrfPredictor.html). We followed a widely accepted secretome prediction pipeline to construct our own differentially expressed secretome database [6,26,27]. The predictions were made following three criteria: (i) presence of N-terminus signal peptide region, (ii) absence of transmembrane helices, and (iii) mature protein length should be smaller than 300 amino acids. SignalP (version 4.1) was used to predict the presence of a secretion signal [28]. We estimated transmembrane helices using the TMHMM webtool (version 2.0, http://www.cbs.dtu.dk/services/TMHMM/) [29]. Filtered sequences were termed as differentially expressed SSPs (DESSPs) of *Puccinia striiformis* f. sp. *tritici* (PstDESSPs). The same filtering pipeline was applied to total proteins obtained from the 'Broad Institute database' for *Pst* and *Pst* relatives (*Pgt* and *Ptt*) (http://www.broadinstitute.org/). Four secretome datasets (PstDESSPs, PstSSPs, PgtSSPs, PttSSPs) were subjected to effector prediction via the EffectorP 1.0 and 2.0 programs (http://effectorp.csiro.au/) [30,31]. The Blastp program was used (E-value < 10^{-5} or $1e^{-5}$) for characterization of PstDESSPs by scanning against the following databases: the lipase engineering database (http://www.led.uni-stuttgart.de/), the protease and protease inhibitors (MEROPs) database (http://merops.sanger.ac.uk/) [32], the carbohydrate-active enzymes (CAZymes) database (http://csbl.bmb.uga.edu/dbCAN/) [33], the fungal peroxidase database (fPoxDB, http://peroxidase.riceblast.snu.ac.kr/) [34], and the pathogen-host interaction database (http://phi-blast.phi-base.org/) [35]. The ClustalW program was executed for multiple alignment analysis of PstDESSPs (https://www.ebi.ac.uk/Tools/msa/clustalo/) [36]. The iTOL webtool was used to construct a phylogenetic tree using multiple alignment results (https://itol.embl.de/) [37]. Small secreted cysteine-rich effector proteins were assessed as cysteine-rich if the protein possessed four cysteine residues or more. A comparison between infection- and haustoria-expressed genes was made using a similarity search using Blastp (E-value cutoff ≤ $1e^{-5}$ or 10^{-5}) on publicly available data [8,9]. We scanned for PstDESSPs overlapping with *Avr* determinant candidates originally interpreted using correlation analysis by Xia and colleagues [10]. Subcellular localization predictions were done on mature proteins (without N-terminus signal peptides) to forecast translocation sites in the host plant using four different tools: TargetP 1.1 (http://www.cbs.dtu.dk/services/TargetP/) [38], Localizer (http://localizer.csiro.au/) [39], WolfPSORT (https://wolfpsort.hgc.jp/) [40], and ApoplastP (http://apoplastp.csiro.au/) [41]. All the analyses, annotations, and predictions regarding PstDESSPs are displayed in Supplementary Table S3.

Cloning of Unigene17495 (Pstg10917)

The cloning primers were designed using the Pstg10917 sequence with an N-terminus CACC site in forward primers (with and without signal peptide (SP)) and no stop codon in reverse primers (Supplementary Table S4). Total RNA samples of infected Avocet-S leaves (10 dpi) used in transcriptome sequencing were sourced to synthesize cDNA after DNase treatment. The Transcriptor First Strand cDNA Synthesis Kit (Roche) was used following the manufacturer's

protocol, except that we preferred using a mixture of both random hexamer primers and oligo(dT) primers at equal volumes to the reaction primer. The effector of interest, with and without SP, was amplified from cDNA, cloned into the pENTR/D-TOPO vector (Invitrogen), and then subsequently cloned into the pK7FWG2 expression vector [42] utilizing Gateway Cloning. Both colony PCR and sequencing analysis validated the cloning of constructs at the proper positions. Then, the plasmid constructs were electroporated into *Agrobacterium tumefaciens* GV3101.

A. tumefaciens-mediated gene transfer

A. tumefaciens infiltration experiments were performed using the previously described protocol [43] with slight alterations. *A. tumefaciens* GV3101 strains carrying expression vectors were plated on selective agar media. After 2 days, bacteria were scratched and suspended in sterile water. Bacteria were collected by centrifugation at 4000 rpm for 4 min at room temperature (24–25°C). Bacteria were washed two times with sterile water and two times with *A. tumefaciens*-mediated gene transfer-induction medium (10 mM MES, pH 5.6, 10 mM $MgCl_2$). The final cell concentration was adjusted to 0.4 at A_{600nm} for infiltration unless otherwise stated. The inoculums were injected into *N. benthamiana* leaves using a needleless syringe. The expressions of the gene of interests were observed in the next 2–4 days.

Confocal microscopy

A. tumefaciens GV3101 cells carrying pK7FWG2/Pstg_10917 and pK7FWG2/ΔSP-Pstg_10917 were injected into *N. benthamiana* leaves (4–5 weeks old) for transient expression. After 2–3 days, leaves were cut into small pieces of 3–5 mm length and soaked in tap water. A Leica 385 TCS SP5 confocal microscope (Leica Microsystems, Germany) was used for subcellular localization imaging. The wavelengths used for imaging GFP fluorescence were excitation at 488 nm and emission at 495–550 nm. Alternatively, the excitation and emission wavelengths for chloroplast autofluorescence were in the far-infrared (>800 nm).

Cell death suppression assay

The pTRBO/GFP plasmid for GFP expression (courtesy of the Kamoun Lab, Sainsbury Laboratory, Norwich, U.K.) and the pK7FWG2/SP-GFP plasmid [44] for SP-GFP were used as negative controls in the cell death suppression assays. The *A. tumefaciens* GV3101 cells carrying the negative control plasmids (GFP and SP-GFP) or the effector of interest (with and without SP) were injected into a *N. benthamiana* leaf. After 24 h, *A. tumefaciens* GV3101 carrying cell death elicitor pGR106/Inf1 (a gift from the Bozkurt lab, Imperial College, London, U.K.) was injected in the same spot without damaging leaf tissue. Cell death was measured at 4 dpi. The infiltration ring area was assessed if the brown region was less than 25% considered as strong suppression, 25–50% as moderate, 50–75% as weak, and more than 75% as no suppression at all.

Results and discussion
Sequencing of *Pst*-infected wheat leaves and *de novo* transcriptome assembly

We have collected total RNA from *Pst*-infected wheat leaves of susceptible and resistant seedlings to analyze gene expression differences between the compatible and incompatible pathogen–host interactions. The comparative transcriptome strategy was employed in similar studies; for instance, Sonah and colleagues identified virulence factors of *Leptosphaeria maculans* on susceptible and resistant canola cultivars, whereas Gupta and colleagues, on the pea powdery mildew *Erysiphe pisi* pathogen of *Medicago truncatula* [45,46]. Similarly, we managed to gather data from Pst pathogen generated on both susceptible (AvocetS) and resistant (AvocetY10) genotypes. The sequence reads have been deposited in the NCBI BioProject database under the accession number PRJNA545454. A total of 9204786 clean reads for *Pst*-infected susceptible wheat (Avocet-S_PST) and 6899596 clean reads for *Pst*-infected resistant wheat cultivar (Avocet-YR10_PST) were obtained after quality filtering. Their Q20 values were 98.16 and 97.90%, respectively (Table 1).

The clean reads were assembled into more than 100000 contigs with an average length of greater than 200 nucleotides using Trinity software (Table 2). Unigenes were clustered based on their level of similarity with each other. A cluster was defined as containing unigenes with more than 70% similarity. We generated 12956 distinct unigene clusters for Avocet-YR10_PST and 13564 for Avocet-S_PST (Table 2). A total of 42275 and 55004 unigenes came from single genes within the two sets, respectively. These data are represented in Table 2 as the distinct singleton. The length distribution of assembled contigs and unigenes is provided in Supplementary Figure S1.

Table 1 Sequencing statistics for *Pst* inoculated/infected wheat transcriptome

Samples	Avocet-S_*Pst*	Avocet-Yr10_*Pst*
Total clean reads	9204786	6899596
Total clean nucleotides (nt)	828430740	620963640
Q20	98.16%	97.90%
GC	51.44%	54.04%

Q20 is the proportion of nucleotides with a quality value larger than 20.
GC is the proportion of guanine and cytosine nucleotides among total nucleotides.

Table 2 *De novo* assembly statistics for *Pst* inoculated/infected wheat transcriptome

	Samples	Number	Length (nt)	Mean length (nt)	N50	Distinct clusters	Distinct singleton
Contig	Avocet-Yr10_*Pst*	105866	23189064	219	249	-	-
	Avocet-S_*Pst*	128680	29848685	232	272	-	-
Unigene	Avocet-Yr10_*Pst*	55231	18753197	340	371	12956	42275
	Avocet-S_*Pst*	68568	24329132	355	394	13564	55004
	All	61105	26978874	442	490	15016	46089

Distinct clusters are similar (more than 70%) unigenes, and these unigenes may originate from the same gene or homologous gene, whereas distinct singletons represent the unigenes come from a single gene.

Table 3 Annotation statistics of the assembled unigenes of the transcriptome sequencing project

Sequence file	NR	NT	Swiss-Prot	KEGG	COG	GO	All
All-Unigene.fa	24933	9945	16318	16275	13278	10015	26843

Functional annotation and classification of *de novo* transcriptome

The unigenes obtained from both Avocet-S_PST and Avocet-YR10_PST were merged into a single integrated dataset to conduct an overall interpretation of transcriptome sequencing. The unigenes were aligned against the following databases: NR, NT, Swiss-Prot, KEGG, COG, and GO for annotation via the Blastx program (e-value < 0.00001). Overall, 26843 of 61105 unigenes, corresponding to 43.9%, were annotated in one or more databases (Table 3). The detailed annotation results are listed in Supplementary Table S1. The NR database provided the highest number of annotations. Approximately half of the annotated unigenes shared high similarity (>60%) and aligned to either *Pst* genes (41.4%) or genes of other fungal species (Figure 2). The results of COG classification and GO functional annotation of the unigenes are presented in Supplementary Figure S2.

Discovery of DEGs

We used the FRKM (RPKM) method [21] to determine DEGs between compatible and incompatible interactions. We identified 10550 DEGs with a FDR of less than 0.001 and an expression difference of at least two-fold (Figure 3).

We mapped the DEGs to our custom database constructed by merging the pathogen and host genomes. Local Blastx analysis (e-value < 0.00001) revealed that 6072 of the DEGs aligned best to the *Pucciniales* genome, and 3028 of the DEGs aligned best to the wheat genome. The remaining DEGs yielded no significant homology to either the pathogen or the host; hence, they are considered to be unknown novel DEGs. The 6220 *Pst*-mapped DEGs (PstDEGs) were selected as the differentially expressed, pathogen-associated candidate genes to proceed with for secretome analysis.

The identified PstDEGs were subjected to Blast2GO analysis to elucidate their functional classification (Supplementary Table S2). A graphical view of the molecular functions and biological processes of the PstDEGs is displayed in Figure 4. We were able to annotate the majority (75.8%) of PstDEGs with their corresponding functional classification. The distribution of molecular function between proteins with catalytic activity (61.73%) and binding activity (61.07%) attributes were practically equal in the Blast2GO level 2 analysis (GO specificity increases as the 'level' of the analysis gets higher). During level 3 analysis of the PstDEGs, we observed that the proteins with catalytic activity belonged to the hydrolase, transferase, and oxidoreductase categories (Figure 4A). The Blast2GO analysis illustrates that the pathogen undergoes radical changes during disease progression. A large number of PstDEGs (6072 genes
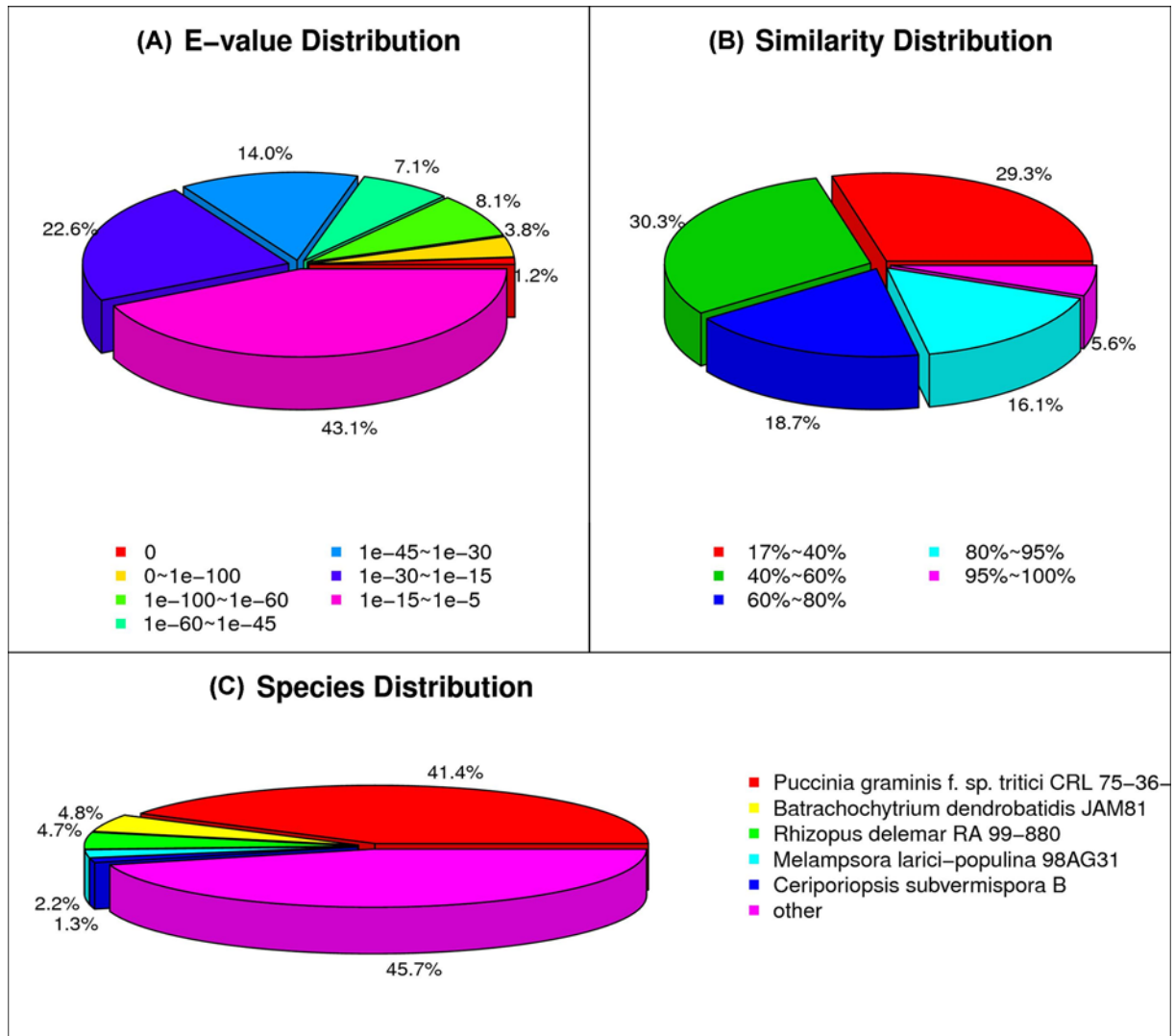
Figure 2. Statistics of the NR classification of unigenes
(**A**) The e-value distribution of the alignment results of NR annotation. (**B**) The similarity distribution and (**C**) the species distribution of the results of NR annotation.
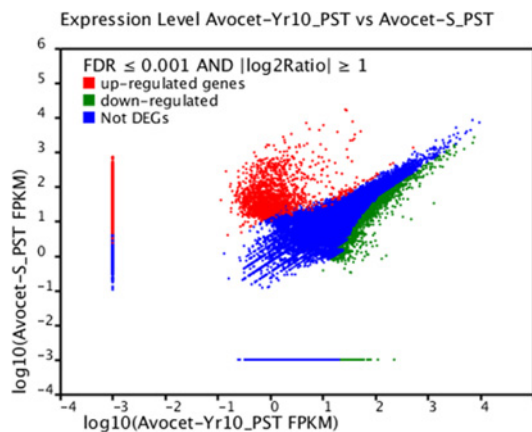

Figure 3. Distribution of DEGs
Red: up-regulated genes (6851), green: down-regulated genes (3699), and blue: no differential expression observed with the parameters FDR $\leq$ 0.001 and $|$log2 ratio$| \geq$ 1.
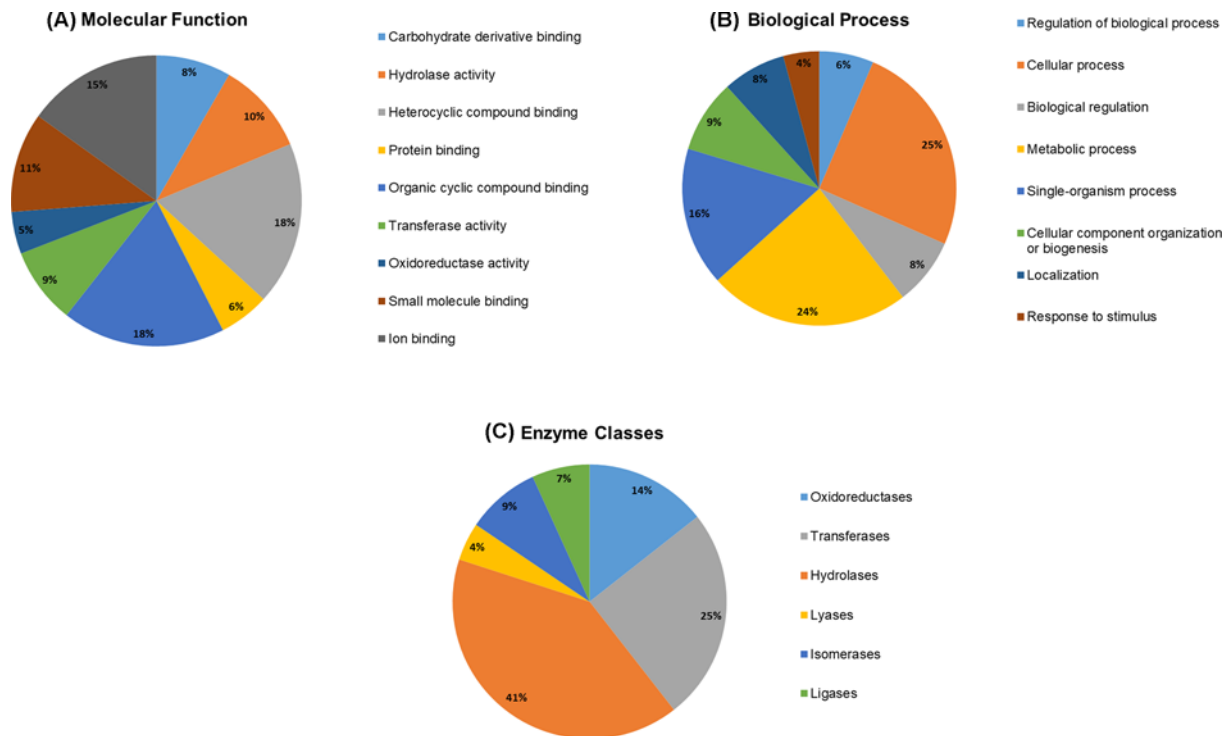
Figure 4. Functional annotations of PstDEGs via the Blast2Go program
(**A**) Molecular function analysis (Level 3) of PstDEGs. (**B**) Biological process classification (Level 2) of PstDEGs. (**C**) Enzyme class distribution of PstDEGs.

compared with the total predicted protein number of *Pst* genes which is over 20000) were sorted into various categories of biological processes (Figure 4B). Therefore, we wanted to determine the number of PstDEGs that sorted as enzymes. In total, we cataloged 1848 PstDEGs as enzymes. Among them, hydrolases were the most common, followed by transferases and oxidoreductases (Figure 4C).

Identification and characterization of the differentially expressed secretome

We made use of a previously constructed pipeline defined and validated by several research groups [6,23,24]. We extracted the secretome repertoire of 227 SSPs of *Puccinia striiformis*, which are referred to as PstDESSPs, by selecting all of the DEGs identified in our transcriptome data (6072 DEGs). We identified other SSPs by applying the same pipeline on all of the available genome sequences of *Puccinia* species: *Pst, Pgt,* and *Ptt* (http://www.broadinstitute. org/). The mRNA levels are not directly proportional to the level of protein expressions; nevertheless, transcriptome and proteome expression values are correlated [47]. Moreover, there are reports stating that the overall predicted CSEPs compose a significant proportion of the genome [10,48]. In a similar fashion, we hypothesize that the number of CSEPs during pathogenesis should denote a large amount as well. We then compared all of the identified secretomes in a method similar to that used in the study published by Xia and colleagues, 2017 [10]. Among 227 small secreted proteins, 94 were predicted as effectors using 'Effector 2.0' software, and were then classified as Pst-Small Secreted Candidate Effectors (PstSSCEs) (Figure 5A). The percentage of PstDESSPs within the dataset of all DEGs was 3.7% (227 out of 6072). This ratio appeared significantly smaller when compared with the hypothetical prediction of small secreted proteins within PST-78 (6.5%, 1332 out of 20482) and the other *Pucciniale* genomes. Therefore, we concluded that the number of actively involved small secreted proteins at 10 dpi is indeed much lower than what was hypothetically expected. This outcome indicates that the hypothetical number of PstDESSPs can be exaggerated when obtained using only transcriptome predictions from genome sequences.

Certain types of gene functions are useful for the fungus in pathogen–host interactions, including CAZymes, proteases, peroxidases, and lipases. The PstDESSPs were compared against several reliable public databases (MEROPs, dbCAN, LED, fPoxDB) in order to predict their functions. We identified four proteases and protease inhibitors, four
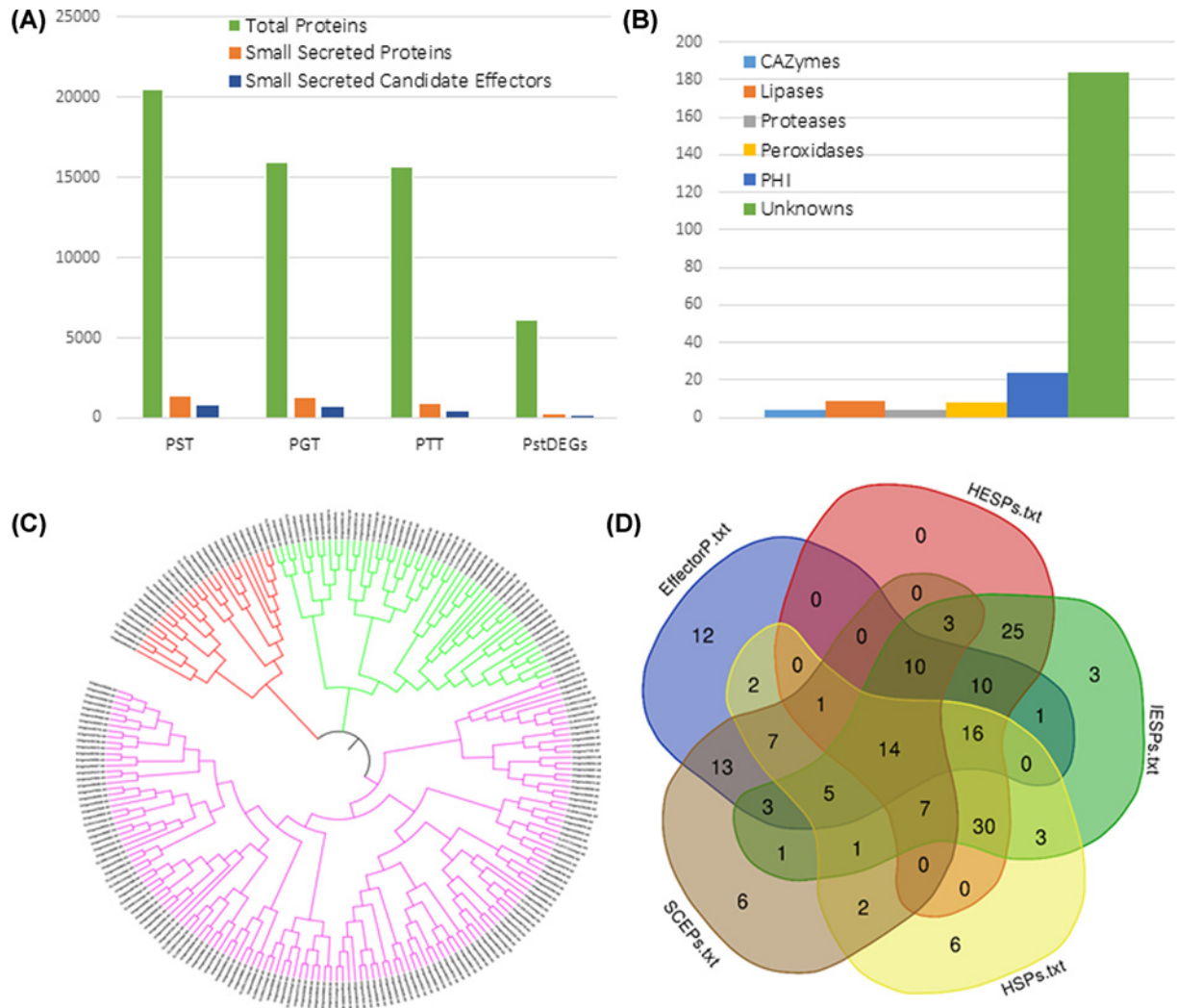
Figure 5. Characterization of the PstDESSPs

(**A**) The number of total proteins, small secreted proteins, and small secreted candidate effectors among our data (PstDEGs) and three wheat rust fungi genomes: *Pst, Pgt* (*P. graminis* f. sp. *tritici*), and *Ptt* (*P. triticina*). (**B**) Distribution of PstDESSPs for functional annotations. (**C**) Phylogenetic analysis based on the protein sequence similarity of PstDESSPs. (**D**) Characterization and comparison of PstDESSPs in five categories: small cysteine-rich effector proteins (SCEPs), predicted to be an effector candidate by a machine learning algorithm (EffectorP 2.0-positive), and consistent with previous sequencing reports in the literature of candidate rust effectors in haustoria (HESPs) or infection (IESPs) expressed secreted proteins [8], and HSPs [9].

CAZymes, nine lipases, and eight peroxidases among the PstDESSPs (Figure 5B). We searched the Pathogen–Host Interaction (PHI) database to predict the virulence attribute of each PstDESSP based on their sequence similarity. A total of 24 of the PstDESSPs exhibited significant similarity to at least one entry in the PHI database. For example, Unigene31938 (Pstg04010) had a predicted 'copper-zinc superoxide dismutase' domain match with PHI383 (*C. albicans*) and PHI6412 (*P. striiformis*) (Supplementary Table S3). Although the level of similarity barely exceeded the cut-off value, both these entries have a superoxide dismutase domain. PHI6412 (PsSOD1) has been reported to increase the resistance of the pathogen against host-derived oxidative stress [49]. The results suggest that all of the PstDESSPs that matched with entries in the PHI database are worthy enough of further exploration employing *in vivo* systems. Hence, a comparison against multiple databases provides a better way to discover CSEPs and understand the importance of the PstDESSPs in Supplementary Table S3. Furthermore, a considerable percentage (80%) of the PstDESSPs remain uncharacterized (Figure 5B), due to the tendency of the candidate effectors to only show homology to the pathogenicity-related domains and not to other typical functional domains.

To understand the relationship between PstDESSPs, we generated a phylogenetic tree using an analysis of multiple sequence alignments. The outcome suggests that there are no significant conserved sequences among the 227 PstDESSPs. In general, we identified three main branches (or classes) of PstDESSPs based on sequence similarities. The purple branch in Figure 5C possesses a substantial amount of the PstDESSPs and is also subdivided into two smaller arms. No significant conserved motifs were detected among the PstDESSPs, except for the Y/F/WxC motif (Supplementary Table S3), which is common to the powdery mildew [50]. The search for the conserved sequences, correlations, and motifs are at the center of the studies using genome-wide analysis or transcriptome-based analysis of CSEPs [51]. Since the discovery of RXLR or LXLFLAK motifs aid the mining of large effector families in oomycetes [52]. Only [YFW]xC motif has been discovered in rust and mildew fungus; however, there is no report for its contribution to the function or localization [48,50,51]. Similarly, we here checked the sequence resemblance and presented it as a phylogenetic tree. Motif mining through MEME motif finder also did not give any significant repeating motif pattern. Hence, we reported the negative result in parallel with the literature.

The PstDESSPs determined in the present study were compared with other published expression datasets of *Pst* to investigate whether the SSPs in the present study were unique or similar to haustorial differentially expressed effector candidates. Cantu and colleagues studied the expression data for effectors expressed in infected leaf and haustoria samples and provided a pioneer list of SSP tribes in their analysis [7,8]. We analyzed the similarity between tribes of SSPs via the Blastp program (E-value < 0.00001). A total of 116 PstDESSPs were found to share significant homology to haustoria-expressed secreted effector proteins (HESPs), and 132 of PstDESSPs were found to be homologous to infected leaves-expressed secreted effector proteins (IESPs) (Supplementary Table S3). The remaining 94 PstDESSPs that did not match with HESPs or IESPs were found to be differentially expressed at 10 dpi in the present study. Comparison with the list of HSPs published by Garnica and colleagues indicated that 94 of the PstDESSPs are similar to HSPs [9]. Overall, the number of PstDESSPs that we identified that are not similar to those in previously reported datasets is 31, signifying their exclusivity to the present study. Xia and colleagues scrutinized the effector candidates for avirulence (Avr) by comparing the expression datasets of Cantu and colleagues, 2013 [8], [10]. We compared our list of PstDESSPs with the list of Avr candidates, and the results suggest that three previously overlooked Avr candidates and six known Avr candidates are present among our PstDESSP dataset (Supplementary Table S5).

The cysteine content of a candidate effector protein has long been a determinant for making effector predictions [8–10,23,24]. Although several effectors have been discovered to be not rich in their cysteine composition, such as AvrM [53] and AvrL567 [54], we examined the PstDESSPs in terms of cysteine possession. We identified 64 cysteine-rich small-secreted effector proteins (SCEPs) that have four or more cysteine amino acid residues in their mature protein length (not including the N-terminal signal peptide) (Supplementary Table S5).

To find the most promising subset of candidate effectors, we focused on the PstDESSPs that produced positive results in the EffectorP analysis, had cysteine residue counts of more than three, and had homologous matches within the previously mentioned reports (HESPs, IESPs, and HSPs). A total of 14 PstDESSPs met all five criteria and were defined as reliable CSEPs for further investigations (Figure 5D). A more detailed description of the features of the 14 CSEPs is provided in Table 4.

Subcellular localization of Unigene17495 (Pstg_10917)

One of the CSEPs from Table 4 was chosen at random to test the reliability of the presented PstDESSPs data. Unigene17495 (Pstg_10917) was predicted to have a transit peptide sequence following the N-terminus secretion signal. Thus, if the effector candidate succeeds in translocation within the host cell, it may be able to partially target the chloroplast, the site of control for programmed cell death (PCD) against both biotic and abiotic stressors. The heterologous expression of Unigene17495 with a GFP tag on *N. benthamiana* leaves showed chloroplast localization (Figure 6). Its localization to the chloroplast despite being predicted for secretion outside the cell was unexpected. However, re-checking the predictions by other algorithms; 'TargetP' in addition to predicting a secretion signal, only after its removal, it also detected a transit peptide within the remainder sequence (111 a.a.). On the other hand, 'Localizer' analysis predicted a transit peptide overlapping the secretion signal region. Therefore, the experimentally shown chloroplast localization, instead of originally predicted secretion outside of the cell, can be explained by the presence of a transit peptide and a possible disruption of the secretion signal due to a particular condition in the vicinity of the molecule. Alternatively, in the natural host, trafficking in both locations may be possible or there is a slight change in the physiological condition that may direct a preference of one target location to another.

Table 4 List of the most promising CSEPs overlapped in all five classes

| ID | CSEP properties | | | Subcellular localization prediction | | | | | |
	Corresponding Pst78 homologs	Conserved domains	Length (a.a.)[1]	Cys[2]	DE[3]	TP[4]	Loc.[5]	WolfPSORT	ApoplastP
CL887.Contig1	PSTG_16009		138	10	Up	-	-	C	AP
CL887.Contig2	PSTG_16009		138	10	Up	-	-	C	AP
Unigene13023	PSTG_03450		100	10	Up	-	-	C	AP
Unigene17495	PSTG_10917		111	6	Up	C	C	M	AP
Unigene28760	PSTG_14090	DPBB_1 superfamily	112	6	Up	-	-	C	AP
Unigene29179	PSTG_04274		82	6	Up	-	-	C	AP
Unigene31294	PSTG_00994		254	10	Up	-	N	N	NonAP
CL5364.Contig2	PSTG_01880		205	6	Up	-	-	ER	NonAP
CL5364.Contig3	PSTG_01881		205	6	Up	-	-	N	AP
Unigene33801	PSTG_05079		92	6	Up	-	-	C	AP
Unigene36354	PSTG_08969		196	8	Up	C	-	C	AP
Unigene36901	PSTG_03879, partial		49	6	Up	-	-	N	AP
Unigene38807	PSTG_14378		191	10	Up	-	-	G	AP
Unigene7930	PSTG_02173		109	4	Up	-	-	C	AP

Abbreviations: AP, apoplastic; C, chloroplast; Cy, cytosolic; E.R., endoplasmic reticulum; G, Golgi; M, mitochondria; N, nucleus; NonAP, non-apoplastic.
All predictions were performed using mature protein sequences under the assumption that mature proteins translocate into the host cell.
[1]Length represents the length of mature proteins without N-terminus signal peptide part.
[2]Cysteine residue number in the mature protein.
[3]Differential expression compatible interaction over incompatible interaction.
[4]The results of TargetP (TP) program.
[5]Localizer (Loc.) program.

Unigene17495 (Pstg10917) suppresses INF1-induced cell death

Unigene17495 (Pstg10917) is conserved among *Pst* relatives. However, none of the homologs have been previously studied for their function. Interestingly, an effector candidate of the soybean rust pathogen (*Phakopsora pachyrhizi*), PpEC82 [55], resembles Unigene17495 (Pstg_10917) with an E-value of $1 \times e^{-14}$ (10^{-14}). PpEC82 was reported to localize to the cytosol and nucleus with strong aggregation [56] properties, unlike the chloroplast-targeting Unigene17495 (Pstg_10917) as we have demonstrated. Moreover, Qi and colleagues stated that PpEC82 was able to suppress BAX-induced cell death when expressed in yeast cells [56]. Correspondingly, we conducted cell death suppression assays to test whether Unigene17495 (Pstg_10917) has suppression abilities similar to PpEC82. We used a cell death elicitor (INF1) as a cell death inducer and categorized the resulting suppression of cell death observations into four levels: strong suppression, moderate suppression, weak suppression, or no suppression at all. Across 15 replicates of Unigene17495-treated *N. benthamiana* leaves, we observed three cases of strong, three cases of moderate, and three cases of weak cell death suppression (Figure 7, Supplementary Figure S3). The other six leaves presented no significant level of cell death suppression. In all 15 trials, leaves treated with the negative controls (GFP and SP-GFP) resulted in tissue collapse and cell death in the infiltrated area. Therefore, we concluded that Unigene17495 (Pstg_10917) without SP, was able to partially suppress programmed cell death triggered by the INF1 elicitor (Figure 7). We need to emphasize that not all attempts produced equal levels of cell death suppression. Although the expression of the effector was verified by microscope analysis at 2 dpi for all of the leaves, 6 out of the 15 leaves revealed no detectable suppression of INF1-induced cell death. Fisher's exact test calculated a *P*-value of 0.0007 for the findings, and thus the results were significant *(P*<0.05). The Unigene17495 (Pstg10917) is up-regulated on susceptible cultivar during the compatible interaction. Our findings suggest that Pstg10917 may be involved in overcoming the host's defense mechanisms by modulating or suppressing hypersensitive cell death triggered by pathogen recognition. Together, these data show that Pstg10917ΔSP-GFP (Unigene17495) is capable of significantly suppressing PCD induced by INF1.

It should be noted that the complex nature of heterologous expression systems serves as a double-edged sword due to the oscillations in cell death response with *N. benthamiana*. However, the versatility of the system provides a practical convenience to study candidate effectors. Other candidate effectors of *Pst* have been studied using heterologous expression systems. For example, the *Pst* effector PstHa5a23 was reported to suppress PCD triggered by
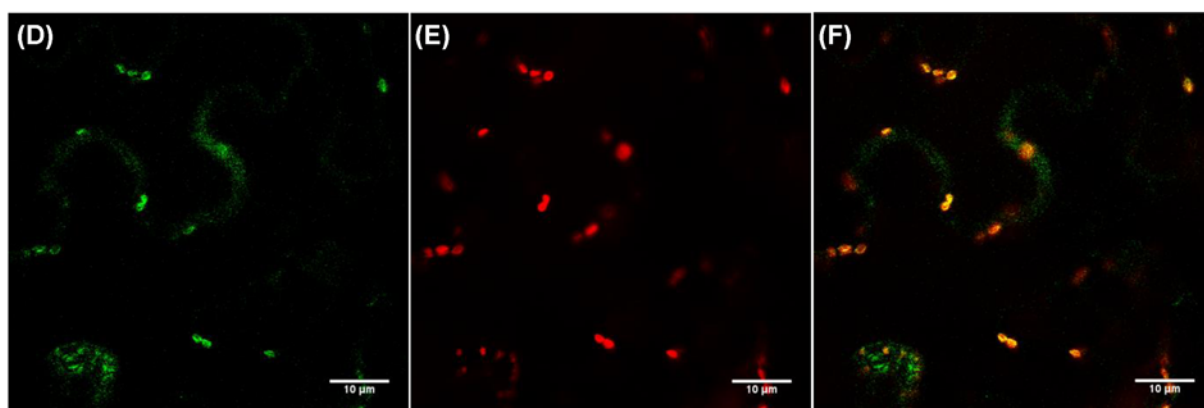
Pstg10917ΔSP-GFP
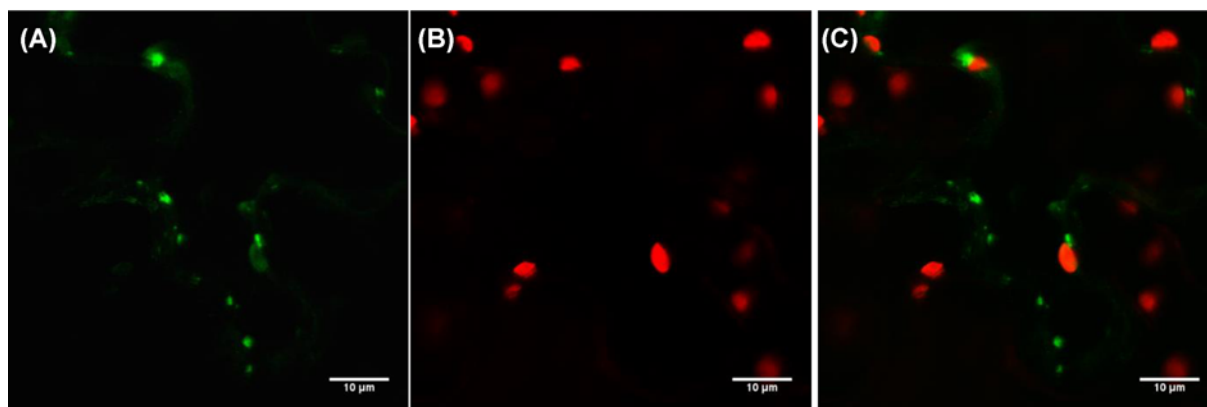
Pstg10917-GFP

Figure 6. Subcellular localization of Unigene17495 (Pstg10917) tagged with GFP at C-terminus

(**A**) Expression of Pstg10917ΔSP-GFP (without N-terminus secretion signal). (**B**,**E**) Chloroplast autofluorescence. (**C**) and (**F**) Overlaid images. (**D**) Expression of Pstg10917 (with N-terminus secretion signal). The expression on *N. benthamiana* leaves was visualized under a confocal microscope at 2 dpi.
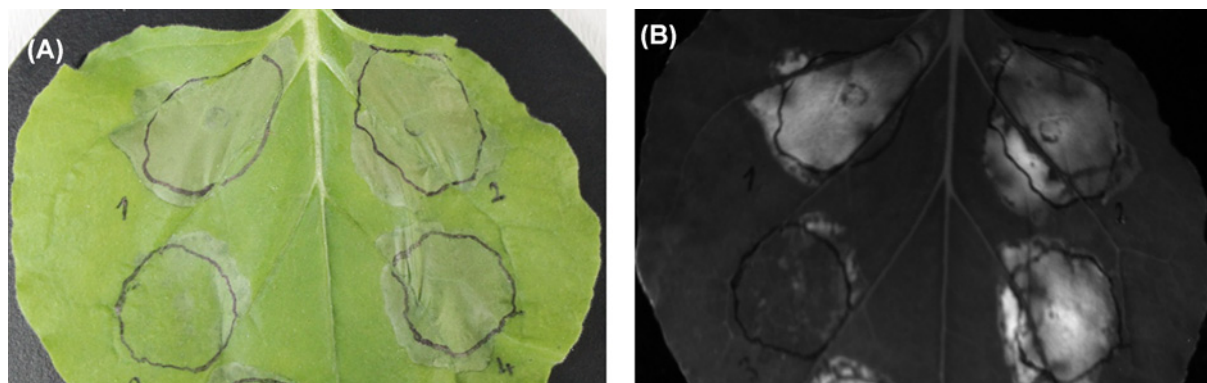

Figure 7. Suppression of cell death mediated by Inf1 expression

(**A**) Daylight and (**B**) UV light exposure. Constructs were expressed with *A. tumefaciens*; (1) GFP, (2) SP-GFP, (3) Unigene17495 without SP (Pstg10917ΔSP-GFP) and (4) Unigene17495 (Pstg10917-GFP). Cell death elicitor (Inf1) was infiltrated into the same region after 24 h. Photos were taken at 4 days after Inf1 challenge.

BAX, INF1, and two mitogen-activated protein kinases (MKK1 and NPK1) involved in host cell resistance [57]. PstHa5a23 is a homolog to Pstg_00676 identified in our PstDESSPs list. In a recent publication, Pstg_14695 and Pstg_09266 were able to suppress PCD as part of effector-triggered immunity [58]. Furthermore, Petre and colleagues utilized mass spectroscopy analysis to decipher host cell factors that interact with GFP-tagged candidate effectors in co-immunoprecipitation samples [59]. Eleven of the twenty candidate effectors reported in that study are homologous to our PstDESSPs. These recent reports support the reliability of our PstDESSPs data as a trustworthy collection of candidate effectors that can be used for effector function studies in the future.

Conclusion

The obligate biotrophic fungus *Pst* causes yellow rust disease by establishing an interface between the host and pathogen using small secreted proteins. Transcriptome sequencing has proven a useful strategy in elucidating the cross-talk of proteins between the host and pathogen, including the interaction between wheat and *Pst*. We compared compatible and incompatible interaction points in order to identify the DEGs (PstDEGs) at 10 dpi. *In silico* predictions highlighted the SSPs within the group of PstDEGs. The annotations, comparisons, and characterizations contributed to the identification of a list of the most promising subset of SSPs, that function as phytopathogen effectors. The presented SSP data represent candidates for the functional analysis to understand yellow rust disease in wheat better. In particular, we were able to determine that Unigene17495 (Pstg10917) of our PstDESSPs list is localized to the chloroplasts of infected plant cells and is capable of halting programmed cell death. The remaining 13 promising candidates are to be investigated in future studies to test whether they are indeed effectors of *Pst*.

Data Availability
All raw data and sequence reads are uploaded on the NCBI BioProject database under the accession number PRJNA545454. The annotations and analysis results are available in Supplementary files.

Competing Interests
The authors declare that there are no competing interests associated with the manuscript.

Author Contribution
Ahmet Caglar Ozketen: Methodology, Investigation, Data Curation, Validation, Writing – original draft preparation, Writing – review and editing. Ayse Andac-Ozketen: Investigation, Validation. Bayantes Dagvadorj: Investigation. Burak Demiralay: Data Curation. Mahinur S. Akkaya: Methodology, Funding acquisition, Supervision, Writing – review and editing.

Abbreviations
BGI, Beijing Genomics Institute; CAZyme, carbohydrate-active enzyme; COG, Clusters of Orthologous Groups; CSEP, candidate-secreted effector protein; DEG, differentially expressed gene/unigene; DESSP, Differentially Expressed Small-Secreted Protein; dpi, days post-inoculation; EST, expressed sequence tag; FDR, false discovery rate; FPKM, fragments per kb per million reads; GO, gene ontology; HESP, haustoria-expressed secreted effector protein; HSP, haustoria-secreted protein; IESP, infection-expressed secreted proteins; KEGG, Kyoto Encyclopedia of Genes and Genomes; NGS, next-generation sequencing; PAMP, pathogen-associated molecular pattern; *Pgt*, *Puccinia graminis* f. sp. *tritici*; PHI, Pathogen–Host Interaction database; *Pst*, *Puccinia striiformis* f. sp. *tritici*; PTI, PAMP-triggered immunity; Ptt, *Puccinia triticina*; RIN, RNA integrity number; RPKM, reads per kb per million reads; SCEP, small cysteine-rich effector protein; SSP, small-secreted protein.

References

1   Stotz, H.U., Mitrousia, G.K., de Wit, P.J.G.M. and Fitt, B.D.L. (2014) Effector-triggered defence against apoplastic fungal pathogens. *Trends Plant Sci.* **19**, 491–500, https://doi.org/10.1016/j.tplants.2014.04.009

2   Garnica, D.P., Nemri, A., Upadhyaya, N.M., Rathjen, J.P. and Dodds, P.N. (2014) The ins and outs of rust haustoria. *PLoS Pathog.* **10**, 10–13, https://doi.org/10.1371/journal.ppat.1004329

3   De Jonge, R., Bolton, M.D. and Thomma, B.P.H.J. (2011) How filamentous pathogens co-opt plants: the ins and outs of fungal effectors. *Curr. Opin. Plant Biol.* **14**, 400–406, https://doi.org/10.1016/j.pbi.2011.03.005

4   Boller, T. and Felix, G. (2009) A renaissance of elicitors: perception of microbe-associated molecular patterns and danger signals by pattern-recognition receptors. *Annu. Rev. Plant Biol.* **60**, 379–407, https://doi.org/10.1146/annurev.arplant.57.032905.105346

5   Yin, C., Chen, X., Wang, X., Han, Q., Kang, Z. and Hulbert, S.H. (2009) Generation and analysis of expression sequence tags from haustoria of the wheat stripe rust fungus Puccinia striiformis f. sp. Tritici. *BMC Genomics* **10**, 1–9, https://doi.org/10.1186/1471-2164-10-626

6   Duplessis, S., Hacquard, S., Delaruelle, C., Tisserant, E., Frey, P., Martin, F. et al. (2011) Melampsora larici-populina transcript profiling during germination and timecourse infection of poplar leaves reveals dynamic expression patterns associated with virulence and biotrophy. *Mol. Plant Microbe Interact.* **24**, 808–818, https://doi.org/10.1094/MPMI-01-11-0006

7   Cantu, D., Govindarajulu, M., Kozik, A., Wang, M., Chen, X., Kojima, K.K. et al. (2011) Next generation sequencing provides rapid access to the genome of Puccinia striiformis f. sp. tritici, the causal agent of wheat stripe rust. *PLoS ONE* **6**, 4–11, https://doi.org/10.1371/journal.pone.0024230

8   Cantu, D., Segovia, V., MacLean, D., Bayles, R., Chen, X., Kamoun, S. et al. (2013) Genome analyses of the wheat yellow (stripe) rust pathogen Puccinia striiformis f. sp. tritici reveal polymorphic and haustorial expressed secreted proteins as candidate effectors. *BMC Genomics* **14**, 270, https://doi.org/10.1186/1471-2164-14-270

9   Garnica, D.P., Upadhyaya, N.M., Dodds, P.N. and Rathjen, J.P. (2013) Strategies for wheat stripe rust pathogenicity identified by transcriptome sequencing. *PLoS ONE* **8**, e67150, https://doi.org/10.1371/journal.pone.0067150

10 Xia, C., Wang, M., Cornejo, O.E., Jiwan, D.A., See, D.R. and Chen, X. (2017) Secretome characterization and correlation analysis reveal putative pathogenicity mechanisms and identify candidate avirulence genes in the wheat stripe rust Fungus Puccinia striiformis f. sp. tritici. *Front. Microbiol.* **8**, 2394, https://doi.org/10.3389/fmicb.2017.02394

11 Yildirim-Ersoy, F., Ridout, C.J. and Akkaya, M.S. (2011) Detection of physically interacting proteins with the CC and NB-ARC domains of a putative yellow rust resistance protein, Yr10, in wheat. *J. Plant. Dis. Prot.* **118**, 119–126, https://doi.org/10.1007/BF03356391

12 Liu, W., Frick, M., Huel, R., Nykiforuk, C.L., Wang, X., Gaudet, D.A. et al. (2014) The stripe rust resistance gene Yr10 encodes an evolutionary-conserved and unique CC-NBS-LRR sequence in wheat. *Mol. Plant* **7**, 1740–1755, https://doi.org/10.1093/mp/ssu112

13 Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I. et al. (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **29**, 644–652, https://doi.org/10.1038/nbt.1883

14 Pruitt, K.D., Tatusova, T. and Maglott, D.R. (2007) NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.* **35**, D61–D65, https://doi.org/10.1093/nar/gkl842

15 Bairoch, A. and 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

16 Kanehisa, M. and Goto, S. (2000) KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30, https://doi.org/10.1093/nar/28.1.27

17 Tatusov, R.L., Fedorova, N.D., Jackson, J.D., Jacobs, A.R., Kiryutin, B., Koonin E, V. et al. (2003) The COG database: an updated version includes eukaryotes. *BMC Bioinformatics* **4**, 41, https://doi.org/10.1186/1471-2105-4-41

18 Iseli, C., Jongeneel, C.V. and Bucher, P. (1999) ESTScan: a program for detecting, evaluating, and reconstructing potential coding regions in EST sequences. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* **99**, 138–148

19 Conesa, A., Götz, S., García-Gómez, J.M., Terol, J., Talón, M. and Robles, M. (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **21**, 3674–3676, https://doi.org/10.1093/bioinformatics/bti610

20 Ye, J., Fang, L., Zheng, H., Zhang, Y., Chen, J., Zhang, Z. et al. (2006) WEGO: a web tool for plotting GO annotations. *Nucleic Acids Res.* **34**, W293–W297, https://doi.org/10.1093/nar/gkl031

21 Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L. and Wold, B. (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* **5**, 621, https://doi.org/10.1038/nmeth.1226

22 Kanehisa, M., Araki, M., Goto, S., Hattori, M., Hirakawa, M., Itoh, M. et al. (2008) KEGG for linking genomes to life and the environment. *Nucleic Acids Res.* **36**, D480–D484

23 Jia, J., Zhao, S., Kong, X., Li, Y., Zhao, G., He, W. et al. (2013) Aegilops tauschii draft genome sequence reveals a gene repertoire for wheat adaptation. *Nature* **496**, 91–95, https://doi.org/10.1038/nature12028

24 Clavijo, B.J., Venturini, L., Schudoma, C., Accinelli, G.G., Kaithakottil, G., Wright, J. et al. (2017) An improved assembly and annotation of the allohexaploid wheat genome identifies complete families of agronomic genes and provides genomic evidence for chromosomal translocations. *Genome Res.* **27**, 885–896, https://doi.org/10.1101/gr.217117.116

25 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**, 3389–3402, https://doi.org/10.1093/nar/25.17.3389

26 Saunders, D.G.O., Win, J., Cano, L.M., Szabo, L.J., Kamoun, S. and Raffaele, S. (2012) Using hierarchical clustering of secreted protein families to classify and rank candidate effectors of rust fungi. *PLoS ONE* **7**, e29847, https://doi.org/10.1371/journal.pone.0029847

27 Hacquard, S., Joly, D.L., Lin, Y.-C., Tisserant, E., Feau, N., Delaruelle, C. et al. (2012) A comprehensive analysis of genes encoding small secreted proteins identifies candidate effectors in *Melampsora larici-populina* (Poplar Leaf Rust). *Mol. Plant Microbe Interact.* **25**, 279–293, https://doi.org/10.1094/MPMI-09-11-0238

28 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, https://doi.org/10.1038/nmeth.1701

29 Möller, S., Croning, M.D.R. and Apweiler, R. (2001) Evaluation of methods for the prediction of membrane spanning regions. *Bioinformatics* **17**, 646–653, https://doi.org/10.1093/bioinformatics/17.7.646

30 Sperschneider, J., Gardiner, D.M., Dodds, P.N., Tini, F., Covarelli, L., Singh, K.B. et al. (2016) EffectorP: predicting fungal effector proteins from secretomes using machine learning. *New Phytol.* **210**, 743–761, https://doi.org/10.1111/nph.13794

31 Sperschneider, J., Dodds, P.N., Gardiner, D.M., Singh, K.B. and Taylor, J.M. (2018) Improved prediction of fungal effector proteins from secretomes with EffectorP 2.0. *Mol. Plant Pathol.* **19**, 2094–2110, https://doi.org/10.1111/mpp.12682

32 Rawlings, N.D., Barrett, A.J., Thomas, P.D., Huang, X., Bateman, A. and Finn, R.D. (2018) The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database. *Nucleic Acids Res.* **46**, D624–D632, https://doi.org/10.1093/nar/gkx1134

33 Yin, Y., Mao, X., Yang, J., Chen, X., Mao, F. and Xu, Y. (2012) DbCAN: a web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res.* **40**, W445–W451, https://doi.org/10.1093/nar/gks479

34 Choi, J., Détry, N., Kim, K.T., Asiegbu, F.O., Valkonen, J.P. and Lee, Y.H. (2014) FPoxDB: Fungal peroxidase database for comparative genomics. *BMC Microbiol.* **14**, 117, https://doi.org/10.1186/1471-2180-14-117

35 Urban, M., Pant, R., Raghunath, A., Irvine, A.G., Pedro, H. and Hammond-Kosack, K.E. (2015) The Pathogen-Host Interactions database (PHI-base): additions and future developments. *Nucleic Acids Res.* **43**, D645–D655, https://doi.org/10.1093/nar/gku1165

36 Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W. et al. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539, https://doi.org/10.1038/msb.2011.75

37 Letunic, I. and Bork, P. (2016) Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* **44**, W242–W245, https://doi.org/10.1093/nar/gkw290

38 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, https://doi.org/10.1006/jmbi.2000.3903

39 Sperschneider, J., Catanzariti, A.M., Deboer, K., Petre, B., Gardiner, D.M., Singh, K.B. et al. (2017) LOCALIZER: subcellular localization prediction of both plant and effector proteins in the plant cell. *Sci. Rep.* **7**, 1–14, https://doi.org/10.1038/srep44598

40 Horton, P., Park, K.J., Obayashi, T., Fujita, N., Harada, H., Adams-Collier, C.J. et al. (2007) WoLF PSORT: protein localization predictor. *Nucleic Acids Res.* **35**, W585–W587, https://doi.org/10.1093/nar/gkm259

41 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, https://doi.org/10.1111/nph.14946

42 Karimi, M., Inzé, D. and Depicker, A. (2002) GATEWAY™ vectors for Agrobacterium-mediated plant transformation. *Trends Plant Sci.* **7**, 193–195, https://doi.org/10.1016/S1360-1385(02)02251-3

43 Win, J., Kamoun, S. and Jones, A.M.E. (2011) Purification of effector-target protein complexes via transient expression in Nicotiana benthamiana. *Methods Mol. Biol.* **712**, 181–194, https://doi.org/10.1007/978-1-61737-998-7˙15

44 Dagvadorj, B., Ozketen, A.C., Andac, A., Duggan, C., Bozkurt, T.O. and Akkaya, M.S. (2017) A Puccinia striiformis f. sp.Tritici secreted protein activates plant immunity at the cell surface. *Sci. Rep.* **7**, 1–10, https://doi.org/10.1038/s41598-017-01100-z

45 Sonah, H., Zhang, X., Deshmukh, R.K., Hossein Borhan, M., Dilantha Fernando, W.G. and Bélanger, R.R. (2016) Comparative transcriptomic analysis of virulence factors in leptosphaeria maculans during compatible and incompatible interactions with canola. *Front. Plant Sci.* **7**, 1784, https://doi.org/10.3389/fpls.2016.01784

46 Gupta, M., Sharma, G., Saxena, D., Budhwar, R., Vasudevan, M., Gupta, V. et al. (2020) Dual RNA-Seq analysis of Medicago truncatula and the pea powdery mildew Erysiphe pisi uncovers distinct host transcriptional signatures during incompatible and compatible interactions and pathogen effector candidates. *Genomics* **112**, 2130–2145, https://doi.org/10.1016/j.ygeno.2019.12.007

47 Buccitelli, C. and Selbach, M. (2020) mRNAs, proteins and the emerging principles of gene expression control. *Nat. Rev. Genet.* **21**, 630–644, https://doi.org/10.1038/s41576-020-0258-4

48 Kim, K.T., Jeon, J., Choi, J., Cheong, K., Song, H., Choi, G. et al. (2016) Kingdom-wide analysis of fungal small secreted proteins (SSPs) reveals their potential role in host association. *Front. Plant Sci.* **7**, 1–13, https://doi.org/10.3389/fpls.2016.00186

49 Liu, J., Guan, T., Zheng, P., Chen, L., Yang, Y., Huai, B. et al. (2016) An extracellular Zn-only superoxide dismutase from puccinia striiformis confers enhanced resistance to host-derived oxidative stress. *Environ. Microbiol.* **18**, 4118–4135, https://doi.org/10.1111/1462-2920.13451

50 Godfrey, D., Böhlenius, H., Pedersen, C., Zhang, Z., Emmersen, J. and Thordal-Christensen, H. (2010) Powdery mildew fungal effector candidates share N-terminal Y/F/WxC-motif. *BMC Genomics* **11**, 1–13, https://doi.org/10.1186/1471-2164-11-317

51 Petre, B., Joly, D.L. and Duplessis, S. (2014) Effector proteins of rust fungi. *Front. Plant Sci.* **5**, 1–7, https://doi.org/10.3389/fpls.2014.00416

52 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, https://doi.org/10.1105/tpc.107.051037

53 Catanzariti, A.M., Dodds, P.N., Lawrence, G.J., Ayliffe, M.A. and Ellis, J.G. (2006) Haustorially expressed secreted proteins from flax rust are highly enriched for avirulence elicitors. *Plant Cell* **18**, 243–256, https://doi.org/10.1105/tpc.105.035980

54 Dodds, P.N., Lawrence, G.J., Catanzariti, A.M., Ayliffe, M.A. and Ellis, J.G. (2004) The Melampsora lini AvrL567 avirulence genes are expressed in haustoria and their products are recognized inside plant cells. *Plant Cell* **16**, 755–768, https://doi.org/10.1105/tpc.020040

55 Link, T.I., Lang, P., Scheffler, B.E., Duke, M.V., Graham, M.A., Cooper, B. et al. (2014) The haustorial transcriptomes of Uromyces appendiculatus and Phakopsora pachyrhizi and their candidate effector families. *Mol. Plant Pathol.* **15**, 379–393, https://doi.org/10.1111/mpp.12099

56 Qi, M., Grayczyk, J.P., Seitz, J.M., Lee, Y., Link, T.I., Choi, D. et al. (2018) Suppression or Activation of Immune Responses by Predicted Secreted Proteins of the Soybean Rust Pathogen Phakopsora pachyrhizi. *Mol. Plant Microbe Interact.* **31**, 163–174, https://doi.org/10.1094/MPMI-07-17-0173-FI

57 Cheng, Y., Wu, K., Yao, J., Li, S., Wang, X., Huang, L. et al. (2017) PSTha5a23, a candidate effector from the obligate biotrophic pathogen Puccinia striiformis f. sp. tritici, is involved in plant defense suppression and rust pathogenicity. *Environ. Microbiol.* **19**, 1717–1729, https://doi.org/10.1111/1462-2920.13610

58 Ramachandran, S.R., Yin, C., Kud, J., Tanaka, K., Mahoney, A.K., Xiao, F. et al. (2017) Effectors from wheat rust fungi suppress multiple plant defense responses. *Phytopathology* **107**, 75–83, https://doi.org/10.1094/PHYTO-02-16-0083-R

59 Petre, B., Saunders, D.G.O., Sklenar, J., Lorrain, C., Krasileva, K.V., Win, J. et al. (2016) Heterologous expression screens in nicotiana benthamiana identify a candidate effector of the wheat yellow rust pathogen that associates with processing bodies. *PLoS ONE* **11**, 1–16, https://doi.org/10.1371/journal.pone.0149035