RESEARCH ARTICLE

Genome mining of the citrus pathogen *Elsinoë fawcettii*; prediction and prioritisation of candidate effectors, cell wall degrading enzymes and secondary metabolite gene clusters

Sarah Jeffress[1], **Kiruba Arun-Chinnappa**[1], **Ben Stodart**[2], **Niloofar Vaghefi**[1], **Yu Pei Tan**[3], **Gavin Ash**[1,2]*

1 Centre for Crop Health, Institute for Life Sciences and the Environment, Research and Innovation Division, University of Southern Queensland, Toowoomba, QLD, Australia, **2** Graham Centre for Agricultural Innovation, (Charles Sturt University and NSW Department of Primary Industries), School of Agricultural and Wine Sciences, Charles Sturt University, Wagga Wagga, NSW, Australia, **3** Department of Agriculture and Fisheries, Queensland Government, Brisbane, QLD, Australia

* gavin.ash@usq.edu.au

Abstract

Elsinoë fawcettii, a necrotrophic fungal pathogen, causes citrus scab on numerous citrus varieties around the world. Known pathotypes of *E. fawcettii* are based on host range; additionally, cryptic pathotypes have been reported and more novel pathotypes are thought to exist. *E. fawcettii* produces elsinochrome, a non-host selective toxin which contributes to virulence. However, the mechanisms involved in potential pathogen-host interactions occurring prior to the production of elsinochrome are unknown, yet the host-specificity observed among pathotypes suggests a reliance upon such mechanisms. In this study we have generated a whole genome sequencing project for *E. fawcettii*, producing an annotated draft assembly 26.01 Mb in size, with 10,080 predicted gene models and low (0.37%) coverage of transposable elements. A small proportion of the assembly showed evidence of AT-rich regions, potentially indicating genomic regions with increased plasticity. Using a variety of computational tools, we mined the *E. fawcettii* genome for potential virulence genes as candidates for future investigation. A total of 1,280 secreted proteins and 276 candidate effectors were predicted and compared to those of other necrotrophic (*Botrytis cinerea*, *Parastagonospora nodorum*, *Pyrenophora tritici-repentis*, *Sclerotinia sclerotiorum* and *Zymoseptoria tritici*), hemibiotrophic (*Leptosphaeria maculans*, *Magnaporthe oryzae*, *Rhynchosporium commune* and *Verticillium dahliae*) and biotrophic (*Ustilago maydis*) plant pathogens. Genomic and proteomic features of known fungal effectors were analysed and used to guide the prioritisation of 120 candidate effectors of *E. fawcettii*. Additionally, 378 carbohydrate-active enzymes were predicted and analysed for likely secretion and sequence similarity with known virulence genes. Furthermore, secondary metabolite prediction indicated nine additional genes potentially involved in the elsinochrome biosynthesis gene cluster than previously described. A further 21 secondary metabolite clusters were

predicted, some with similarity to known toxin producing gene clusters. The candidate virulence genes predicted in this study provide a comprehensive resource for future experimental investigation into the pathogenesis of *E. fawcettii*.

Introduction

Elsinoë fawcettii Bitancourt & Jenkins, a necrotrophic fungal species within the Ascomycota phylum (class Dothideomycetes, subclass Dothideomycetidae, order Myriangiales), is a filamentous phytopathogen which causes a necrotic disease, known as citrus scab, to the leaves and fruit of a variety of citrus crops around the world. Susceptible citrus varieties include lemon (*Citrus limon*), rough lemon (*C. jambhiri*), sour orange (*C. aurantium*), Rangpur lime (*C. limonia*), Temple and Murcott tangors (*C. sinensis* x *C. reticulata*), Satsuma mandarin (*C. unshiu*), grapefruit (*C. paradisi*), Cleopatra mandarin (*C. reshni*), clementine (*C. clementina*), yuzu (*C. junos*), kinkoji (*C. obovoidea*), pomelo (*C. grandis*) and Jiangjinsuanju (*C. sunki*) [1–9]. Numerous pathotypes of *E. fawcettii* are defined by host range, including the Florida Broad Host Range (FBHR), Florida Narrow Host Range (FNHR), Tyron's, Lemon, Jinguel, SRGC and SM, while cryptic and novel pathotypes are also reported [1, 3, 10]. Only the Tyron's pathotype (which infects Eureka lemon, Rough lemon, clementine, Rangpur lime and Cleopatra mandarin) and the Lemon pathotype (which only infects Eureka lemon, Rough lemon, Rangpur lime) have been described in Australia [2, 3, 7], however *E. fawcettii* has reportedly been isolated from kumquat (*Fortunella* sp.), tea plant (*Camellia sinensis*) and mango (*Mangifera indica*) [11], indicating a wider range of pathotypes to be present in Australia. Additional species of *Elsinoë* found causing disease in Australia include *E. ampelina*, which causes anthracnose to grapes [12] and two *E. australis* pathotypes; one which causes scab disease to jojoba (*Simmondsia chinensis*) [13] and a second found on rare occasions on finger lime (*C. australasica*) in Queensland forest areas [14]. Species of *Elsinoë* causing crop disease in countries neighbouring Australia include *E. batatas*, which causes large yield losses in sweet potato crops in Papua New Guinea [15, 16] and *E. pyri*, which infects apples in organic orchards in New Zealand [17]. Around the world there are reportedly 75 *Elsinoë* species, the majority of which appear to be host specific [18]. While citrus scab is not thought to affect yield, it reduces the value of affected fruit on the fresh market. Australia is known for producing high quality citrus fruits for local consumption and export, and so understandably, there is great interest in protecting this valuable commodity from disease.

Elsinoë fawcettii is commonly described as an anamorph, reproducing asexually. Hyaline and spindle shaped conidia are produced from the centre of necrotic citrus scab lesions [19, 20]. Conidia are dispersed by water splash, requiring temperatures between 23.5–27˚C with four hours of water contact for effective host infection. Therefore, disease is favoured by warm weather with overhead watering systems or rain [21]. Only young plant tissues are vulnerable to infection; leaves are susceptible from first shoots through to half expanded and similarly fruit for 6 to 8 weeks after petal fall, while mature plants are resistant to disease [19]. Cuticle, epidermal cells and mesophyll tissue are degraded within 1 to 2 days of inoculation, hyphal colonisation proceeds and within 3 to 4 days symptoms are visible [20, 22]. After formation of necrotic scab lesions on fruit, twigs and leaves, conidia are produced from the scab pustules providing inoculum for further spread. Within 5 days, host cell walls become lignified separating infected regions from healthy cells, which is thought to limit internal spread of the pathogen [20]. The necrosis that occurs during infection is produced in response to elsinochrome, a well-known secondary metabolite (SM) of species of *Elsinoë*. Elsinochromes are red or orange

pigments which can be produced in culture [23, 24]. In aerobic and light-activated conditions, reactive oxygen species are produced in response to elsinochromes in a non-host selective manner, generating an environment of cellular toxicity [25]. Elsinochrome production is required for full virulence of *E. fawcettii*, specifically the *EfPKS1* and *TSF1* genes are vital within the elsinochrome gene cluster [26, 27]. However, two points indicate that *E. fawcettii* pathogenesis is more complex than simply the result of necrotic toxin production: (I) the production of elsinochrome appears to be variable and does not correlate with virulence [28]; and (II) elsinochrome is a non-host selective toxin, yet *Elsinoë* species and *E. fawcettii* pathotypes cause disease in a host specific manner. Host-specific virulence factors targeted for interaction with distinct host proteins to overcome immune defences, prior to elsinochrome production, could explain the observed host specificity. Candidate virulence genes may include effectors and cell wall degrading enzymes. Effectors are secreted pathogen proteins, targeted to either the host cytoplasm or apoplast, which enable the pathogen to evade recognition receptor activities of the host's defence system and, if successful, infection proceeds. Resistant hosts, however, recognise pathogen effectors using resistant (R) genes which elicit plant effector-triggered immunity and pathogenesis is unsuccessful [29, 30]. While it was previously thought that necrotrophic fungal pathogens would use only a repertoire of carbohydrate-active enzymes (CAZymes) or SMs to infect host plants [31], there is increased awareness of their utilisation of secreted protein effectors [32–37], highlighting the importance of protein effector identification in all fungal pathogens. Frequently shared features of effectors include; a signal peptide at the N-terminal and no transmembrane helices or glycosylphosphatidylinositol (GPI) anchors. Other features less frequently shared include; small size, cysteine rich, amino acid polymorphism, repetitive regions, gene duplication, no conserved protein domains, coding sequence found nearby to transposable elements, and absence in non-pathogenic strains [38–45]. Furthermore, some appear to be unique to a species for example the necrosis-inducing protein effectors NIP1, NIP2 and NIP3 of *Rhynchosporium commune* [46] and three avirulence effectors AvrLm1, AvrLm6 and AvrLm4-7 of *Leptosphaeria maculans* [47]. Others have orthologous genes or similar domains in numerous species for example the chorismate mutase effector, Cmu1, of *Ustilago maydis* [48] and the cell death-inducing effector, MoCDIP4, of *Magnaporthe oryzae* [49]. Understandably, with such a large variety of potential features, effector identification remains challenging. Effectors are found in biotrophs, for example *U. maydis* [50–53], hemibiotrophs, such as *L. maculans* [54–56], *M. oryzae* [57, 58], *R. commune* [46] and *Verticillium dahliae* [59–61], necrotrophs, for example *Botrytis cinerea* [62, 63], *Parastagonospora nodorum* [34, 42, 64], *Pyrenophora tritici-repentis* [65], *Sclerotinia sclerotiorum* [32] and also the hemibiotroph/latent necrotroph *Zymoseptoria tritici* [66]. Genomic location has potential to be an identifying feature of virulence genes in some species, for example pathogenicity-related genes of *L. maculans*, including those coding for secreted proteins and genes potentially involved in SM biosynthesis, are found at higher rates in AT-rich genomic regions in comparison to GC-equilibrated blocks [47]. It is thought that effectors and their target host proteins co-evolve, in a constant arms race [67], presenting genomic regions with higher levels of plasticity as potential niches which harbour effector genes.

Another group of virulence factors likely to play a role in *E. fawcettii* pathogenesis are cell wall degrading enzymes (CWDE), these are CAZymes, including glycoside hydrolases, polysaccharide lyases and carbohydrate esterases, which can be secreted from fungal pathogens and promote cleavage of plant cell wall components [68–70]. Cell wall components, such as cellulose, hemicelluloses (xyloglucan and arabinoxylan) and pectin (rhamnogalacturonan I, homogalacturonan, xylogalacturonan, arabinan and rhamnogalacturonan II) [71], are targets for pathogens to degrade for nutrients and/or to overcome the physical barriers presented by their host. CWDE can include polygalacturonases, pectate lyases, and pectinesterases which

promote pectin degradation [72–78], glucanases (also known as cellulase) which breaks links between glucose residues [79] and xylanases which cleave links in the xylosyl backbone of xyloglucan [80–82].

Elsinoë fawcettii effectors and/or CWDE which interact with certain host plant cell wall components could explain the observed host specificity of pathotypes. Computational prediction of genes coding for such virulence factors can lead to many candidate effectors (CE) and potential CWDE, leading to an overabundance of candidates which require prioritisation. This study aimed to generate an assembly of the *E. fawcettii* isolate, BRIP 53147a, through whole genome shotgun (WGS) sequencing, to identify candidate virulence genes and appropriately shortlist these predictions to improve the focus of future experimental validation procedures. Computational methods involving genomic, proteomic and comparative analyses enabled the prediction and prioritisation of CE and CWDE which may be interacting with the host plant and overcoming immune defences prior to the biosynthesis of elsinochrome. Additional genes potentially involved in the elsinochrome gene cluster were also predicted, as were additional SM clusters which may be impacting virulence of *E. fawcettii*.

Materials and methods

Sequencing, assembly, gene prediction, annotation and genomic analyses

Elsinoë fawcettii (BRIP 53147a), collected from *C. limon* in Montville, Queensland, Australia, was obtained from DAF Biological Collections [11]. The isolate was cultured on potato dextrose agar (Difco) and incubated at 23 to 25°C for two months. Whole genomic DNA was extracted using the DNeasy Plant Mini kit (QIAGEN) according to the manufacturer's protocol. Paired-end libraries, with a mean insert size of approximately 330 bp, were prepared according to Illumina Nextera™ DNA Flex Library Prep Reference Guide using a Nextera™ DNA Flex Library Prep Kit and Nextera™ DNA CD Indexes. WGS sequencing was performed on Illumina MiSeq platform (600-cycles) at the molecular laboratories of the Centre for Crop Health, USQ. Assembly was performed on the Galaxy-Melbourne/GVL 4.0.0 webserver [83]. Raw reads were quality checked using FastQC (v0.11.5) [84] and trimmed using Trimmomatic (v0.36) [85] with the following parameters: TruSeq3 adapter sequences were removed using default settings, reads were cropped to remove 20 bases from the leading end and 65 bases from the trailing end of each read, minimum quality of leading and trailing bases was set to 30, a sliding window of four bases was used to retain those with an average quality of 30 and the minimum length read retained was 31 bases. *De novo* assembly was performed in two steps, first using Velvet (v1.2.10) [86] and VelvetOptimiser (v2.2.5) [87] with input k-mer size range of 81–101 (step size of 2). Secondly, SPAdes (v3.11.1) [88] was run on trimmed reads with the following parameters: read error correction, careful correction, automatic k-mer values, automatic coverage cutoff and Velvet contigs (>500 bp in length), from the previous step, included as trusted contigs. Contigs >500 bp in length were retained. Reads were mapped back to the assembly using Bowtie2 (v2.2.4) [89] and Picard toolkit (v2.7.1) [90] and visualised using IGV (v2.3.92) [91]. The estimated genome size was determined using Kmergenie (v1.6715) [92] on Galaxy-Australia (v19.09) [93] and GenomeScope [94]. The genome assembly was checked for completeness with BUSCO (v2.0) [95] using the Dothideomycetes orthoDB (v10) dataset [96]. The extent and location of AT-rich regions was determined using OcculterCut (v1.1) [97] with default parameters and mitochondrial contigs removed.

The prediction of genes and transposable elements (TE) was performed on the GenSAS (v6.0) web platform [98], using GeneMarkES (v4.33) [99], with fungal mode, for gene prediction and RepeatMasker (v4.0.7) [100], using the NCBI search engine and slow speed sensitivity, for the prediction of TE. Predicted gene models containing short exons, missing a start or

stop codon or which overlapped a TE region were removed from the predicted proteome. The genome was searched for Simple Sequence Repeats (SSR) using the Microsatellite Identification tool (MISA) [101], with the SSR motif minimum length parameters being 10 for mono, 6 for di, and 5 for tri, tetra, penta and hexa motifs.

Annotation was performed using BLASTP (v2.7.1+) [102] to query the *E. fawcettii* predicted proteome against the Swiss-Prot Ascomycota database (release 2018_08) [103] with an e-value of 1e-06 and word size of 3. BLAST results were loaded into Blast2GO Basic (v5.2.1) [104], with InterProScan, mapping and annotation steps being performed with default parameters, except HSP-hit coverage cutoff was set to 50% to increase stringency during annotation. Further annotation was achieved using HmmScan in HMMER (v3.2.1) [105] to query the predicted proteome against the Protein Family Database (Pfam) (release 32) [106]. GC% content of the coding DNA sequence (CDS) of each gene was determined using nucBed from Bedtools (v2.27.1) [107]. Predicted proteins were searched for polyamino acid (polyAA) repeats of at least five consecutive amino acid residues using the FIMO motif search tool [108] within the Meme suite (v5.0.2) [109]. The Whole Genome Shotgun project was deposited at DDBJ/ENA/GenBank under the accession SDJM00000000. The version described in this paper is version SDJM00000000. Raw reads were deposited under the SRA accession PRJNA496356.

Phylogenetic analysis

Two analyses were conducted, the first included three isolates of *Elsinoë fawcettii* (BRIP 53147a, DAR 70024 and SM16-1), and individual isolates of *E. ampelina*, *E. australis*, *U. maydis*, *L. maculans*, *M. oryzae*, *R. commune*, *V. dahliae*, *B. cinerea*, *Parastagonospora nodorum*, *Pyrenophora tritici-repentis*, *S. sclerotiorum* and *Z. tritici* and utilised partial TEF1-α and RPB2 regions which were obtained using BLASTN (v2.7.1+) [102] on each assembly; *Spizellomyces punctatus* was included as the outgroup. The second used ITS and partial TEF1-α sequences, obtained from GenBank, of 12 *E. fawcettii* pathotypes, 11 closely related *Elsinoë* species, and *Myriangium hispanicum* as the outgroup, for phylogenetic analysis with *E. fawcettii* (BRIP 53147a). Genome locations and GenBank accessions of all sequences are provided in S1 Table. Sequences for each locus were aligned using MUSCLE [110] with a gap open penalty of -400, concatenated and used to perform maximum likelihood analysis in MEGA7 [111] based on the General Time Reversible model [112] with partial deletion of 90% and 1000 bootstrap replicates. The initial tree for each maximum likelihood analysis was automatically selected using Neighbor-Join and BioNJ on the matrix of pairwise distances estimated using the Maximum Composite Likelihood method. A discrete Gamma distribution utilising 4 categories (+G, parameter = 0.5348 (Fig 1) and 0.4095 (Fig 2)) was used and the rate variation model allowed some sites to be invariable (+*I*, 15.4278% sites (Fig 1) and 26.6862% sites (Fig 2)). The character matrix and tree were combined and converted to nexus format using Mesquite (v3.6) [113] prior to TreeBASE submission (Fig 1 TreeBASE reviewer access: http://purl.org/phylo/treebase/phylows/study/TB2:S26086?x-access-code=26e5270bba657f28e8d78a0849503953&format=html

Fig 2 TreeBASE reviewer access: http://purl.org/phylo/treebase/phylows/study/TB2:S26087?x-access-code=8130d199a2304fe8bd684df1cc2ebacc&format=html). *E. fawcettii* (BRIP 53147a) ITS and partial TEF1-α sequences (accessions MN784182 and MN787508) were submitted to GenBank.

Sequence information

Genome assemblies and predicted proteomes included in the comparative analysis were obtained from GenBank. These included *U. maydis* (accession GCF_000328475.2, no. of
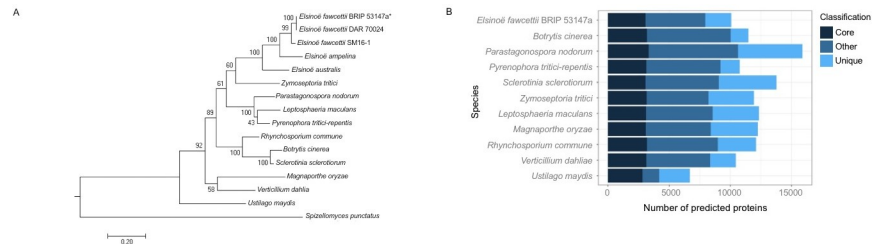
Fig 1. Species comparison. (A) Maximum likelihood phylogenetic tree of *E. fawcettii* BRIP 53147a with recently sequenced *Elsinoë* isolates and species included in the comparative study. The phylogenetic tree was inferred from a concatenated dataset including partial TEF1α and RPB2 regions. *Spizellomyces punctatus* was used as the outgroup. The branch length indicates the number of nucleotide substitutions per site, bootstrap values are shown at nodes and the isolate analysed in this study is denoted with asterisk (*). (B) Comparison of gene classifications among the proteomes of 11 fungal pathogens. Genes were categorised using orthoMCL group IDs, or proteinortho if no group was assigned. Genes were considered; (I) core if they were shared by all 11 species; (II) "other" if they were shared by at least two species, but not all; (III) unique if they were found in only one of the 11 species.

https://doi.org/10.1371/journal.pone.0227396.g001

scaffolds = 27) [114], *L. maculans* (accession GCF_000230375.1, no. of scaffolds = 76) [115], *M. oryzae* (accession GCF_000002495.2, no. of scaffolds = 53) [116], *R. commune* (accession GCA_900074885.1, no. of scaffolds = 164) [117], *V. dahliae* (accession GCF_000150675.1, no. of scaffolds = 55) [118], *B. cinerea* (accession GCF_000143535.2, no. of scaffolds = 18) [119], *Parastagonospora nodorum* (accession GCF_000146915.1, no. of scaffolds = 108) [120], *Pyrenophora tritici-repentis* (accession GCA_003231415.1, no. of scaffolds = 3964) [121], *Sclerotinia sclerotiorum* (accession GCF_000146945.2, no. of scaffolds = 37) [122] and *Z. tritici* (accession GCA_900184115.1, no. of scaffolds = 20) [123]. Additionally, genome assemblies of *E. fawcettii* DAR 70024 (accession GCA_007556565.1, no. of scaffolds = 53), *E. fawcettii* SM16-1 (accession GCA_007556535.1, no. of scaffolds = 1,266), *E. australis* Ea1 (accession GCA_007556505.1, no. of scaffolds = 21) [124] and *E. ampelina* (accession GCA_005959805.1, no. of scaffolds = 13) [125], were obtained from GenBank and gene prediction performed as for *E. fawcettii* BRIP 53147a. TE were identified in each assembly, as previously described, and predicted genes which overlapped them were similarly removed from predicted proteomes. Sequences of experimentally verified effector proteins were obtained from the EffectorP 2.0 study [126].

Prediction of secretome and effectors

Secretome and effector prediction was performed on the predicted proteomes of *E. fawcettii* and 10 fungal species known to contain effector proteins. Secretome prediction for each species began with a set of proteins predicted as secreted by either SignalP (v4.1) [127], Phobius [128] or ProtComp-AN (v6) [129]. This set was run through both the TMHMM Server (v2.0) [130] and PredGPI [131] to predict proteins with transmembrane helices and GPI-anchors, respectively. Those proteins with >1 helix or with 1 helix beyond the first 60 amino acids were removed, as were those with "highly probable" or "probable" GPI anchors. Remaining proteins formed the predicted secretome and were subjected to candidate effector prediction using EffectorP (v1.0 and v2.0) [126,132].

Genomic, proteomic and known effector analyses

Sequences of 42 experimentally verified effector proteins, which showed >98% similarity to proteins from the 10 species included in this study, and which appeared in both the predicted secretome and candidate effector list for the respective species, were utilised in the known
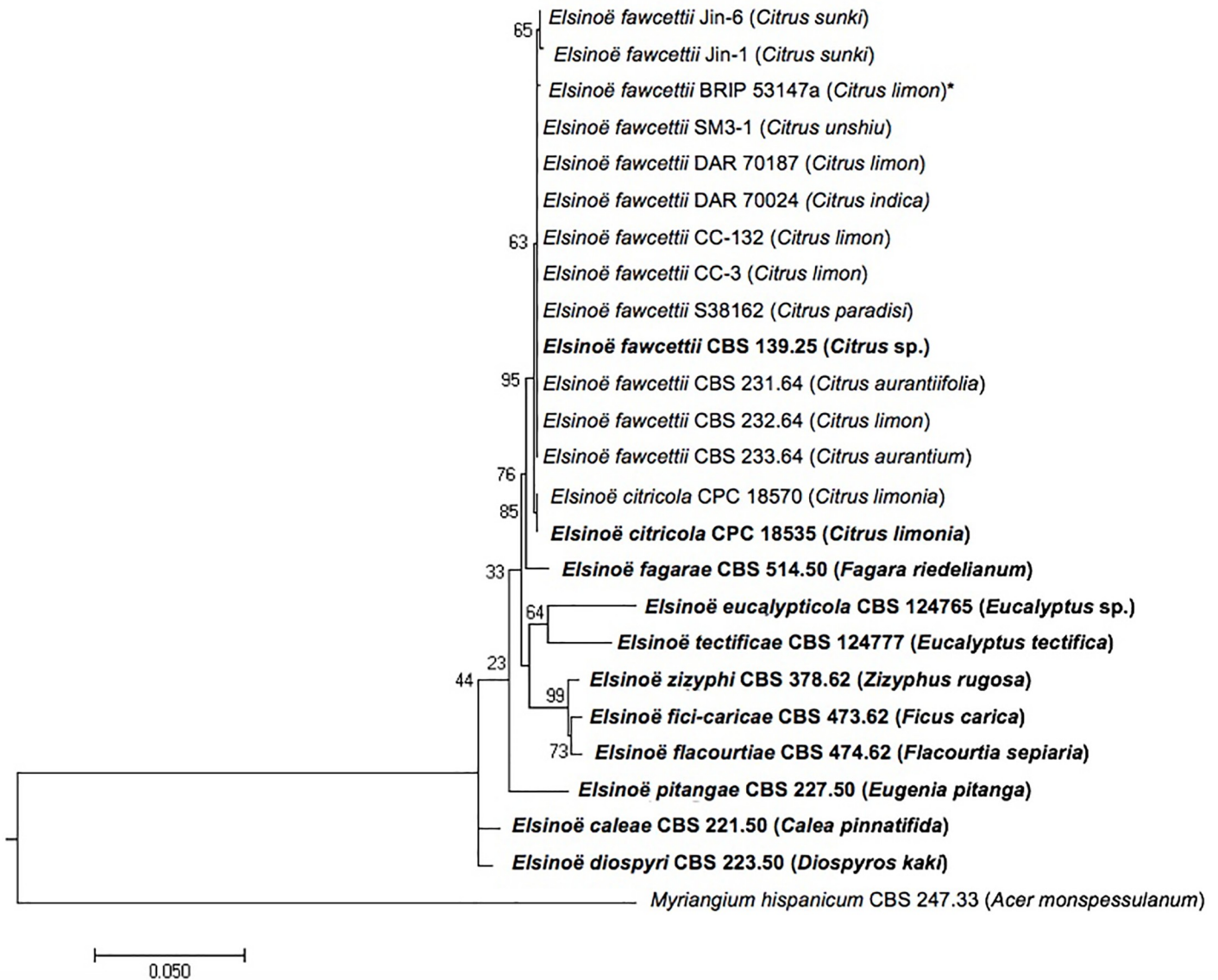
Fig 2. Maximum likelihood phylogenetic tree of *E. fawcettii* isolates and closely related species. The phylogenetic tree was inferred from a concatenated dataset including ITS and partial TEF1-α regions. *Myriangium hispanicum* was used as the outgroup. The branch length indicates the number of nucleotide substitutions per site, bootstrap values are shown at nodes, host in parentheses, new isolate described in the current study denoted with asterisk (*) and type strains are in bold.

effector analysis. The following analyses were performed on the proteome/genome of each species. Results relating to the 42 known effectors were compared to results of all proteins from each species. Length of the intergenic flanking region (IFR) was determined as the number of bases between the CDS of two adjacent genes. Genes were labelled as gene-dense if the IFR on each side was less than 1500 bp, genes on a contig end were not included among gene-dense labelled genes. Genes with IFR greater than or equal to 1500 bp were labelled as gene-sparse genes. SM clusters were predicted by passing genome assemblies and annotation files through antiSMASH fungal (v4.2.0) [133] using the Known Cluster Blast setting. Core, unique and other genes for each species were determined by grouping proteins into ortholog groups using the orthoMCL algorithm [134] followed by ProteinOrtho (v5.16b/v6.0.14) [135] on remaining unclassified genes. For the purposes of CE prioritisation, each *Elsinoë* proteome was compared

individually to the proteomes of the 10 fungal species included in the study during ortholog classification. Core genes were those shared by all species included in the comparison, unique genes were found in only one species and other genes were those shared by at least two species, but not all. Additionally, ortholog classification, using the method described above, was performed on the five species of *Elsinoë* together, to determine potential *E. fawcettii-* and isolate-specific genes. GC% content of the CDS of each gene was determined as described above, Q_1 and Q_3 values were determined for each species using R [136]. HmmScan [105] of all protein sequences against the Pfam database [106] was performed as described above. Genomic AT-rich region identification was performed using OcculterCut (v1.1) [97] as described above. For genomes with identified AT-rich regions, the distance between genes and their closest AT-rich region edge was determined using Bedtools closestBed [107], as was the distance between genes and the closest TE.

Prioritisation of candidate effectors

CE of each species were prioritised using an optimised scoring system based on the analysis of known effectors in 10 fungal species. All were scored out of at least four points, corresponding to one point allocated for each of the following conditions: (I) not labelled as gene-dense; (II) no involvement in predicted SM clusters; (III) labelled as either unique to the species or allocated to the same orthoMCL group as a known effector; and (IV) GC% of CDS was either below the Q_1 value or above the Q_3 value of the respective species. For species with genomes which had >2% TE coverage or >25% AT-rich region coverage, CE were scored out of five points. Those genomes which had both >2% TE and >25% AT-rich region coverage, CE were scored out of six points. Hence, all candidate effectors were scored out of n (four, five or six) points, those CE which obtained a score of n or n-1 points were labelled as prioritised CE. P-values, to test overrepresentation of SP and CE among *E. fawcettii*, were determined using the one-tailed Fisher's exact test in R [136].

Prediction of other virulence genes

SM clusters were predicted using antiSMASH fungal (v4.2.0) [133] as described above. CAZymes were predicted by passing the predicted proteomes through the dbCAN2 meta server [137] and selecting three tools including HMMER scan against the dbCAN HMM database [138], Diamond [139] search against the Carbohydrate-Active enZYmes (CAZy) database [140] and Hotpep query against the Peptide Pattern Recognition library [141]. Predicted CAZymes were taken as those with positive results for at least two out of the three tools. Potential pathogenesis-related proteins were identified by querying the predicted proteomes against the Pathogen Host Interactions Database (PHI-base) (v4.6, release Oct 2018) [142] using BlastP (v2.7.1) [102] analyses with an e-value of 1e-06 and a query coverage hsp of 70%, those results with >40% similarity were retained. Prioritised candidate CWDE were shortlisted from the predicted CAZymes to those which were predicted as secreted and obtained hits to plant associated fungal pathogenicity-related genes in PHI-base which showed evidence of reduced virulence in knockout or mutant experiments.

Results and discussion

Genome assembly and features

The genome assembly of *E. fawcettii* (BRIP 53147a), deposited at DDBJ/ENA/GenBank (accession SDJM00000000), was sequenced using paired-end Illumina WGS sequencing technology. Assembly of reads produced a draft genome 26.01 Mb in size with a coverage of 193x (Table 1)

Table 1. Features of *Elsinoë fawcettii* (BRIP 53147a) genome assembly.

General Features	
Assembly length (bp)	26,011,141
Coverage	193x
Number of contigs	286
Mean GC content (%)	52.3
N50 (bp)	694,004
Mean contig length (bp)	90,948
Minimum contig length (bp)	501
Maximum contig length (bp)	2,345,732
Coverage of interspersed repeats (bp)	95,654 (0.37%)
Coverage of short simple repeats (bp)	6868 (0.026%)
Number of predicted gene models	10,080
Number of contigs containing predicted genes	141
Mean gene length (bp)	1,573
Mean number of exons per gene	2.35
Number of genes containing a polyAA repeat	1,073
Mean GC content of CDS (%)	54.7

https://doi.org/10.1371/journal.pone.0227396.t001

and consisted of 286 contigs greater than 500 bp in length, with an N50 of 662,293 bp, a mean contig length of 90,948 bp and an overall GC content of 52.3%. The estimated genome size based on k-mer counts of trimmed reads was 27.25–27.28 Mb, indicating approximately 4.7% of the genome may be missing from the current assembly. Running the assembly against the Dothideomycetes orthoDB (v10) [96] showed 94.2% of complete single copy genes were found in the *E. fawcettii* assembly, indicating a high degree of coding DNA sequence completeness. The genome of *E. fawcettii* is comparable in size to other fungal genomes, including *Eurotium rubrum* (26.21 Mb) [143], *Xylona heveae* (24.34 Mb) [144] and *Acidomyces richmondensis* (29.3 Mb) [145], however it is smaller than the average Ascomycota genome size of 36.91 Mb [146]. When analysed against the 10 fungal species included in this comparative analysis (*B. cinerea*, *L. maculans*, *M. oryzae*, *Parastagonospora nodorum*, *Pyrenophora tritici-repentis*, *R. commune*, *V. dahliae*, *S. sclerotiorum*, *U. maydis* and *Z. tritici*), the *E. fawcettii* assembly is the second smallest, after the *U. maydis* assembly at 19.6 Mb. It is comparable in size to two *E. fawcettii* genomes recently published, being 26.65 Mb (SM16-1) and 26.32 Mb (DAR 70024) [124]. TE identification, by analysis against Repbase (release 18.02) [147], showed a coverage of only 0.37%, indicating a low proportion of the *E. fawcettii* genome is represented by currently known TE, this is a likely contributor to its comparatively small genome size. This low TE coverage may also be the result of a fragmented genome [148]. It is possible, should long read sequencing of this isolate be completed in the future, TE coverage may appear higher.

The *E. fawcettii* genome has less predicted gene models than the average Ascomycota genome of 11129.45 [146]. Gene prediction produced 10,080 gene models, which is comparable to the number of genes predicted for the recently published *E. fawcettii* genomes, specifically 10,340 (SM16-1) and 9930 (DAR 70024) genes [124]. A total of 5,636 (55.91%) genes were annotated, while 4,444 (44.09%) were labelled as coding for hypothetical proteins. The average gene length was 1,573 bp with an average of 2.35 exons per gene, there were 3,280 single exon genes. The mean GC content of CDS was 54.7%, which was 2.4% higher than the overall GC content and showed a wide variation in range, with the lowest scoring gene at 44.29% GC and the highest being 71.53%, thus exposing a spectrum on which genes may be differentiated. Hmmscan [105] analysis of the predicted proteome against the Pfam database

[106] revealed a high proportion (70.1% = 7,069) of genes with at least one hit to a Pfam model. The same analysis performed on the proteomes of the 10 fungal species included in the comparative analysis gave results ranging from 48.6% for *S. sclerotiorum*, with the lowest proportion of Pfam hits, to 74.9% for *U. maydis* with the highest, and a mean of 62.1% over the 11 species (S2 Table).

Phylogenetic analysis showed all species included in this comparative study are distinct from one another, with only the three *E. fawcettii* isolates being closely related (Fig 1A). Analysis of orthologous genes among *E. fawcettii* and the 10 comparative species (other *Elsinoë* species not included) indicated 3,077 (30.5%) of the predicted genes of *E. fawcettii* were core genes, finding hits through OrthoMCL or ProteinOrtho in all 11 species (S2 Table). There were 4,874 (48.4%) *E. fawcettii* genes found in at least one other species but not all and were therefore labelled as "other" genes. Lastly, the remaining 2,129 (21.1%) were found in only the *E. fawcettii* proteome, 140 of these, however, obtained a hit to an orthoMCL group and were therefore set aside and not considered as unique proteins in subsequent analyses, leaving 1,989 (19.7%) genes presumed to be *Elsinoë*-specific and therefore potentially involved in either *Elsinoë*- or *E. fawcettii*-specific pathogenesis pathways. The comparative analysis among the core, unique and other genes of the 11 species (S2 Table) (Fig 1B) indicated that *U. maydis* was set apart from the other species by showing the lowest proportion of "other" genes and the highest proportion of unique genes, this was expected as *U. maydis* was the only biotroph and Basidiomycete among the group, and is seen separated from the Ascomycota clade in the phylogenetic analysis (Fig 1A). E. *fawcettii* showed a below average percentage of unique genes which may be expected due its smaller than average sized genome and proteome. When comparing predicted genes of *E. fawcettii* (BRIP 53147a) to those of *E. ampelina*, *E. australis* and two *E. fawcettii* isolates (Table 2), 75.70% (7,631) of genes were labelled as core genes, indicating the majority of genes appeared in all five isolates. A further 12.37% (1,247) were classed as accessory, being found in more than one species, but not all five isolates. *E. fawcettii*-specific genes, found in at least two *E. fawcettii* isolates, accounted for 10.72% (1081), while *E. fawcettii* BRIP 53147a-specific genes made up 1.2% (121) of genes. The results of Table 2 indicated that the predicted gene repertoire of *E. fawcettii* isolates BRIP 53147a and DAR 70024 were closely aligned, with isolate SM16-1 set apart with a higher proportion of unique genes. As SM16-1 is classified as the FBHR pathotype and DAR 70024 as the Tyron's pathotype, SM16-1 therefore has the ability to infect a wider variety of host plants [3], these additional unique genes of *E. fawcettii* SM16-1 may contribute to its greater host range.

The overall GC content of *E. fawcettii* was 52.3%. However, when taking AT-rich regions into consideration, 98.97% of the genome had an average GC content of 52.8%, while the remaining 1.03% consisted of AT-rich regions with an average GC content of 33.8%. AT-rich regions are sections of DNA that are scattered throughout the genome and have a significantly higher AT content compared to adjacent GC equilibrated blocks [97]. The presence of AT-rich regions in genomes varies widely, for example *Sclerotinia sclerotiorum* does not show evidence of AT-rich regions [149], while 36% of the *L. maculans* genome is covered by AT-rich regions which have an average GC content of 33.9% [47]. AT-rich regions are thought to

Table 2. Summary gene classifications among five *Elsinoë* species.

Classification	*E. fawcettii* BRIP 53147a	*E. fawcettii* DAR 70024	*E. fawcettii* SM16-1	*E. australis* Ea-1	*E. ampelina* YL-1
Core	75.70%	74.79%	73.08%	80.35%	75.21%
Accessory	12.37%	12.96%	12.34%	9.67%	12.58%
E. fawcettii-specific	10.72%	10.84%	8.16%	N/A	N/A
Unique	1.20%	1.26%	6.43%	9.98%	12.21%

develop in, and nearby to, regions containing TE repeats, through Repeat-Induced Point mutation (RIP), a mechanism used to inhibit the destructive actions of TE against an organism's genome. Through a fungal genome defence mechanism causing cytosine to thymine polymorphisms, a TE repeat sequence is inhibited from further movement and potential destruction of necessary genes. This same type of polymorphism can also occur in genes nearby to TE regions [150–153], potentially providing numerous genomic locations with increased plasticity scattered throughout the genome. While RIP occurs during the sexual phase it has also been observed in asexual fungi and is thought to indicate a species reproductive history or potential [154]. AT-rich regions are present within the *E. fawcettii* genome, however the extent of their coverage in the present assembly is low, 59 regions with an average GC content of 33.8% cover only 1.03% of the genome. Sixteen regions are found overlapping TE, while four are found within 2 Kb of a TE region, meaning 33.9% of the AT-rich regions potentially represent RIP-affected regions. The remaining 66.1%, found either >2Kb away or on a contig that does not contain a predicted TE region, are potentially RIP-affected regions where the TE is no longer recognisable. The AT-rich regions of *E. fawcettii* are not scattered evenly throughout the genome, instead 29/59 (49.2%) are situated at the end of a contig and 15/59 (25.4%) cover the entire length of a contig, specifically contigs not containing genes. Two further AT-rich regions were located between the end of a contig and the beginning of the first gene and so were grouped with those located at the end of a contig. The remaining 13 regions (22.0%) were situated within a contig with genes residing on both sides. Hence, the majority either made up the end of a contig which contained genes or filled entire contigs which did not contain genes, meaning it is likely that the sequence of many *E. fawcettii* AT-rich regions contain sections of such low complexity that contig breaks result, a hypothesis which could be tested in the future using long read sequencing technology. Eight predicted genes at least partially overlap these regions and 57 are located within 2 Kb, a finding which has potential significance as AT-rich regions have been known to harbour effector genes in fungal pathogens [155, 156]. There was a large range of diversity of AT-rich region coverage among the fungal pathogens analysed in the current study; *S. sclerotiorum*, *Pyrenophora tritici-repentis*, *M. oryzae* and *U. maydis* showed no AT-rich regions; *V. dahliae* (1.5%), *B. cinerea* (4.9%), *Parastagonospora nodorum* (6.6%) and *Z. tritici* (17.3%) showed lower degrees of AT-rich coverage; while *R. commune* (29.5%) and *L. maculans* (37%) showed the greatest extent. These levels of AT-rich coverage did not appear to correlate with pathogen classification as necrotrophic, hemibiotrophic or biotrophic, nor as host-specific or broad-host range pathogens. The genomic location of AT-rich regions was, however, further included in the known effectors and candidate effectors analyses.

Identification and analysis of SSR in the *E. fawcettii* genome located 400 regions covering 6,868 bp (0.026%), 164 (41%) of which were contained within a predicted gene. Furthermore, polyAA repeats, of at least five identical and adjacent residues, were identified within 1,073 predicted protein sequences. The presence of repetitive sequences has been noted in fungal effectors [33, 45, 157] and implicated in the function and evolution of pathogenicity-related genes of other plant-associated microorganisms [158]. Hence, SSR- and polyAA-containing proteins were retained for cross-referencing against candidate effectors.

Phylogenetic analysis of partial ITS and TEF1-α regions of *E. fawcettii* (BRIP 53147a) in comparison with other *E. fawcettii* isolates and closely related *Elsinoë* species (Fig 2) indicates *E. fawcettii* (BRIP 53147a) closely aligns with the *E. fawcettii* clade. Substitutions appearing in the Jingeul pathotype isolates are not seen in isolate BRIP 53147a. One G to A substitution in the TEF1-α region sets isolate BRIP 53147a apart from the other *E. fawcettii* isolates (S3 Table), a base which is at the 3rd position of a Glu codon and hence does not result in a translational difference. This substitution in the BRIP 52147a isolate appeared with a high degree of

confidence, 100% of sequence reads aligned back to the assembly and a coverage of 241x, at this point, agreed with the substitution. While it is thought that isolate BRIP 53147a belongs to either the Lemon or Tyron's pathotype, the only two pathotypes reported in Australia [2, 3, 7], it is yet to be determined which or if it constitutes a new pathotype of its own. Aside from the one base substitution in the TEF1-α region, there would be some expected differences throughout the genomes of the *E. fawcettii* BRIP 53147a isolate and the other *E. fawcettii* isolates due to differences in collection details, such as geographical location, year and host specificity. Specifically, isolate BRIP 53147a was collected in Montville, Queensland in 2009, while the other Australian isolates, DAR 70187 and DAR 70024, belonging to the Lemon and Tyron's pathotypes, were collected 15 years earlier in Somersby and Narara in NSW, respectively [7], both a distance of almost 1000 km away. Several isolates from Fig 2 have been tested for host pathogenicity leading to the designation of specific pathotypes [3], as opposed to relying on only sequence data and thus illustrating the importance of experimental validation prior to pathotype or species classification. For example, Jin-1 and Jin-6 are classified as the Jingeul pathotype, SM3-1 as FBHR, S38162 as FNHR, CC-132 as SRGC, DAR 70187 and CC-3 as the Lemon pathotype, and DAR 70024 as Tyron's pathotype [3]. Host specificity experimentation for the *E. fawcettii* BRIP 53147a isolate is a suggested future step, as is the whole genome sequencing and analysis of further *E. fawcettii* isolates for comparison. The comprehensive host pathogenicity testing of 61 *E. fawcettii* isolates and their subsequent classification into six pathotypes [3] coupled with genomic sequencing data analysis would provide a wealth of knowledge of potential host-specific pathogenicity-related genes and mutations.

Prediction of secretome and effectors

A total of 1,280 genes (12.7% of the proteome) were predicted to code for secreted proteins (SP) in the *E. fawcettii* (BRIP 53147a) genome (Table 3). Using the discovery pipeline outlined in Fig 3, classically secreted proteins with a detectable signal peptide were predicted by either SignalP and/or Phobius providing 1,449 proteins, while ProtComp identified a further 120 as potential non-classically secreted proteins. Of these 1,569 proteins, 186 were removed as they were predicted to contain transmembrane helices, an indication that while targeted for secretion, the protein likely functions while situated in the cell membrane. A further 103 were removed as they contained a predicted GPI anchor, also suggesting they associate with the cell membrane to perform their function, leaving a total of 1,280 proteins identified as likely SP. To enable comparison of the species' predicted secretomes and CE, the same prediction pipeline (Fig 3) was used on the proteomes of 14 further fungal species included in the analysis (Table 3), essentially utilising additional *Elsinoë* genomes and genomes which contain known protein effectors for comparison. The proportion of predicted SP in the *E. fawcettii* proteome was similar to that of other necrotrophic fungal pathogens, which ranged from *B. cinerea* at 11.3% to *Parastagonospora nodorum* at 13.9%.

Known effectors were frequently identified by the CE pipeline (Fig 3), with 43/45 (95.6%) correctly predicted as being secreted and 42/45 (93.3%) also predicted as effectors (Table 3). This high proportion of predicted effectors is due to 23 being utilised as positive training data for EffectorP (v2), the unbiased sensitivity and specificity of EffectorP (v2) is reportedly 84.5% and 82.8%, respectively [126]. Those known effectors which were tested but not identified as SP included Vdlsc1 (*V. dahliae*) and MoCDIP2 (*M. oryzae*). Vdlsc1 lacks an N-terminal signal peptide and is unconventionally secreted [159], however it was not identified as a non-classically secreted protein. MoCDIP2 was removed as it obtained a GPI-anchor hit. Additionally, Eff1-1 (*U. maydis*) was predicted as secreted but not as a candidate effector, Eff1-1, along with MoCDIP2, are both known false negatives of EffectorP 2.0 [126].

Table 3. Predicted Secreted Proteins (SP), Candidate Effectors (CE) and known effectors.

Species	Total proteins[*]	SP (% of total)	CE (% of SP)	Known effectors correctly predicted as SP and CE	Known effectors not predicted as SP and CE
Necrotrophs:					
Elsinoë fawcettii BRIP 53147a	10,080	1,280 (12.7%)	276 (21.6%)	-	
Elsinoë fawcettii DAR 70024	10,223	1,291	274	-	
Elsinoë fawcettii SM16-1	10,519	1,393	317	-	
Elsinoë australis	9,253	1,091	235	-	
Elsinoë ampelina	9,804	1,167	270	-	
Botrytis cinerea	11,481	1,294 (11.3%)	285 (22.0%)	NEP1	
Parastagonospora nodorum	15,878	2,206 (13.9%)	932 (42.2%)	Tox1, ToxA	
Pyrenophora tritici-repentis	10,771	1,298 (12.1%)	388 (29.9%)	ToxB	
Sclerotinia sclerotiorum	13,770	1,707 (12.4%)	692 (40.5%)	SsSSVP1	
Zymoseptoria tritici	11,936	1,514 (12.7%)	597 (39.4%)	Zt6, AvrStb6	
Hemibiotrophs:					
Leptosphaeria maculans	12,337	1,883 (15.3%)	787 (41.8%)	AvrLM6, AvrLM11, AvrLM4-7	
Magnaporthe oryzae	12,236	2,263 (18.5%)	1055 (46.6%)	SPD10, Msp1, BAS1, SPD4, SPD2, MoCDIP3, MoCDIP4, AVR-Pik, MoCDIP1, Bas107, BAS2, BAS3, BAS4, Avr-Pita1, Bas162, MoHEG13, SPD7, MC69, AvrPi9, AvrPiz-t, SPD9, MoCDIP5	MoCDIP2
Rhynchosporium commune	12,100	1,510 (12.5%)	509 (33.7%)	NIP1, NIP2, NIP3	
Verticillium dahliae	10,441	1,407 (13.5%)	413 (29.4%)	PevD1, VdSCP7	Vdlsc1
Biotroph:					
Ustilago maydis	6,692	856 (12.8%)	256 (29.9%)	Pit2, Pep1, See1, Cmu1, Tin2	Eff1-1

[*]Not including gene models which overlap a predicted TE region

The total number of CE identified for *E. fawcettii* (BRIP 53147a) was 276, meaning only 21.6% of SP gained CE classification, this was the lowest proportion out of all 11 species analysed (Table 3). This may be explained by the potential favouring of EffectorP towards SP of species on which it was trained. To further investigate this potential, results of EffectorP for the 11 species were compared to the results of an alternate candidate effector search; SP with a protein length less than the species' median and with no Pfam hit other than to that of a known effector (S4 Table). While this second method resulted in the identification of a higher number of CE for each species, *E. fawcettii* still obtained the lowest proportion of CE out of predicted SP. It also highlighted the advantage of using EffectorP to narrow down an extensive catalogue of SP, as opposed to identifying CE based on arbitrary features. However, the CE predicted by EffectorP still range in the hundreds (Table 3), it was therefore beneficial to further shortlist candidates for prioritisation. To achieve this, known effectors which were correctly predicted as both SP and as CE (Table 3) were retained for further analysis to generate an optimised prioritisation scoring system. In the current study, positive results from either version of EffectorP (1.0/2.0) [126,132] formed the CE set, while this provided a larger group
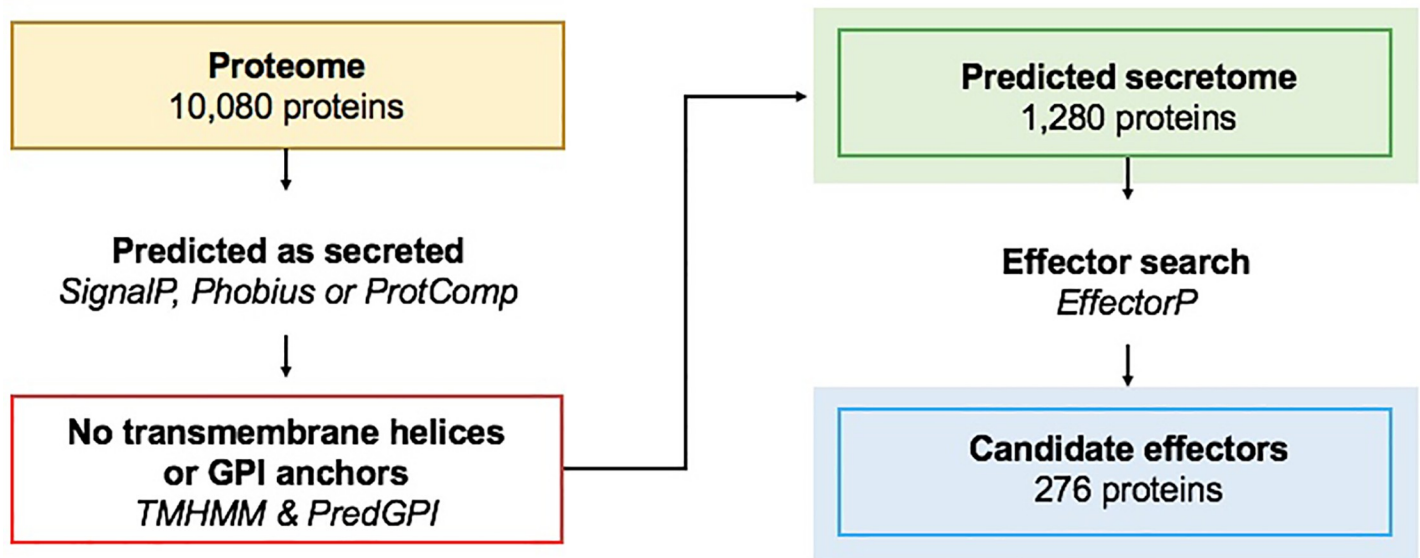
Fig 3. Pipeline for the discovery of the predicted secretome and candidate effectors. The secretome search started with the predicted proteins of a species, proteins were predicted as secreted using at least one of three tools, proteins with predicted transmembrane helices or GPI-anchors were removed. Candidate effectors were predicted using EffectorP (v1.0 and v2.0). The number of proteins shown for the predicted proteome, secretome and effectome refers to the *Elsinoë fawcettii* BRIP 53147a genome.

for subsequent prioritisation, it removed potential user discrimination based on arbitrary features. EffectorP 1.0 has previously been shown to predict effectors with a shorter average sequence length compared to version 2.0 and, additionally, selecting only CEs predicted by both versions' favours proteins with a higher cysteine content [126]. By utilising a CE set predicted by either version potentially allows greater variety of CE prediction.

Known effector analysis

A total of 42 known effectors from 10 fungal species were analysed for; (I) gene density; (II) GC content; (III) involvement in SM clusters; (IV) uniqueness; (V) distance to the closest TE; and (VI) distance to the closest AT-rich regions (Table 4). Results were compared to those of all predicted genes from each of the same 10 species (S5 Table). Features observed at a higher rate among the known effector group compared with each species' proteome were used to prioritise CE using a point allocation system. (I) Genes were labelled as gene-dense if the IFRs on both sides were less than 1500 bp. The proportions of gene-dense genes ranged from 17.66% (*Pyrenophora tritici-repentis*) to 70.23% (*L. maculans*) (S5 Table) with a mean of 49.4%, in contrast to 9/42 (21.4%) known effectors (Table 4). This provided grounds to allocate one point to each known effector which was not labelled as gene-dense. (II) GC content of the CDS of each gene was determined and median values calculated for each species, revealing the GC percentage of 32/42 (76.2%) known effectors fell either below the Q_1 value or above the Q_3 value for the respective species. When compared to an expected 50% in the upper and lower quartiles, this provided reason for the allocation of one point to known effectors should they fall in these two quartiles. (III) No overlap was observed between known effectors and the predicted SM clusters within each species, giving strong reason for the allocation of one point to known effectors that were not included in SM clusters. (IV) Analysis of gene classification (core, unique or other) for each known effector highlighted that 41/42 (97.6%) were either unique to the species (31/42) or were assigned an orthoMCL group ID of a known effector

Table 4. Features of known fungal effectors used to guide candidate effector prioritisation.

Effector	Gene density class	CDS GC %	Within SM gene cluster	Ortholog class	Distance to TE	Distance to AT-rich region	Total possible points (*n* points)	Points scored
Necrotrophic:								
***Botrytis cinerea*:**								
NEP1	Sparse	>Q_3	No	Other[A]	16,435	N/A	5	5
***Parastagonospora nodorum*:**								
Tox1		<Q_1	No	Unique	N/A	N/A	4	4
ToxA	Sparse	<Q_1	No	Unique	N/A	N/A	4	4
***Pyrenophora tritici-repentis*:**								
ToxB		<Q_1	No	Unique	N/A	N/A	4	4
***Sclerotinia sclerotiorum*:**								
SsSSVP1	Sparse	Q_2[B]	No	Unique	8,521	N/A	5	4
***Zymoseptoria tritici*:**								
Zt6	Dense[B]	>Q_3	No	Core[A]	14,100	N/A	5	4
AvrStb6	Sparse	>Q_3	No	Unique	3,166	N/A	5	5
Hemibiotrophic:								
***Leptosphaeria maculans*:**								
AvrLM6		<Q_1	No	Unique	3,766	0	6	6
AvrLM11	Sparse	<Q_1	No	Unique	2,467	0	6	6
AvrLM4-7	Sparse	<Q_1	No	Unique	891	0	6	6
***Magnaporthe oryzae*:**								
SPD10	Dense[B]	Q_2[B]	No	Unique	8,747	N/A	5	3[C]
Msp1	Dense[B]	>Q_3	No	Other[A]	39,744[B]	N/A	5	3[C]
BAS1	Sparse	<Q_1	No	Unique	249	N/A	5	5
SPD4		<Q_1	No	Unique	1,038	N/A	5	5
SPD2	Dense[B]	>Q_3	No	Unique	17,554	N/A	5	4
MoCDIP3	Sparse	>Q_3	No	Unique	168	N/A	5	5
MoCDIP4	Sparse	>Q_3	No	Other[A]	238	N/A	5	5
AVR-Pik	Sparse	<Q_1	No	Unique	442	N/A	5	5
MoCDIP1	Sparse	>Q_3	No	Other[A]	68,564[B]	N/A	5	4
Bas107		<Q_1	No	Unique	7,541	N/A	5	5
BAS2	Dense[B]	Q_2[B]	No	Other[B]	4,583	N/A	5	2[C]
BAS4	Sparse	Q_2[B]	No	Unique	3,898	N/A	5	4
BAS3		Q_2[B]	No	Unique	12,126	N/A	5	4
Avr-Pita1		<Q_1	No	Other[A]	299	N/A	5	5
Bas162		<Q_1	No	Unique	8,604	N/A	5	5
MoHEG13		<Q_1	No	Unique	5,888	N/A	5	5
SPD7		<Q_1	No	Unique	8,963	N/A	5	5
MC69	Sparse	>Q_3	No	Other[A]	18,884	N/A	5	5
AvrPi9	Dense[B]	>Q_3	No	Other[A]	5,031	N/A	5	4
AvrPiz-t		Q_2[B]	No	Unique	465	N/A	5	4
SPD9		Q_2[B]	No	Unique	3,433	N/A	5	4
MoCDIP5	Dense[B]	>Q_3	No	Other[A]	5,123	N/A	5	4
***Rhynchosporium commune*:**								
NIP3		Q_2[B]	No	Unique	32,352[B]	1,368	6	4[C]
NIP1	Sparse	>Q_3	No	Unique	2,611	1,814	6	6
NIP2	Sparse	>Q_3	No	Unique	1,860	6,572	6	6
***Verticillium dahliae*:**								

(*Continued*)

Table 4. (Continued)

Effector	Gene density class	CDS GC %	Within SM gene cluster	Ortholog class	Distance to TE	Distance to AT-rich region	Total possible points (*n* points)	Points scored
PevD1		$>Q_3$	No	Other[A]	N/A	N/A	4	4
VdSCP7		Q_2[B]	No	Unique	N/A	N/A	4	3
Biotrophic:								
Ustilago maydis:								
Pit2		$<Q_1$	No	Unique	N/A	N/A	4	4
Pep1	Sparse	Q_2[B]	No	Unique	N/A	N/A	4	3
See1	Dense[B]	$<Q_1$	No	Unique	N/A	N/A	4	3
Cmu1	Dense[B]	$>Q_3$	No	Unique	N/A	N/A	4	3
Tin2		$<Q_1$	No	Unique	N/A	N/A	4	4

[A] Allocated the same orthoMCL group ID as a known effector

[B] Possible point not allocated

[C] Less than *n*-1 points scored

(10/42). In contrast, the proportion of unique genes for each species was much lower, ranging from 11.9% (*B. cinerea*) to 33.7% (*S. sclerotiorum*), with an average of 25.4%. The proportion of genes allocated an orthoMCL of a known effector was similarly low at less than 0.3% for all species. Thus, a point was allocated to known effectors that were either unique to the species or obtained the same orthoMCL ID of a known effector. (V) Those genomes with >2% TE coverage also showed a high proportion of known effectors in the close vicinity of TE. Specifically, 29/32 (90.6%) known effectors from *Z. tritici*, *S. sclerotiorum*, *B. cinerea*, *R. commune*, *L. maculans* and *M. oryzae* were within 20 Kb of a TE region, compared to an average of 47.1% of genes within 20 Kb of a TE for the same six species. This led to the allocation of one point for known effectors less than 20 Kb from a TE for species with >2% TE coverage. (VI) Lastly, of the genomes analysed, only those consisting of >25% AT-rich regions, being *R. commune* and *L. maculans*, were found to have a noticeable association between the location of known effectors and AT-rich regions. The distance of all known effectors to the closest AT-rich region, of these two species, were found to be less than the Q_1 value for each species. Hence, known effectors with these specifications, in species with >25% AT-rich region coverage, were allocated one point. It can be seen that depending on the degree of TE and AT-rich region coverage, each species' known effectors may be scored out of four, five or six points, henceforth referred to as "*n* points". Over the 10 species with known effectors which were analysed, Table 4 illustrates a total of 38/42 (90.5%) known effectors obtained *n* or *n*-1 points, revealing a process which could be used to prioritise the many CE predicted for the *E. fawcettii* genome.

Prioritisation of candidate effectors

CE were prioritised using the method described above for the analysis of known effectors. The three *E. fawcettii* isolates, *E. ampelina*, *E. australis*, *Parastagonospora nodorum*, *Pyrenophora tritici-repentis*, *V. dahlia* and *U. maydis* each had <2% TE coverage and <25% coverage of AT-rich regions, their CE were therefore scored out of four points. *Z. tritici*, *S. sclerotiorum*, *B. cinerea* and *M. oryzae* had >2% TE coverage but <25% coverage of AT-rich regions and so were scored out of five points. Only the assemblies of *R. commune* and *L. maculans* showed >2% TE and >25% AT-rich regions, and as such their CE were scored out of six points. By using *n* or *n*-1 points as an acceptable score for CE prioritisation, CE of the 15 pathogens

Fig 4. Candidate effector prioritisation features and points. The candidate effectors (CE) of all genomes analysed were scored using features shown in the blue box. Additional features were considered for CE from genomes with >2% TE coverage (red box) and >25% AT-rich region coverage (green box).

https://doi.org/10.1371/journal.pone.0227396.g004

could be reduced, by 31.96% - 77.13% (average 54.39%) (S6 Table), with species that were scored out of more points achieving higher reductions.

Applying the method outlined in Fig 4 to the CE of *E. fawcettii* led to the prioritisation of 120 CE, a reduction of 56.5%, for future experimental validation. This is a comparable reduction to that of the other necrotrophic pathogens (Fig 5, S6 Table), for which six out of seven known effectors were retained within the shortlisted CE. Features of the 120 CE of *E. fawcettii* (BRIP 53147a) (S7 Table) indicated many were small in size, had a high GC content, had a high proportion of cysteine residues and were more likely to be classified as gene-sparse. The median protein length was 180 aa, compared to 409 aa for all *E. fawcettii* predicted genes. The mean GC content was 57.05% and the mean cysteine content was 2.9%, compared to 54.16% and 1.2%, respectively for all predicted genes of *E. fawcettii*. The high proportion (18.3%) of gene-sparse genes among prioritised CE was expected, as CE which were not classified as gene-dense were favoured during the prioritisation process, however high proportions of gene-sparse genes were also observed among the SP and CE (Table 5). Specifically, 4.0% of all *E. fawcettii* (BRIP 53147a) predicted genes were classed as gene-sparse, 72.6% as gene-dense and the remaining 23.4% classed as neither. In comparison, 4.8% of SP ($p = 0.09548$) and 8.3% of CE ($p = 0.00082$) were classed as gene-sparse, potentially indicating a preference for gene-sparse locations by CE and proteins likely secreted by the pathogen. PolyAA repeat-containing proteins were not overrepresented among the prioritised CE, only six (5.0%) were found to contain five or more consecutive amino acids, compared to 10.6% of all proteins. Additionally, no CE were found to contain SSR suggesting that diversity of *E. fawcettii* effector sequences is not being generated through an increased mutational rate related to short repetitive sequences. Furthermore, the prioritised CE were found scattered throughout the genome over 45 of the 141 gene-containing contigs and did not appear to cluster together. While AT-rich regions were not taken into consideration during the prioritisation of *E. fawcettii* CE, due to a low AT-rich coverage of 1.03%, it should be noted that a higher proportion of CE were found among genes on the end of a contig, and significantly more SP and CE were located within 2 Kb of an AT-rich region, than expected. Out of the 252 genes found at the end of a contig 14 (5.6%, $p = 0.00947$) were CE, compared to 2.7% out of all *E. fawcettii* proteins. Additionally, of the 57 genes found within 2 Kb of an AT-rich region (including those found to overlap an AT-rich region), 12 (21.1%, $p = 0.05148$) were SP and six (10.5%, $p = 0.00449$) were CE (S7 Table). This suggests that genomic regions near contig breaks, such as sequences of low complexity or
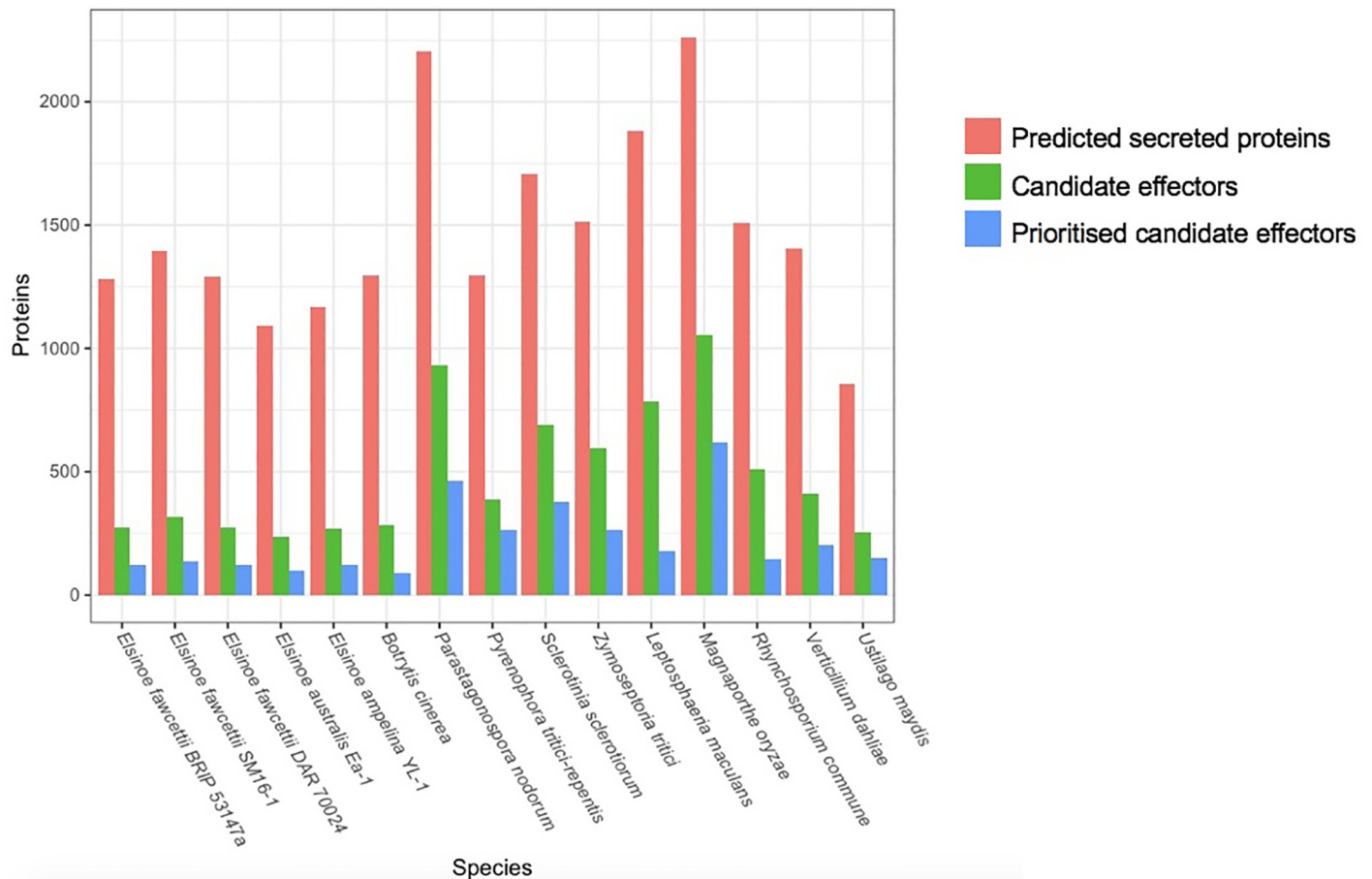
Fig 5. Comparison of numbers of secreted proteins, candidate effectors and prioritised candidate effectors among 15 fungal pathogens. Secreted proteins and candidate effectors were predicted using the pipeline in Fig 3. Prioritised candidate effectors were determined using features shown in Fig 4.

https://doi.org/10.1371/journal.pone.0227396.g005

regions under-represented by short read sequencing technology, and AT-rich regions may be indicators within the *E. fawcettii* genome of nearby effector genes. Interestingly, SP and CE were not overrepresented among genes found within 2 Kb of a predicted TE region, of the 120 genes found in these regions 12 (10%) were SP and 3 (2.5%) were CE, both slightly less than their proportions across the whole genome. This suggested while potential effector genes are more likely to be found near AT-rich regions, a nearby predictable TE region was not necessary. Thus, *E. fawcettii*, a necrotrophic pathogen not considered at first thought to utilise protein effectors to increase virulence, shows a subtle, yet intriguing, pattern of CE near AT-rich regions, at contig ends and in more gene-sparse locations. This potentially points towards a set of virulence-related genes being maintained in specific genomic locations and therefore suggesting their potential significance.

Table 5. Gene density classification of *Elsinoë fawcettii* (BRIP 53147a) predicted proteins.

Classification	All predicted proteins	Secreted proteins	Candidate effectors	Prioritised candidate effectors
Gene-sparse	4.0%	4.8%	8.3%	18.3%
Gene-dense	72.6%	68.5%	64.5%	35.0%
Neither	23.4%	26.7%	27.2%	46.6%

https://doi.org/10.1371/journal.pone.0227396.t005

While analysing proteins using the features mentioned above can shortlist CE, awareness of limitations should be considered. For example, only prioritising CE which are unique to a species, or obtain the same orthoMCL hit as a known effector, limits the identification of novel effectors which may be utilised by multiple species. Hence, a blast search of *E. fawcettii* CE against CE of the 10 other fungal pathogens was conducted and indicated 16 (5.8%) *E. fawcettii* CE had >70% similarity to at least one candidate effector of another species (S7 Table). Five of these 16 proteins were prioritised CE, one of which had 72.9% similarity to MoCDIP1 (*M. oryzae*), a known effector which is expressed in planta and induces host cell death [49], thus highlighting this CE for further investigation. Cross referencing the *E. fawcettii* (BRIP 53147a) CE with PHI-base showed 7.2% (20) obtained a hit with >40% similarity, the majority (17/20) of which were core genes found among the *Elsinoë* species studied (S7 Table), indicating a subset of CE which may be of benefit to multiple pathogens.

Prediction and prioritisation of cell wall degrading enzymes

Further potential pathogenicity-related genes of *E. fawcettii* which deserve attention include CWDE. The *E. fawcettii* (BRIP 53147a) proteome showed 378 (3.75%) predicted CAZymes (S8 Table), comparable to the proportion of CAZymes seen in the other 10 pathogen genomes, which ranged from 2.8% (*S. sclerotiorum*) to 4.3% (*V. dahliae*) (S2 Table). Of the total *E. fawcettii* CAZymes, 203 (53.7%) were also predicted as secreted, highlighting numerous potential CWDE secreted by the pathogen and targeted for interaction with host carbohydrates. It would be beneficial to compare these potential CWDE with transcriptomic data once available, however, currently they can be cross-referenced against the Pfam database. Analysis of the 203 potential CWDE revealed frequently appearing Pfam hits to pectate lyase and pectinesterase (19 hits), the glycosyl hydrolases family 28 of pectin-degrading polygalacturonases (11 hits) and the glycosyl hydrolases family 43 of hemicellulose-degrading beta-xylosidases (10 hits). Hemicellulose- and pectin-degrading enzymes target plant cell wall components including xyloglucans and pectin's, respectively [68], both found in high proportions in the primary cell wall, potentially revealing an arsenal of CWDE of *E. fawcettii* which are targeted towards young plant tissues. Polygalacturonases break bonds between polygalacturonic acid residues, thereby degrading pectin, while beta-xylosidases hydrolyse xylan, a hemicellulose component of the cell wall. It is possible that the CWDE of *E. fawcettii* have the ability to degrade components of a growing cell wall, however as the host cell wall matures, the *E. fawcettii* CWDE repertoire becomes less effective, perhaps explaining why only young plant tissues are susceptible to citrus scab. The 203 potential CWDE were also cross-referenced against PHI-base, resulting in the prioritisation of 21 proteins which had similarity to known virulence factors of plant pathogens (Table 6, S8 Table), thus highlighting candidate virulence genes of *E. fawcettii* for future experimental investigation. Among these 21 proteins were 14 predicted pectin-degrading enzymes, including two with similarity to polygalacturonase genes, specifically *pg1* (53.7%) and *pgx6* (66.4%) of *Fusarium oxysporum* which have been shown to reduce pathogen virulence when both are mutated simultaneously [74]; two showed similarity (61.6% and 41.8%) to the *PecA* polygalacturonase gene of *Aspergillus flavus*, a CWDE which primarily degrades pectin, and has been shown to improve pathogen invasion and increase spread during infection [73]; one with similarity to the pectin methylesterase *Bcpme1* gene of *B. cinerea* [78]; four with similarity (45.7% - 63.5%) to *PelA* and *PelD*, two pectate lyase virulence factors of *Nectria haematococca* [75]; and a further five obtained a pectate lyase Pfam hit, of which four showed similarity (40.3% - 53.5%) to the *Pnl1* pectin lyase gene of citrus pathogen *Penicillium digitatum* [76] and one with 58.4% similarity to *PelB* pectate lyase B gene of *Colletotrichum gloeosporioides*, seen to affect virulence on avocado [77]. A further five prioritised candidate CWDE,

Table 6. Predicted function of prioritised candidate cell wall degrading enzymes of *Elsinoë fawcettii*.

Gene accession	PHI-base hit	Similarity (%)	Top Pfam hit
Predicted pectin-degrading enzymes:			
KAF4548260	PGX6 *Fusarium oxysporum* (PHI:4880)	66.39	Glycosyl hydrolases family 28 (GH28)
KAF4556463	PG1 *F. oxysporum* (PHI:4879)	53.69	GH28
KAF4550523	PECA *Aspergillus flavus* (PHI:88)	61.64	GH28
KAF4547067	PECA *A. flavus* (PHI:88)	41.80	GH28
KAF4547800	BCPME1 *Botrytis cinerea* (PHI:278)	47.97	Pectinesterase
KAF4549166	PelD *Nectria haematococca* (PHI:180)	47.27	Pectate lyase (PL)
KAF4552448	PelD *N. haematococca* (PHI:180)	63.45	PL
KAF4550092	PelA *N. haematococca* (PHI:179)	46.38	PL
KAF4548090	PelA *N. haematococca* (PHI:179)	45.69	PL
KAF4549258	PNL1 *Penicillium digitatum* (PHI:3226)	53.46	PL
KAF4556657	PNL1 *P. digitatum* (PHI:3226)	44.74	PL
KAF4555488	PNL1 *P. digitatum* (PHI:3226)	41.70	PL
KAF4556483	PNL1 *P. digitatum* (PHI:3226)	40.33	PL
KAF4549223	PELB *Colletotrichum gloeosporioides* (PHI:222)	58.40	PL
Predicted Hemicellulose-degrading enzymes:			
KAF4552838	Endo-1,4-beta-xylanase *Magnaporthe oryzae* (PHI:2204)	61.56	Glycosyl hydrolase family 10 (GH10)
KAF4550100	Endo-1,4-beta-xylanase *M. oryzae* (PHI:2204)	57.69	GH10
KAF4555167	Endo-1,4-beta-xylanase *M. oryzae* (PHI:2208)	46.67	GH10
KAF4547778	Endo-1,4-beta-xylanase I *M. oryzae* (PHI:2214)	58.87	Glycosyl hydrolases family 11 (GH11)
KAF4556368	Endo-1,4-beta-xylanase I *M. oryzae* (PHI:2213)	56.72	GH11
Predicted Cellulose-degrading enzymes:			
KAF4547532	GLU1 *Pyrenophora tritici-repentis* (PHI:3859)	52.89	Cellulase—glycosyl hydrolase family 5 (GH5)
KAF4552889	GLU1 *P. tritici-repentis* (PHI:3859)	51.93	Cellulase–GH5

classed as hemicellulose-degrading enzymes, showed similarity (46.7% - 61.6%) to the endo-1,4-beta-xylanases (glycosyl hydrolase families 10 and 11) of *M. oryzae*, the knockdown of which is seen to reduce pathogenicity [80]. The remaining two prioritised CWDE, classed as cellulose-degrading enzymes, showed 51.9% and 52.9% similarity to the *Glu1* glucanase gene, a known virulence factor of wheat pathogen *Pyrenophora tritici-repentis* [79]. The similarities seen between these predicted secreted CAZymes and known virulence factors provides a collection of likely CWDE of *E. fawcettii* for future investigation. Unlike SP or CE, predicted CWDE of *E. fawcettii* were not overrepresented among genes found at the contig end or within 2 Kb of an AT-rich region (S8 Table). There was some crossover between CE and CWDE, with three *E. fawcettii* (BRIP 53147a) proteins being labelled as both prioritised CE and prioritised CWDE, thus providing some CE with potential carbohydrate-interacting functions.

Prediction of secondary metabolite clusters

Much research surrounding *E. fawcettii* has focused on the SM elsinochrome, which contributes to the formation of necrotic lesions [25–28]. Analysis of the *E. fawcettii* (BRIP 53147a) genome assembly enabled the prediction of further genes potentially involved in the elsinochrome gene cluster than previously described, as well as the prediction of additional SM clusters throughout the assembly. In total, there were 22 predicted SM clusters, involving 404 (4.0%) genes (Table 7, S9 Table). Comparing this to the results of the 10 comparative species showed that the number of predicted SM clusters varies widely among the pathogens, from 13 clusters (*U. maydis*) to 53 clusters (*M. oryzae*) (Fig 6). This wide variety among fungal species,

Table 7. Predicted Secondary Metabolite (SM) gene clusters of *Elsinoë fawcettii*.

Cluster #	SM class	Genomic location (number of genes involved)	Similarity to known SM biosynthetic gene clusters		
			Known SM cluster gene (GenBank accession)	Similarity (%)	*E. fawcettii* GenBank accession
1	T1PKS	SDJM01000001, 641093:686753 (15 genes)	**Elsinochrome A/B/C:**		
			EfHP1 hypothetical protein (ABZ82009.1)	97	KAF4556313
			ESC reductase (ABZ01830.1)	100	KAF4556314
			Transcription factor (ABZ01831.1)	98	KAF4556315
			Polyketide synthase (ABU63483.1)	99	KAF4556316
			ESC prefoldin protein subunit 3 (ABZ01833.1)	100	KAF4556317
			ECT1 transporter (ABZ82008.1)	70	KAF4556318
2	terpene-T1PKS	SDJM01000001, 1100227:1205433 (43 genes)	**PR toxin:**		
			Short-chain dehydrogenase/reductase SDR (CDM31317.1)	54	KAF4556505
			Aristolochene synthase (CDM31315.1)	60	KAF4556513
			FAD-binding, type 2 (CDM31316.1)	42	KAF4556518
3	other	SDJM01000002, 204508:248496 (18 genes)			
4	other	SDJM01000002, 1497538:1541073 (22 genes)			
5	terpene	SDJM01000003, 564086:586459 (10 genes)			
6	terpene	SDJM01000003, 907579:930486 (11 genes)			
7	other	SDJM01000004, 582204:627436 (23 genes)			
8	other	SDJM01000006, 282237:328303 (19 genes)			
9	other	SDJM01000006, 329430:373960 (19 genes)			
10	other	SDJM01000006, 783514:830534 (17 genes)			
11	terpene	SDJM01000007, 20929:44027 (11 genes)			
12	T1PKS	SDJM01000007, 199413: 248702 (25 genes)	**Trypacidin:**		
			Putative toxin biosynthesis regulatory protein AflJ (EAL89340.1)	43	KAF4553274
			Hypothetical protein (EAL89347.1)	72	KAF4553277
			Putative metallo-beta-lactamase domain protein (EAL89338.1)	57	KAF4553279
			Putative polyketide synthase (EAL89339.1)	59	KAF4553280
			Pestheic acid:		
			PtaD (AGO59044.1)	57	KAF4553277
			PtaB (AGO59041.1)	63	KAF4553279
			PtaA (AGO59040.1)	59	KAF4553280
13	NRPS	SDJM01000008, 153859:208507 (21 genes)			
14	other	SDJM01000009, 468080:512558 (18 genes)			
15	NRPS	SDJM01000015, 163571:217225 (15 genes)			
16	terpene	SDJM01000020, 268495:289017 (9 genes)			

(*Continued*)

Table 7. (Continued)

Cluster #	SM class	Genomic location (number of genes involved)	Similarity to known SM biosynthetic gene clusters		
			Known SM cluster gene (GenBank accession)	Similarity (%)	*E. fawcettii* GenBank accession
17	T1PKS	SDJM01000025, 66786:116804 (18 genes)			
18	T1PKS	SDJM01000028, 107087:155682 (17 genes)	**Cercosporin:**		
			Polyketide synthase (AAT69682.1)	53	KAF4548432
			Cercosporin toxin biosynthesis protein (ABC79591.2)	52	KAF4548433
			Oxidoreductase (ABK64184.1)	41	KAF4548434
			O-methyltransferase (ABK64180.1)	61	KAF4548436
			Oxidoreductase (ABK64182.1)	60	KAF4548439
19	T3PKS	SDJM01000034, 15555:58185 (18 genes)			
20	NRPS	SDJM01000035, 41518:95090 (22 genes)			
21	NRPS	SDJM01000037, 59764:106480 (19 genes)			
22	other	SDJM01000059, 16530:45727 (15 genes)			

in particular an overrepresentation of SM clusters among hemibiotrophs and necrotrophs has been seen before [160]. From the comparative analysis, it appears *E. fawcettii* has a lower variety of secondary metabolite clusters compared to the other necrotrophs and hemibiotrophs, particularly for Type I polyketide synthase (T1PKS) clusters. Blast analysis of the previously determined *E. fawcettii* elsinochrome cluster [27] against the *E. fawcetti* proteome indicated high similarities in amino acid sequence for six genes of the predicted T1PKS SM cluster 1 (S9 Table). Specifically, the predicted core biosynthetic gene of cluster 1 (accession KAF4556316) showed 98.6% similarity to the *E. fawcettii* polyketide synthase (*EfPKS1*) gene (accession ABU63483.1). An additional predicted biosynthetic gene (accession KAF4556314) had 99.6%
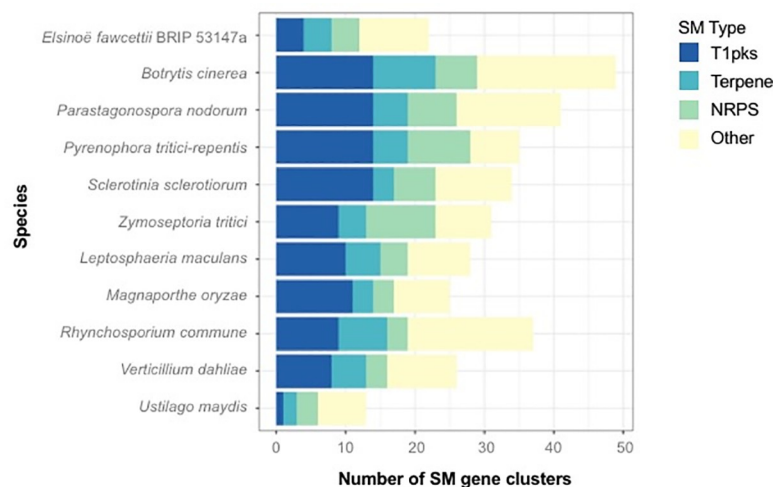

Fig 6. Comparison of numbers of predicted secondary metabolite gene clusters among 11 fungal species. Numbers of SM gene clusters, shown on the x axis, are divided into SM types; (I) Type I Polyketide synthase (T1PKS); (II) terpene; (III) non-ribosomal peptide synthetase (NRPS); and (IV) other, which contains all clusters identified by antiSMASH as either Type 3 Polyketide synthase (T3PKS), terpene-T1PKS, indole-T1PKS-NRPS, T1PKS-NRPS, indole-T1PKS, T1PKS-terpene-NRPS, indole, siderophore, lantipeptide, T3PKS-T1PKS or other.

similarity to the *E. fawcettii* ESC reductase (*RDT1*) gene (accession ABZ01830) and the predicted transport-related gene (accession KAF4556318) showed 70.3% similarity to the *E. fawcettii* ECT1 transporter (*ECT1*) gene (accession ABZ82008). Additional genes within the *E. fawcettii* SM cluster 1 obtained hits to the *E. fawcettii* elsinochrome cluster [27], specifically KAF4556317, KAF4556315 and KAF4556313 had high (97.4% - 100%) similarity to *PRF1* prefoldin protein subunit 3 (accession ABZ01833.1), *TSF1* transcription factor (accession ABZ01831.1) and *EfHP1* coding a hypothetical protein (accession ABZ82009.1). Hence, SM cluster 1 contains the two genes, *EfPKS1* and *TSF1*, which have been shown to be essential in elsinochrome production, as well as four genes (*RDT1*, *PRF1*, *ECT1* and *EfHP1*) also thought to be involved in elsinochrome biosynthesis [26, 27]. SM cluster 1 appears to lack four genes, being *OXR1*, *EfHP2*, *EfHP3* and *EfHP4*, which have all been reported to code for hypothetical proteins and not thought to be involved in biosynthesis [27]. However, to further investigate these omissions, BLAST analysis querying the nucleotide sequences of the elsinochrome cluster [27] against the contigs of the *E. fawcettii* genome assembly indicated regions with high similarities (99.3% - 99.7%) consistent with the location of predicted SM cluster 1 on contig SDJM01000001. The CDS of all four gene regions, however, were found to overlap with either each other or with other predicted genes. As no overlapping genes were predicted by GeneMark-ES on this isolate, it is thought the use of alternate gene model prediction programs between the studies may be a contributing factor for these differences. Further investigation through future transcriptomics analyses of *E. fawcettii* may provide resolution. Interestingly, SM cluster 1 consisted of an additional nine genes to the elsinochrome cluster previously described [27], all of which lay in a cluster adjacent to *ECT1*. Several of these additional genes obtained Pfam hits such as the THUMP domain, peptidase M3, Apolipoprotein O, Gar1/Naf1 RNA binding region and Endonuclease/Exonuclease/phosphatase family, suggesting these additional neighbouring proteins may perform functions such as RNA binding and modification, peptide cleavage, lipid binding and intracellular signalling, thus providing further genes for future investigation into the elsinochrome biosynthesis pathway.

An additional predicted SM cluster deserving of further investigation was SM cluster 2, a terpene-T1PKS, located 415,394 bp from the elsinochrome SM cluster 1 on contig SDJM01000001. This cluster shows sequence similarity to three proteins within the PR toxin biosynthetic gene cluster, namely aristolochene synthase (accession CDM31315.1) with 60% similarity to KAF4556513, short-chain dehydrogenase/reductase (accession CDM31317.1) with 54% similarity KAF4556505 and the type 2 FAD-binding protein (accession CDM31316.1) with 42% similarity to KAF4556518. The PR toxin is produced by the saprobe *Penicillium roqueforti*, a known contaminant of silages [161], while the mechanisms of its likely role in plant degeneration are unknown [162], PR toxin is seen to induce necrosis in human intestinal epithelial cells and monocytic immune cells [163] and exhibits mutagenic activity towards rats [164]. Thus, indicating the potential production of a toxin by *E. fawcettii* with DNA-binding capabilities. Another predicted SM gene cluster of interest was the T1PKS SM cluster 12. Three genes of cluster 12 (KAF4553277, KAF4553279 and KAF4553280) showed similarity to multiple known biosynthetic genes clusters; including the pestheic acid biosynthetic gene cluster of *Pestalotiopsis fici* [165] thought to function as a plant growth regulator [166] and the Trypacidin biosynthetic gene cluster of *Aspergillus fumigatus*, which produces a SM toxic to human lung cells [167]. Lastly, SM cluster 18 is predicted to code for five proteins with sequence similarity to those of the cercosporin biosynthetic gene cluster of *Cercospora nicotianae* [168]. Specifically, KAF4548432 (53% similarity to polyketide synthase, accession AAT69682.1), KAF4548433 (52% similarity to cercosporin toxin biosynthesis protein, accession ABC79591.2), KAF4548434 (41% similarity to oxidoreductase, accession ABK64184.1), KAF4548436 (61% similarity to O-methyltransferase, accession ABK64180.1) and

KAF4548439 (60% similarity to oxidoreductase, accession ABK64182.1). Cercosporin, similar to elsinochrome, is a fungal toxin which promotes the generation of reactive oxygen species in the presence of light, killing plant cells [169]. Cercosporin produced by *C. nicotianae* has been shown to cause necrotic lesions on tobacco leaves [170] and is also produced by the apple pathogen *Colletotrichum fioriniae* [171]. While it has been shown that elsinochrome production is important for full virulence by *E. fawcettii* [26, 27], biosynthesis of further SMs, such as cluster 2, 12 or 18, may be beneficial to pathogenesis by potentially disrupting host plant signalling, causing additional necrosis or inhibiting competing microbes.

Analysis of the distances between predicted SM genes and TE indicated no TE were in the close vicinity of SM cluster 1 (elsinochrome), the closest TE to the edge of the cluster was 199,748 bp or 77 genes away. This lack of association was seen among all *E. fawcettii* predicted SM clusters, with seven clusters predicted on contigs without identified TE (S9 Table). Of those clusters which did lie on contigs with TE, genes were an average distance of 236,556 bp away, suggesting recent activity of known TE was unlikely to be involved in the formation of *E. fawcettii* SM clusters. The closest AT-rich region to SM cluster 1 was a distance of 90,363 bp, while this was less than the mean distance (257,863 bp), this indication of potential TE degradation by RIP is still quite distant. In contrast to multiple SP and CE seen in the close vicinity of AT-rich regions, there were no genes from predicted SM clusters within 2 Kb of an AT-rich region, suggesting genes involved in SM production may benefit from residing in more stable genomic regions.

Conclusion

The WGS sequencing, genome mining and comparative analyses conducted in this study illustrates the potential that exists within the genome of *E. fawcettii* for virulence factors such as protein effectors and CWDE. The identification of these potential pathogenicity-related genes is a first step in determining further mechanisms utilised by *E. fawcettii* in addition to elsinochrome production, thus enabling this pathogen to defeat plant immune strategies in a host-specific manner. This study provides predicted virulence genes for future experimental investigation of *E. fawcettii* pathogenesis pathways, as well as establishing a comprehensive genomic resource for use in future studies to determine improved methods of control and screening of this pathogen.

Supporting information

S1 Table. GenBank accessions and genomic locations for RPB2, ITS and TEF1-α sequences included in the phylogenetic analyses with *elsinoë fawcettii* isolate (BRIP 53147a).
(DOCX)

S2 Table. Comparison of predicted gene classifications among *elsinoë fawcettii* and 10 other species; pfam hits, predicted CAZymes and core/unique/other genes.
(XLSX)

S3 Table. Sequence alignment of partial ITS and TEF1-α regions of *elsinoë fawcettii* (BRIP 53147a) in comparison with other isolates of *E. fawcettii* and closely related *elsinoë* species.
(TXT)

S4 Table. Comparison of results of EffectorP predicted candidate effectors and alternate candidate effector search among 11 species.
(XLSX)

S5 Table. Genomic and proteomic analyses of 15 species for use in known effector analysis and candidate effector prioritisation.
(XLSX)

S6 Table. Comparison of numbers of predicted secreted proteins, candidate effectors and prioritised candidate effectors among 15 species.
(XLSX)

S7 Table. Features and GenBank accessions of 276 *elsinoë fawcettii* BRIP 53147a candidate effectors.
(XLSX)

S8 Table. Features and GenBank accessions of 378 *elsinoë fawcettii* predicted CAZymes.
(XLSX)

S9 Table. Features and GenBank accessions of 404 *elsinoë fawcettii* genes with predicted involvement in secondary metabolite clusters.
(XLSX)

Acknowledgments

Author Contributions

Conceptualization: Sarah Jeffress, Gavin Ash.

Data curation: Sarah Jeffress, Niloofar Vaghefi.

Formal analysis: Sarah Jeffress, Kiruba Arun-Chinnappa, Ben Stodart, Niloofar Vaghefi.

Investigation: Ben Stodart.

Methodology: Sarah Jeffress, Niloofar Vaghefi, Gavin Ash.

Project administration: Gavin Ash.

Resources: Yu Pei Tan, Gavin Ash.

Supervision: Kiruba Arun-Chinnappa, Ben Stodart, Gavin Ash.

Validation: Sarah Jeffress.

Visualization: Niloofar Vaghefi.

Writing – original draft: Sarah Jeffress.

Writing – review & editing: Kiruba Arun-Chinnappa, Ben Stodart, Niloofar Vaghefi, Yu Pei Tan, Gavin Ash.

References

1. Xin H, Huang F, Zhang T, Xu J, Hyde KD, Li H. Pathotypes and genetic diversity of Chinese collections of *Elsinoë fawcettii* causing citrus scab. Journal of Integrative Agriculture. 2014; 13(6):1293–302.

2. Hyun J, Timmer L, Lee SC, Yun SH, Ko SW, Kim KS. Pathological characterization and molecular analysis of *Elsinoe* isolates causing scab diseases of citrus in Jeju Island in Korea. Plant Disease. 2001; 85(9):1013–7.

3. Hyun J, Yi S, MacKenzie S, Timmer L, Kim K, Kang S, et al. Pathotypes and genetic relationship of worldwide collections of *Elsinoë* spp. causing scab diseases of citrus. Phytopathology. 2009; 99 (6):721–8.

4. Tan M, Timmer L, Broadbent P, Priest M, Cain P. Differentiation by molecular analysis of *Elsinoë* spp. causing scab diseases of citrus and its epidemiological implications. Phytopathology. 1996; 86 (10):1039–44.

5. Bitancourt AA, Jenkins AE. *Elsinoe fawcettii*, the perfect stage of the Citrus scab fungus. Phytopathology. 1936; 26(4):393–5.

6. Bitancourt AA, Jenkins AE. Sweet orange fruit scab caused by *Elsinoë australis*. Journal of Agricultural Research. 1937; 54:0001–18.

7. Timmer L, Priest M, Broadbent P, Tan M. Morphological and pathological characterization of species of *Elsinoë* causing scab diseases of citrus. Phytopathology. 1996; 86(10):1032–8.

8. Whiteside J. Biological characteristics of *Elsinoe fawcettii* pertaining to the epidemiology of sour orange scab. Phytopathology. 1975; 65(10):1170–5.

9. Whiteside J. Pathogenicity of two biotypes of *Elsinoë fawcettii* to sweet orange and some other cultivars. Phytopathology. 1978; 68(1).

10. Wang LY, Liao HL, Bau HJ, Chung KR. Characterization of pathogenic variants of *Elsinoë fawcettii* of citrus implies the presence of new pathotypes and cryptic species in Florida. Canadian Journal of Plant Pathology. 2009; 31(1):28–37.

11. Queensland Government. Department of Agriculture and Fisheries. DAF Biological Collections [cited 2019 May 10]. Available from: https://collections.daf.qld.gov.au/web/home.html

12. dos Santos RF, Spósito MB, Ayres MR, Sosnowski MR. Phylogeny, morphology and pathogenicity of *Elsinoë ampelina*, the causal agent of grapevine anthracnose in Brazil and Australia. Journal of Phytopathology. 2018; 166(3):187–98.

13. Ash G, Stodart B, Hyun JW. Black scab of jojoba (*Simmondsia chinensis*) in Australia caused by a putative new pathotype of *Elsinoë australis*. Plant disease. 2012; 96(5):629–34.

14. Miles AK, Tan YP, Shivas RG, Drenth A. Novel pathotypes of *Elsinoë australis* associated with *Citrus australasica* and *Simmondsia chinensis* in Australia. Tropical Plant Pathology. 2015; 40(1):26–34.

15. Kokoa P. Review of sweet potato diseases in PNG. Food Security for Papua New Guinea Proceedings of the Papua New Guinea Food and Nutrition 2000 Conference ACIAR Proceedings; 2001.

16. Mau YS. Resistance response of fifteen sweet potato genotypes to scab disease (*Sphaceloma batatas*) in two growing sites in East Nusa Tenggara, Indonesia. Tropical Drylands. 2018; 2(1):5–11.

17. Scheper R, Wood P, Fisher B. Isolation, spore production and Koch's postulates of *Elsinoe pyri*. NZ Plant Prot. 2013; 66:308–16.

18. Fan X, Barreto R, Groenewald J, Bezerra J, Pereira O, Cheewangkoon R, et al. Phylogeny and taxonomy of the scab and spot anthracnose fungus *Elsinoë* (Myriangiales, Dothideomycetes). Studies in Mycology. 2017; 87:1–41.

19. Timmer LW, Garnsey SM, Graham JH. Compendium of Citrus Diseases. 2nd ed. The American Phytopathological Society. 2000.

20. Paudyal DP, Hyun JW. Physical changes in satsuma mandarin leaf after infection of *Elsinoë fawcettii* causing citrus scab disease. The Plant Pathology Journal. 2015; 31(4):421.

21. Agostini J, Bushong P, Bhatia A, Timmer L. Influence of environmental factors on severity of citrus scab and melanose. Plant Disease. 2003; 87(9):1102–6.

22. Chung KR. *Elsinoë fawcettii and Elsinoë australis*: the fungal pathogens causing citrus scab. Molecular Plant Pathology. 2011; 12(2):123–35.

23. Weiss U, Ziffer H, Batterham T, Blumer M, Hackeng W, Copier H, et al. Pigments of *Elsinoë* species: Pigment production by *Elsinoë* species; isolation of pure elsinochromes A, B, and C. Canadian Journal of Microbiology. 1965; 11(1):57–66.

24. Weiss U, Merlini L, Nasini G. Naturally occurring perylenequinones. Progress in the Chemistry of Organic Natural Products: Springer; 1987; 52:1–71.

25. Liao HL, Chung KR. Cellular toxicity of elsinochrome phytotoxins produced by the pathogenic fungus, *Elsinoë fawcettii* causing citrus scab. New Phytologist. 2008; 177(1):239–50.

26. Liao HL, Chung KR. Genetic dissection defines the roles of elsinochrome phytotoxin for fungal pathogenesis and conidiation of the citrus pathogen *Elsinoë fawcettii*. Molecular Plant-Microbe Interactions. 2008; 21(4):469–79.

27. Chung KR, Liao HL. Determination of a transcriptional regulator-like gene involved in biosynthesis of elsinochrome phytotoxin by the citrus scab fungus, *Elsinoë fawcettii*. Microbiology. 2008; 154 (11):3556–66.

28. Wang LY, Bau HJ, Chung KR. Accumulation of Elsinochrome phytotoxin does not correlate with fungal virulence among *Elsinoë fawcettii* isolates in Florida. Journal of Phytopathology. 2009; 157(10):602–8.

29. Hogenhout SA, Van Der Hoorn RAL, Terauchi R, Kamoun S. Emerging concepts in effector biology of plant-associated organisms. Molecular Plant-Microbe Interactions. 2009; 22(2):115–22. https://doi.org/10.1094/MPMI-22-2-0115 PMID: 19132864

30. Kamoun S. A catalogue of the effector secretome of plant pathogenic oomycetes. Annu Rev Phytopathol. 2006; 44:41–60.

31. Bolton MD, Thomma BPHJ, Nelson BD. *Sclerotinia sclerotiorum* (Lib.) de Bary: biology and molecular traits of a cosmopolitan pathogen. Oxford, UK: Blackwell Science Ltd; 2006. p. 1–16.

32. Lyu X, Shen C, Fu Y, Xie J, Jiang D, Li G, et al. A small secreted virulence-related protein is essential for the necrotrophic interactions of *Sclerotinia sclerotiorum* with its host plants. PLoS Pathogens. 2016; 12(2):e1005435. https://doi.org/10.1371/journal.ppat.1005435 PMID: 26828434

33. Yu Y, Xiao J, Zhu W, Yang Y, Mei J, Bi C, et al. Ss-Rhs1, a secretory Rhs repeat-containing protein, is required for the virulence of *Sclerotinia sclerotiorum*. Molecular Plant Pathology. 2017; 18(8):1052–61. https://doi.org/10.1111/mpp.12459 PMID: 27392818

34. Friesen TL, Faris JD, Solomon PS, Oliver RP. Host-specific toxins: effectors of necrotrophic pathogenicity. Oxford, UK: Blackwell Publishing Ltd; 2008. p. 1421–8.

35. Rodriguez-Moreno L, Ebert MK, Bolton MD, Thomma BPHJ. Tools of the crook- infection strategies of fungal plant pathogens. Plant Journal. 2018; 93(4):664–74. https://doi.org/10.1111/tpj.13810 PMID: 29277938

36. Wang X, Jiang N, Liu J, Liu W, Wang G-L. The role of effectors and host immunity in plant–necrotrophic fungal interactions. Taylor & Francis; 2014. p. 722–32.

37. Lo Presti L, Lanver D, Schweizer G, Tanaka S, Liang L, Tollot M, et al. Fungal Effectors and Plant Susceptibility. Annual Review of Plant Biology. 2015; 66(1):513–45. https://doi.org/10.1146/annurev-arplant-043014-114623 PMID: 25923844

38. Stergiopoulos I, de Wit PJGM. Fungal Effector Proteins. Annual Review of Phytopathology. 2009; 47 (1):233–63. https://doi.org/10.1146/annurev.phyto.112408.132637 PMID: 19400631

39. Manning VA, Pandelova I, Dhillon B, Wilhelm LJ, Goodwin SB, Berlin AM, et al. Comparative genomics of a plant-pathogenic fungus, *Pyrenophora tritici-repentis*, reveals transduplication and the impact of repeat elements on pathogenicity and population divergence. G3 (Bethesda, Md). 2013; 3 (1):41–63. https://doi.org/10.1534/g3.112.004044 PMID: 23316438

40. Sperschneider J, Dodds PN, Gardiner DM, Manners JM, Singh KB, Taylor JM. Advances and challenges in computational prediction of effectors from plant pathogenic fungi. PLoS Pathogens. 2015; 11 (5):e1004806. https://doi.org/10.1371/journal.ppat.1004806 PMID: 26020524

41. Martinez JP, Oesch NW, Ciuffetti LM. Characterization of the multiple-copy host-selective toxin gene, ToxB, in pathogenic and nonpathogenic isolates *of Pyrenaphora tritici-repentis*. Molecular Plant-Microbe Interactions. 2004; 17(5):467–74. https://doi.org/10.1094/MPMI.2004.17.5.467 PMID: 15141950

42. Friesen TL, Stukenbrock EH, Liu Z, Meinhardt S, Ling H, Faris JD, et al. Emergence of a new disease as a result of interspecific virulence gene transfer. Nature Genetics. 2006; 38(8):953. https://doi.org/10.1038/ng1839 PMID: 16832356

43. Syme RA, Hane JK, Friesen TL, Oliver RP. Resequencing and comparative genomics of *Stagonospora nodorum*: Sectional gene absence and effector discovery. G3: Genes, Genomes, Genetics. 2013; 3(6):959–69. https://doi.org/10.1534/g3.112.004994 PMID: 23589517

44. Chuma I, Isobe C, Hotta Y, Ibaragi K, Futamata N, Kusaba M, et al. Multiple translocation of the AVR-Pita effector gene among chromosomes of the rice blast fungus *Magnaporthe oryzae* and related species. PLoS Pathogens. 2011; 7(7). https://doi.org/10.1371/journal.ppat.1002147 PMID: 21829350

45. Ve T, Williams SJ, Catanzariti A-M, Rafiqi M, Rahman M, Ellis JG, et al. Structures of the flax-rust effector AvrM reveal insights into the molecular basis of plant-cell entry and effector-triggered immunity. Proceedings of the National Academy of Sciences of the United States. 2013; 110(43):17594. https://doi.org/10.1073/pnas.1307614110 PMID: 24101475

46. Kirsten S, Navarro-Quezada A, Penselin D, Wenzel C, Matern A, Leitner A, et al. Necrosis-inducing proteins of *Rhynchosporium commune*, effectors in quantitative disease resistance. Molecular Plant-Microbe Interactions. 2012; 25(10):1314–25. https://doi.org/10.1094/MPMI-03-12-0065-R PMID: 22712509

47. Rouxel T, Grandaubert J, Hane JK, Hoede C, Van De Wouw AP, Couloux A, et al. Effector diversification within compartments of the *Leptosphaeria maculans* genome affected by Repeat-Induced Point mutations. Nature Communications. 2011; 2(202):202. https://doi.org/10.1038/ncomms1189 PMID: 21326234

48. Djamei A, Schipper K, Rabe F, Ghosh A, Vincon V, Kahnt J, et al. Metabolic priming by a secreted fungal effector. Nature. 2011; 478(7369):395. https://doi.org/10.1038/nature10454 PMID: 21976020

49. Chen S, Songkumarn P, Venu RC, Gowda M, Bellizzi M, Hu J, et al. Identification and characterization of in planta-expressed secreted effector proteins from *Magnaporthe oryzae* that induce cell death in rice. Molecular Plant-Microbe Interactions. 2013; 26(2):191–202. https://doi.org/10.1094/MPMI-05-12-0117-R PMID: 23035914

50. Doehlemann G, van der Linde K, Assmann D, Schwammbach D, Hof A, Mohanty A, et al. Pep1, a secreted effector protein of *Ustilago maydis*, is required for successful invasion of plant cells. PLoS Pathogens. 2009; 5(2). https://doi.org/10.1371/journal.ppat.1000290 PMID: 19197359

51. Doehlemann G, Reissmann S, Aßmann D, Fleckenstein M, Kahmann R. Two linked genes encoding a secreted effector and a membrane protein are essential for *Ustilago maydis*-induced tumour formation. Molecular Microbiology. 2011; 81(3):751–66. https://doi.org/10.1111/j.1365-2958.2011.07728.x PMID: 21692877

52. Mueller AN, Ziemann S, Treitschke S, Assmann D, Doehlemann G. Compatibility in the *Ustilago maydis*-maize interaction requires inhibition of host cysteine proteases by the fungal effector pit2. PLoS Pathogens. 2013; 9(2). https://doi.org/10.1371/journal.ppat.1003177 PMID: 23459172

53. Tanaka S, Brefort T, Neidig N, Djamei A, Kahnt J, Vermerris W, et al. A secreted *Ustilago maydis* effector promotes virulence by targeting anthocyanin biosynthesis in maize. eLife. 2014; 2014(3). https://doi.org/10.7554/eLife.01355.001 PMID: 24473076

54. Fudal I, Ross S, Gout L, Blaise F, Kuhn ML, Eckert MR, et al. Heterochromatin-like regions as ecological niches for avirulence genes in the *Leptosphaeria maculans* genome: Map-based cloning of AvrLm6. Molecular Plant-Microbe Interactions. 2007; 20(4):459–70. https://doi.org/10.1094/MPMI-20-4-0459 PMID: 17427816

55. Parlange F, Daverdin G, Fudal I, Kuhn ML, Balesdent MH, Blaise F, et al. *Leptosphaeria maculans* avirulence gene AvrLm4-7 confers a dual recognition specificity by the Rlm4 and Rlm7 resistance genes of oilseed rape, and circumvents Rlm4 -mediated recognition through a single amino acid change. Molecular Microbiology. 2009; 71(4):851–63. https://doi.org/10.1111/j.1365-2958.2008.06547.x PMID: 19170874

56. Huang Y, Li Z, Evans N, Rouxel T, Fitt B, Balesdent M. Fitness Cost Associated with Loss of the AvrLm4 Avirulence Function in *Leptosphaeria maculans* (phoma stem canker of oilseed rape). European Journal of Plant Pathology. 2006; 114(1):77–89. https://doi.org/10.1007/s10658-005-2643-4

57. Saitoh H, Fujisawa S, Mitsuoka C, Ito A, Hirabuchi A, Ikeda K, et al. Large-scale gene disruption in *Magnaporthe oryzae* identifies MC69, a secreted protein required for infection by monocot and dicot fungal pathogens. PLoS Pathogens. 2012; 8(5):e1002711. https://doi.org/10.1371/journal.ppat.1002711 PMID: 22589729

58. Li W, Wang B, Wu J, Lu G, Hu Y, Zhang X, et al. The *Magnaporthe oryzae* avirulence gene AvrPiz-t encodes a predicted secreted protein that triggers the immunity in rice mediated by the blast resistance gene Piz-t. Molecular Plant-Microbe Interactions. 2009; 22(4):411–20. https://doi.org/10.1094/MPMI-22-4-0411 PMID: 19271956

59. Han L, Liu Z, Liu X, Qiu D. Purification, crystallization and preliminary X-ray diffraction analysis of the effector protein PevD1 from *Verticillium dahliae*. Acta Crystallographica Section F. 2012; 68(7):802–5. https://doi.org/10.1107/S1744309112020556 PMID: 22750869

60. Wang B, Yang X, Zeng H, Liu H, Zhou T, Tan B, et al. The purification and characterization of a novel hypersensitive-like response-inducing elicitor from *Verticillium dahliae* that induces resistance responses in tobacco. Applied Microbiology and Biotechnology. 2012; 93(1):191–201. https://doi.org/10.1007/s00253-011-3405-1 PMID: 21691787

61. Zhou R, Zhu T, Han L, Liu M, Xu M, Liu Y, et al. The asparagine-rich protein NRP interacts with the *Verticillium* effector PevD1 and regulates the subcellular localization of cryptochrome 2. Journal of Experimental Botany. 2017; 68(13):3427–40. https://doi.org/10.1093/jxb/erx192 PMID: 28633330

62. Schouten A, Van Baarlen P, Van Kan JAL. Phytotoxic Nep1-like proteins from the necrotrophic fungus *Botrytis cinerea* associate with membranes and the nucleus of plant cells. New Phytologist. 2008; 177 (2):493. https://doi.org/10.1111/j.1469-8137.2007.02274.x PMID: 18028294

63. Cuesta Arenas Y, Kalkman ERIC, Schouten A, Dieho M, Vredenbregt P, Uwumukiza B, et al. Functional analysis and mode of action of phytotoxic Nep1-like proteins of *Botrytis cinerea*. Physiological and Molecular Plant Pathology. 2010; 74(5):376–86. https://doi.org/10.1016/j.pmpp.2010.06.003

64. Liu ZH, Faris JD, Meinhardt SW, Ali S, Rasmussen JB, Friesen TL. Genetic and physical mapping of a gene conditioning sensitivity in wheat to a partially purified host-selective toxin produced by *Stagonospora nodorum*. Phytopathology. 2004; 94(10):1056–60. https://doi.org/10.1094/PHYTO.2004.94.10.1056 PMID: 18943793

65. Martinez JP, Ottum SA, Ali S, Francl LJ, Ciuffetti LM. Characterization of the ToxB gene from *Pyreno-phora tritici-repentis*. Molecular Plant-Microbe Interactions. 2001; 14(5):675–7. https://doi.org/10.1094/MPMI.2001.14.5.675 PMID: 11332732

66. Zhong Z, Marcel TC, Hartmann FE, Ma X, Plissonneau C, Zala M, et al. A small secreted protein in *Zymoseptoria tritici* is responsible for avirulence on wheat cultivars carrying the Stb6 resistance gene. New Phytologist. 2017; 214(2):619–31. https://doi.org/10.1111/nph.14434 PMID: 28164301

67. Raffaele S, Kamoun S. Genome evolution in filamentous plant pathogens: why bigger can be better. Nature Publishing Group; 2012. p. 417.

68. Kubicek CP, Starr TL, Glass NL. Plant cell wall degrading enzymes and their secretion in plant-patho-genic fungi. Annual Review of Phytopathology. 2014; 52(1):427–51. https://doi.org/10.1146/annurev-phyto-102313-045831 PMID: 25001456

69. King BC, Waxman KD, Nenni NV, Walker LP, Bergstrom GC, Gibson DM. Arsenal of plant cell wall degrading enzymes reflects host preference among plant pathogenic fungi. Biotechnology for Biofuels. 2011; 4:4.

70. Cantarel BI, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B. The Carbohydrate-Active EnZymes database (CAZy): An expert resource for glycogenomics. Nucleic Acids Research. 2009; 37(1):D233–D8. https://doi.org/10.1093/nar/gkn663 PMID: 18838391

71. Cosgrove DJ. Growth of the plant cell wall. Nature Publishing Group; 2005. p. 850.

72. Zuppini A, Navazio L, Sella L, Castiglioni C, Favaron F, Mariani P. An endopolygalacturonase from *Sclerotinia sclerotiorum* induces calcium-mediated signaling and programmed cell death in soybean cells. Molecular Plant-Microbe Interactions. 2005; 18(8):849–55. https://doi.org/10.1094/MPMI-18-0849 PMID: 16134897

73. Shieh MT, Brown RL, Whitehead MP, Cary JW, Cotty PJ, Cleveland TE, et al. Molecular genetic evi-dence for the involvement of a specific polygalacturonase, P2c, in the invasion and spread of *Aspergil-lus flavus* in cotton bolls. Applied and Environmental Microbiology. 1997; 63(9):3548.

74. Bravo Ruiz G, Di Pietro A, Roncero MIG. Combined action of the major secreted exo- and endopolyga-lacturonases is required for full virulence of *Fusarium oxysporum*. Molecular Plant Pathology. 2016; 17(3):339–53. https://doi.org/10.1111/mpp.12283 PMID: 26060046

75. Rogers LM, Kim YK, Guo W, Gonzalez-Candelas L, Li D, Kolattukudy PE. Requirement for either a host- or pectin-induced pectate lyase for infection of *Pisum sativum* by *Nectria hematococca*. Proceed-ings of the National Academy of Sciences of the United States. 2000; 97(17):9813. https://doi.org/10.1073/pnas.160271497 PMID: 10931947

76. López-Pérez M, Ballester AR, González-Candelas L. Identification and functional analysis of *Penicil-lium digitatum* genes putatively involved in virulence towards citrus fruit. Molecular Plant Pathology. 2015; 16(3):262–75. https://doi.org/10.1111/mpp.12179 PMID: 25099378

77. Yakoby N, Beno-Moualem D, Keen NT, Dinoor A, Pines O, Prusky D. *Colletotrichum gloeosporioides* pelB is an important virulence factor in avocado fruit-fungus interaction. Molecular Plant-Microbe Inter-actions. 2001; 14(8):988–95. https://doi.org/10.1094/MPMI.2001.14.8.988 PMID: 11497471

78. Valette-Collet O, Cimerman A, Reignault P, Levis C, Boccara M. Disruption of *Botrytis cinerea* pectin methylesterase gene Bcpme1 reduces virulence on several host plants. Molecular Plant-Microbe Interactions. 2003; 16(4):360–7. https://doi.org/10.1094/MPMI.2003.16.4.360 PMID: 12744465

79. Fu H, Feng J, Aboukhaddour R, Cao T, Hwang SF, Strelkov SE. An exo-1,3-[beta]-glucanase GLU1 contributes to the virulence of the wheat tan spot pathogen *Pyrenophora tritici-repentis*. Fungal Biol-ogy. 2013; 117(10):673.

80. Nguyen QB, Itoh K, Van Vu B, Tosa Y, Nakayashiki H. Simultaneous silencing of endo-β-1,4 xylanase genes reveals their roles in the virulence of *Magnaporthe oryzae*. Molecular Microbiology. 2011; 81(4):1008–19. https://doi.org/10.1111/j.1365-2958.2011.07746.x PMID: 21696466

81. Brito N, Espino JJ, González C. The endo-β-1,4-xylanase Xyn11A is required for virulence in *Botrytis cinerea*. Molecular Plant-Microbe Interactions. 2006; 19(1):25–32. https://doi.org/10.1094/MPMI-19-0025 PMID: 16404950

82. Noda J, Brito N, Gonzalez C. The *Botrytis cinerea* xylanase Xyn11A contributes to virulence with its necrotizing activity, not with its catalytic activity. BMC Plant Biology. 2010; 10:38.

83. Afgan E, Sloggett C, Goonasekera N, Makunin I, Benson D, Crowe M, et al. Genomics Virtual Labora-tory: A Practical Bioinformatics Workbench for the Cloud. PLoS ONE. 2015; 10(10):e0140829. https://doi.org/10.1371/journal.pone.0140829 PMID: 26501966

84. Andrews S. FastQC: a quality control tool for high throughput sequence data. 2010.

85. Bolger AM, Lohse M, Usadel B. Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinfor-matics. 2014; 30(15):2114–20. https://doi.org/10.1093/bioinformatics/btu170 PMID: 24695404

86. Zerbino DR, Birney E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Research. 2008; 18(5):821–9. https://doi.org/10.1101/gr.074492.107 PMID: 18349386

87. Gladman S. VelvetOptimiser. Victorian Bioinformatics Consortium, Clayton, Australia: http://bioinformaticsnetausoftwarevelvetoptimisershtml. 2012.

88. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. Journal of Computational Biology. 2012; 19(5):455–77. https://doi.org/10.1089/cmb.2012.0021 PMID: 22506599

89. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nature methods. 2012; 9 (4):357.

90. The Picard Toolkit. http://broadinstitute.github.io/picard/ [cited 2018].

91. Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Briefings in bioinformatics. 2013; 14(2):178–92.

92. Chikhi R, Medvedev P. Informed and automated k-mer size selection for genome assembly. Bioinformatics. 2014 Jan 1; 30(1):31–7.

93. Afgan E, Baker D, Batut B, Van Den Beek M, Bouvier D, Čech M, et al. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. Nucleic acids research. 2018 Jul 2; 46(W1):W537–44.

94. Vurture GW, Sedlazeck FJ, Nattestad M, Underwood CJ, Fang H, Gurtowski J, et al. GenomeScope: fast reference-free genome profiling from short reads. Bioinformatics. 2017 Jul 15; 33(14):2202–4.

95. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. 2015; 31 (19):3210–2. https://doi.org/10.1093/bioinformatics/btv351 PMID: 26059717

96. Kriventseva EV, Kuznetsov D, Tegenfeldt F, Manni M, Dias R, Simão FA, et al. OrthoDB v10: sampling the diversity of animal, plant, fungal, protist, bacterial and viral genomes for evolutionary and functional annotations of orthologs. Nucleic acids research. 2019 Jan 8; 47(D1):D807–11.

97. Testa AC, Oliver RP, Hane JK. OcculterCut: a comprehensive survey of AT-rich regions in fungal genomes. Genome biology and evolution. 2016; 8(6):2044–64.

98. Lee T, Peace C, Jung S, Zheng P, Main D, Cho I. GenSAS—An online integrated genome sequence annotation pipeline. 2011. p. 1967–73.

99. Lomsadze A, Ter-Hovhannisyan V, Chernoff YO, Borodovsky M. Gene identification in novel eukaryotic genomes by self-training algorithm. Nucleic Acids Research. 2005; 33(20):6494–506. https://doi.org/10.1093/nar/gki937 PMID: 16314312

100. Smit AFA, Hubley R, Green P. RepeatMasker http://repeatmasker.org.

101. Beier S, Thiel T, Münch T, Scholz U, Mascher M. MISA-web: a web server for microsatellite prediction. Bioinformatics (Oxford, England). 2017; 33(16):2583–5. https://doi.org/10.1093/bioinformatics/btx198 PMID: 28398459

102. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic Acids Research. 1997; 25 (17):3389–402. https://doi.org/10.1093/nar/25.17.3389 PMID: 9254694

103. The Uniprot Consortium. UniProt: the universal protein knowledgebase. Nucleic acids research. 2018; 46(5):2699.

104. Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M. Blast2GO: A universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics. 2005; 21 (18):3674–6. https://doi.org/10.1093/bioinformatics/bti610 PMID: 16081474

105. Johnson LS, Eddy SR, Portugaly E. Hidden Markov model speed heuristic and iterative HMM search procedure. BMC Bioinformatics. 2010; 11(1). https://doi.org/10.1186/1471-2105-11-431 PMID: 20718988

106. Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, et al. The Pfam protein families database: Towards a more sustainable future. Nucleic Acids Research. 2016; 44(1):D279–85. https://doi.org/10.1093/nar/gkv1344 PMID: 26673716

107. Quinlan AR. BEDTools: the Swiss-army tool for genome feature analysis. Current protocols in bioinformatics. 2014; 47(1):11.2. 1–.2. 34.

108. Grant CE, Bailey TL, Noble WS. FIMO: Scanning for occurrences of a given motif. Bioinformatics. 2011; 27(7):1017–8. https://doi.org/10.1093/bioinformatics/btr064 PMID: 21330290

109. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, et al. MEME Suite: Tools for motif discovery and searching. Nucleic Acids Research. 2009; 37(2):W202–W8. https://doi.org/10.1093/nar/gkp335 PMID: 19458158

110. Edgar RC. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research. 2004; 32(5):1792–7. https://doi.org/10.1093/nar/gkh340 PMID: 15034147

111. Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Molecular biology and evolution. 2016; 33(7):1870–4. https://doi.org/10.1093/molbev/msw054 PMID: 27004904

112. Nei M, Kumar S. Molecular evolution and phylogenetics. Kumar S, editor. Oxford;: Oxford University Press; 2000.

113. Maddison WP, Maddison DR. Mesquite: a modular sysem for evolutionary analysis. Version 3.6 http://www.mesquiteproject.org2018.

114. Kämper J, Kahmann R, Bölker M, Ma L-J, Brefort T, Saville BJ, et al. Insights from the genome of the biotrophic fungal plant pathogen *Ustilago maydis*. Nature. 2006; 444(7115):97.

115. Rouxel T, Grandaubert J, Hane JK, Hoede C, Van De Wouw AP, Couloux A, et al. Effector diversification within compartments of the *Leptosphaeria maculans* genome affected by Repeat-Induced Point mutations. Nature Communications. 2011; 2(202):202. https://doi.org/10.1038/ncomms1189 PMID: 21326234

116. Dean RA, Talbot NJ, Ebbole DJ, Farman ML, Mitchell TK, Orbach MJ, et al. The genome sequence of the rice blast fungus *Magnaporthe grisea*. Nature. 2005; 434(7036):980.

117. Penselin D, Münsterkötter M, Kirsten S, Felder M, Taudien S, Platzer M, et al. Comparative genomics to explore phylogenetic relationship, cryptic sexual potential and host specificity of *Rhynchosporium* species on grasses. BMC genomics. 2016; 17(1):953.

118. Klosterman S, Subbarao K, Kang S, Veronese P, Gold S, Thomma B, et al. Comparative genomics of the plant vascular wilt pathogens, *Verticillium dahliae* and *Verticillium alboatrum*. Phytopathology. 2010; 100(6):S64–S.

119. Van Kan JA, Stassen JH, Mosbach A, Van Der Lee TA, Faino L, Farmer AD, et al. A gapless genome sequence of the fungus *Botrytis cinerea*. Molecular plant pathology. 2017; 18(1):75–89.

120. Hane JK, Lowe RG, Solomon PS, Tan K-C, Schoch CL, Spatafora JW, et al. Dothideomycete–plant interactions illuminated by genome sequencing and EST analysis of the wheat pathogen *Stagonospora nodorum*. The Plant Cell. 2007; 19(11):3347–68.

121. Moolhuijzen P, See PT, Hane JK, Shi G, Liu Z, Oliver RP, et al. Comparative genomics of the wheat fungal pathogen *Pyrenophora tritici-repentis* reveals chromosomal variations and genome plasticity. BMC genomics. 2018; 19(1):279.

122. Amselem J, Cuomo CA, van Kan JAL, Viaud M, Benito EP, Couloux A, et al. Genomic analysis of the necrotrophic fungal pathogens *Sclerotinia sclerotiorum* and *Botrytis cinerea*. PLoS Genetics. 2011; 7 (8):e1002230. https://doi.org/10.1371/journal.pgen.1002230 PMID: 21876677

123. Plissonneau C, Hartmann FE, Croll D. Pangenome analyses of the wheat pathogen *Zymoseptoria tritici* reveal the structural basis of a highly plastic eukaryotic genome. BMC biology. 2018; 16(1):5.

124. Shanmugam G, Jeon J, Hyun JW. Draft genome sequences of *Elsinoë fawcettii* and *Elsinoë australis* causing scab diseases on citrus. Molecular Plant-Microbe Interactions. 2020 Feb 15; 33(2):135–7.

125. Li Z, Fan YC, Chang PP, Gao LL, Wang X. Genome sequence resource for *Elsinoë ampelina*, the causal organism of grapevine anthracnose. Molecular Plant-Microbe Interactions. 2020 Feb 4(ja).

126. Sperschneider J, Dodds PN, Gardiner DM, Singh KB, Taylor JM. Improved prediction of fungal effector proteins from secretomes with EffectorP 2.0. Molecular Plant Pathology. 2018; 19(9):2094–110. https://doi.org/10.1111/mpp.12682 PMID: 29569316

127. Petersen TN, Brunak S, Heijne GV, Nielsen H. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nature Methods. 2011; 8(10):785. https://doi.org/10.1038/nmeth.1701 PMID: 21959131

128. Käll L, Krogh A, Sonnhammer ELL. A Combined Transmembrane Topology and Signal Peptide Prediction Method. Journal of Molecular Biology. 2004; 338(5):1027–36. https://doi.org/10.1016/j.jmb.2004.03.016 PMID: 15111065

129. SoftBerry Inc. ProtComp v6 [cited 2018 July 19]. Available from: http://www.softberry.com/berry.phtml?topic=fdp.htm&no_menu=on.

130. Krogh A, Larsson B, Von Heijne G, Sonnhammer ELL. Predicting transmembrane protein topology with a hidden markov model: application to complete genomes. Journal of Molecular Biology. 2001; 305(3):567–80. https://doi.org/10.1006/jmbi.2000.4315 PMID: 11152613

131. Pierleoni A, Martelli P, Casadio R. PredGPI: A GPI-anchor predictor. BMC Bioinformatics. 2008; 9(1). https://doi.org/10.1186/1471-2105-9-392 PMID: 18811934

132. Sperschneider J, Gardiner DM, Dodds PN, Tini F, Covarelli L, Singh KB, et al. EffectorP: predicting fungal effector proteins from secretomes using machine learning. New Phytologist. 2016 Apr; 210 (2):743–61.

133. Blin K, Wolf T, Chevrette MG, Lu X, Schwalen CJ, Kautsar SA, et al. AntiSMASH 4.0—improvements in chemistry prediction and gene cluster boundary identification. Nucleic Acids Research. 2017; 45(1): W36–W41. https://doi.org/10.1093/nar/gkx319 PMID: 28460038

134. Fischer S, Brunk BP, Chen F, Gao X, Harb OS, Iodice JB, et al. Using OrthoMCL to assign proteins to OrthoMCL-DB groups or to cluster proteomes into new ortholog groups. Current protocols in bioinformatics. 2011; 35(1):6.12. 1–6. 9.

135. Steiner L, Findeiß S, Lechner M, Marz M, Stadler Peter F, Prohaska Sonja J. Proteinortho: Detection of (Co-)orthologs in large-scale analysis. BMC Bioinformatics. 2011; 12(1):124. https://doi.org/10.1186/1471-2105-12-124 PMID: 21526987

136. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria https://www.R-project.org/2018.

137. Zhang H, Yohe T, Huang L, Entwistle S, Wu P, Yang Z, et al. DbCAN2: A meta server for automated carbohydrate-active enzyme annotation. Nucleic Acids Research. 2018; 46(1):W95–W101. https://doi.org/10.1093/nar/gky418 PMID: 29771380

138. Yin Y, Mao X, Yang J, Chen X, Mao F, Xu Y. DbCAN: A web resource for automated carbohydrate-active enzyme annotation. Nucleic Acids Research. 2012; 40(1):W445–W51. https://doi.org/10.1093/nar/gks479 PMID: 22645317

139. Buchfink B, Xie C, Huson D, H. Fast and sensitive protein alignment using DIAMOND. Nature Methods. 2014; 12(1). https://doi.org/10.1038/nmeth.3176 PMID: 25402007

140. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Research. 2014; 42(1):D490–D5. https://doi.org/10.1093/nar/gkt1178 PMID: 24270786

141. Busk PK, Pilgaard B, Lezyk MJ, Meyer AS, Lange L. Homology to peptide pattern for annotation of carbohydrate-active enzymes and prediction of function.(Report). BMC Bioinformatics. 2017; 18(1). https://doi.org/10.1186/s12859-017-1625-9 PMID: 28403817

142. Urban M, Cuzick A, Rutherford K, Irvine A, Pedro H, Pant R, et al. PHI-base: A new interface and further additions for the multi-species pathogen-host interactions database. Nucleic Acids Research. 2017; 45(1):D604–D10. https://doi.org/10.1093/nar/gkw1089 PMID: 27915230

143. Kis-Papo T, Weig AR, Riley R, Peršoh D, Salamov A, Sun H, et al. Genomic adaptations of the halophilic Dead Sea filamentous fungus *Eurotium rubrum*. Nature Communications. 2014; 5(1). https://doi.org/10.1038/ncomms4745 PMID: 24811710

144. Gazis R, Kuo A, Riley R, Labutti K, Lipzen A, Lin J, et al. The genome *of Xylona heveae* provides a window into fungal endophytism. Fungal Biology. 2016; 120(1):26–42. https://doi.org/10.1016/j.funbio.2015.10.002 PMID: 26693682

145. Rosienski MD, Lee MK, Yu JH, Kaspar CW, Gibbons JG. Genome sequence of the extremely acidophilic fungus *Acidomyces richmondensis* FRIK2901. Microbiology Resource Announcements. 2018; 7 (16). https://doi.org/10.1128/MRA.01314-18 PMID: 30533739

146. Mohanta TK, Bae H. The diversity of fungal genome. BioMed Central Ltd.; 2015.

147. Bao W, Kojima K, Kohany O. Repbase Update, a database of repetitive elements in eukaryotic genomes. Mobile DNA. 2015; 6(1). https://doi.org/10.1186/s13100-015-0041-9 PMID: 26045719

148. Thomma BPHJ, Seidl MF, Shi-Kunne X, Cook DE, Bolton MD, van Kan JAL, et al. Mind the gap; seven reasons to close fragmented genome assemblies. Fungal Genetics and Biology. 2016; 90:24–30. https://doi.org/10.1016/j.fgb.2015.08.010 PMID: 26342853

149. Amselem J, Cuomo CA, van Kan JAL, Viaud M, Benito EP, Couloux A, et al. Genomic analysis of the necrotrophic fungal pathogens *Sclerotinia sclerotiorum* and *Botrytis cinerea*. PLoS Genetics. 2011; 7 (8):e1002230. https://doi.org/10.1371/journal.pgen.1002230 PMID: 21876677

150. Cambareri EB, Jensen BC, Schabtach E, Selker EU. Repeat-induced G-C to A-T mutations in *Neurospora*. (glycine-cysteine to alanine-threonine). Science. 1989; 244(4912):1571. https://doi.org/10.1126/science.2544994 PMID: 2544994

151. Selker EU, Cambareri EB, Jensen BC, Haack KR. Rearrangement of duplicated DNA in specialized cells of *Neurospora*. Cell. 1987; 51(5):741–52. https://doi.org/10.1016/0092-8674(87)90097-3

152. Selker EU. Premeiotic Instability of Repeated Sequences in *Neurospora crassa*. Annual Review of Genetics. 1990; 24(1):579–613. https://doi.org/10.1146/annurev.ge.24.120190.003051 PMID: 2150906

153. Gladyshev E. Repeat-Induced Point Mutation (RIP) and Other Genome Defense Mechanisms in Fungi. Microbiology spectrum. 2017; 5(4).

154. Braumann I, Berg M, Kempken F. Repeat induced point mutation in two asexual fungi, *Aspergillus niger* and *Penicillium chrysogenum*. Current Genetics. 2008; 53(5):287–97. https://doi.org/10.1007/s00294-008-0185-y PMID: 18347798

155. Gout L, Fudal I, Kuhn ML, Blaise F, Eckert M, Cattolico L, et al. Lost in the middle of nowhere: the AvrLm1 avirulence gene of the Dothideomycete *Leptosphaeria maculans*. Molecular Microbiology. 2006; 60(1):67–80. https://doi.org/10.1111/j.1365-2958.2006.05076.x PMID: 16556221

156. Fudal I, Ross S, Brun H, Besnard AL, Ermel M, Kuhn ML, et al. Repeat-Induced Point Mutation (RIP) as an alternative mechanism of evolution toward virulence in *Leptosphaeria maculans*. Molecular Plant-Microbe Interactions. 2009; 22(8):932–41. https://doi.org/10.1094/MPMI-22-8-0932 PMID: 19589069

157. Kloppholz S, Kuhn H, Requena N. A Secreted Fungal Effector of *Glomus intraradices* Promotes Symbiotic Biotrophy. Current Biology. 2011; 21(14):1204–9. https://doi.org/10.1016/j.cub.2011.06.044 PMID: 21757354

158. Mesarich CH, Bowen JK, Hamiaux C, Templeton MD. Repeat-containing protein effectors of plant-associated organisms. Frontiers in Plant Science. 2015; 6. https://doi.org/10.3389/fpls.2015.00872 PMID: 26557126

159. Liu T, Song T, Zhang X, Yuan H, Su L, Li W, et al. Unconventionally secreted effectors of two filamentous pathogens target plant salicylate biosynthesis. Nature Communications. 2014; 5(1). https://doi.org/10.1038/ncomms5686 PMID: 25156390

160. Zuccaro A, Lahrmann U, Langen G. Broad compatibility in fungal root symbioses. Current Opinion in Plant Biology. 2014; 20:135–45. https://doi.org/10.1016/j.pbi.2014.05.013 PMID: 24929298

161. Rasmussen R, Storm I, Rasmussen P, Smedsgaard J, Nielsen K. Multi-mycotoxin analysis of maize silage by LC-MS/MS. Analytical and Bioanalytical Chemistry. 2010; 397(2):765–76. https://doi.org/10.1007/s00216-010-3545-7 PMID: 20213172

162. Dubey MK, Aamir M, Kaushik MS, Khare S, Meena M, Singh S, et al. PR Toxin—Biosynthesis, Genetic Regulation, Toxicological Potential, Prevention and Control Measures: Overview and Challenges. Frontiers in Pharmacology. 2018; 9. https://doi.org/10.3389/fphar.2018.00288 PMID: 29651243

163. Hymery N, Puel O, Tadrist S, Canlet C, Le Scouarnec H, Coton E, et al. Effect of PR toxin on THP1 and Caco-2 cells: an in vitro study. World Mycotoxin Journal. 2017; 10(4):375–86.

164. Polonelli L, Lauriola L, Morace G. Preliminary studies on the carcinogenic effects of *Penicillium roqueforti* toxin (PR toxin) on rats. Mycopathologia. 1982; 78(2):125–7. https://doi.org/10.1007/BF00442636 PMID: 7099243

165. Xu X, Liu L, Zhang F, Wang W, Li J, Guo L, et al. Identification of the First Diphenyl Ether Gene Cluster for Pestheic Acid Biosynthesis in Plant Endophyte *Pestalotiopsis fici*. ChemBioChem. 2014; 15 (2):284–92. https://doi.org/10.1002/cbic.201300626 PMID: 24302702

166. Shimada A, Takahashi I, Kawano T, Kimura Y. Chloroisosulochrin, Chloroisosulochrin Dehydrate, and Pestheic Acid, Plant Growth Regulators, Produced by *Pestalotiopsis theae*. Zeitschrift fur Naturforschung—Section B Journal of Chemical Sciences. 2001; 56(8):797–803. https://doi.org/10.1515/znb-2001-0813

167. Gauthier T, Wang X, Dos Santos JS, Fysikopoulos A, Tadrist S, Canlet C, et al. Trypacidin, a spore-borne toxin from *Aspergillus fumigatus*, is cytotoxic to lung cells. PLoS ONE. 2012; 7(2):e29906. https://doi.org/10.1371/journal.pone.0029906 PMID: 22319557

168. Chen H, Lee MH, Daub ME, Chung KR. Molecular analysis of the cercosporin biosynthetic gene cluster in *Cercospora nicotianae*. Molecular Microbiology. 2007; 64(3):755–70. https://doi.org/10.1111/j.1365-2958.2007.05689.x PMID: 17462021

169. Daub ME, Hangarter RP. Light-induced production of singlet oxygen and superoxide by the fungal toxin, cercosporin. Plant Physiology. 1983; 73(3):855–7.

170. Dekkers KL, You B-J, Gowda VS, Liao H-L, Lee M-H, Bau H-J, et al. The *Cercospora nicotianae* gene encoding dual O-methyltransferase and FAD-dependent monooxygenase domains mediates cercosporin toxin biosynthesis. Fungal Genetics and Biology. 2007; 44(5):444–54. https://doi.org/10.1016/j.fgb.2006.08.005 PMID: 17074519

171. de Jonge R, Ebert MK, Huitt-Roehl CR, Pal P, Suttle JC, Spanner RE, et al. Gene cluster conservation provides insight into cercosporin biosynthesis and extends production to the genus *Colletotrichum*. Proceedings of the National Academy of Sciences of the United States. 2018; 115(24):E5459. https://doi.org/10.1073/pnas.1712798115 PMID: 29844193