Presence of elsinochrome and other putative effectors in select genomes of the plant pathogen *Elsinoë* spp. based on *in silico* analysis

Rhenz Cedrick M. Cequeña and Leilani G. Sumabat-Dacones*

Institute of Biology, College of Science, University of the Philippines, Diliman, Quezon City, 1101 Philippines

ABSTRACT

The genus *Elsinoë* comprises obligate fungal pathogens known to cause scab disease in various crops, including mango, peanut, and sweet potato. In other fungal plant pathogens, secreted proteins or effectors and other secondary metabolites have been characterized as being involved mainly in stimulating host responses, such as the eventual manifestation of distinct symptoms on specific crops or as indications of host specialization. In line with this, variability in the genes conferring for the corresponding effector proteins and other secondary metabolites has also been identified between species of the same fungal genus and has helped in elucidating phylogenetic relationships and in determining host preferences. Available draft genomes of six plant pathogenic *Elsinoë* spp. retrieved from the NCBI were utilized in this study to assess their variations in terms of putative effectors and other secondary metabolites, along with *in silico* descriptions of their potential biological function and evolutionary relationships. Genome prediction revealed, on average, 3000 cytoplasmic effectors and approximately 280 apoplastic effectors across the species. Pathogenicity and virulence effectors such as melanin and elsinochrome were specifically detected in all the *Elsinoë* spp. genomes, with considerable variation between species. Ontology analysis revealed putative genes related to toxin activity, detoxification, and carbohydrate binding in *Elsinoë*. Furthermore, phylogenetic analysis of the internal transcribed spacer (ITS) and elsinochrome biosynthetic genes revealed clustering of *Elsinoë* spp. based on the host. These findings suggest that the disease status of different plant hosts infected by different *Elsinoë* spp. varies.

*Corresponding author
Email Address: lsdacones@up.edu.ph
Date received: November 16, 2023
Date revised: January 6, 2024
Date accepted: January 14, 2024

INTRODUCTION

Plant pathogenic fungi possess genes that confer determinants of virulence and pathogenicity in their corresponding hosts. These genetic factors have been characterized as crucial for successfully defending against the host immune response to take over and completely facilitate host preferential invasion (van der Does and Rep 2007).

Elsinoë, a causative agent of plant diseases manifesting as scabs on infected substrates, is a specialized obligate pathogenic fungus in various hosts, including economically important crops (Fan et al. 2017). Few *Elsinoë* spp. have been reported to be present in the Philippines, such as *E. mangiferae,* which causes mango scab; *E. perseae*, which causes avocado scab; and *E. batatas*, which causes sweet *potato scab* (CropLife Philippines 2018). In mangoes, symptoms are characterized by the presence of dark grayish and irregular spots on fruits and leaves.

KEYWORDS

Elsinochrome, fungal melanin, *Elsinoë* spp., whole-genome analyses, mango scab disease, pathogenicity and virulence genes

Moreover, newly infected leaves might defoliate, and infected fruits are misshapen and often unattractive to consumers, which eventually leads to decreased market value and elimination of export potential (CropLife Philippines 2018). Similarly, avocado and sweet potato scabs pose serious problems in the country (Marais 2004). In fact, the leaf and stem scabbing of sweet potato plants represents the most severe disease among farms in Southeast Asia and the Pacific region, of which up to 50% of the losses are detected in the Philippines alone (CropLife Philippines 2018).

The presence of scabs in problematic produce may be attributed to secondary metabolites secreted by *Elsinoë* spp. In other fungi, cercosporin, for instance, is a photoactive and highly toxic secondary metabolite that aids in disease development in the fungal genus *Cercospora* (de Jonge et al. 2018); its mechanism involves the breakdown of cellular compartments by peroxidation of lipid membranes, leading to phenotypic responses such as necrosis. Furthermore, cercosporin is produced by a cluster of eight polyketide synthase genes (CTB1–CTB8). All of these genes were shown to be required for cercosporin biosynthesis and symptom development (Świderska-Burek et al. 2020). Secondary metabolites in fungi can also facilitate host preferences or specialization, such as those of *Corynespora cassiicola* in the case of cassicolin (Dacones et al. 2022; Lopez et al., 2018). Cassiicolin variants have been linked to host specificity in different lineages and populations of *C. cassiicola*. To date, only two species, namely, *E. facettii* and *E. ampelina,* have been identified to possess the non-host selective toxin elsinochrome, which is a unique secondary metabolite produced by the genus. Previous findings suggest that elsinochrome plays an important role in the necrosis of plant tissues during infection (Jeffress et al. 2020). This secondary metabolite is structurally related to perlylenequinones due to its pentacyclic 4,9-dihydroxy-3,10-perylenequinone core and exhibits toxicity via the generation of reactive oxygen species (ROS) (Hu et al. 2019; Jiao et al. 2019).

Secondary metabolite prediction and identification of their respective gene clusters have been used in *in silico* detection of putative genetic factors relevant to fungal biology and disease establishment. In *Zymoseptoria tritici*, gene clusters of secondary metabolites were initially detected *in silico*, which contributed to the understanding of disease progression (Cairns and Meyer 2017). With the recent availability of whole-genome resources in open-access sequence databases, exploration of *Elsinoë* genomes is instrumental in identifying putative elements and secondary metabolites related to pathogenicity and virulence. Draft genomes of six plant pathogenic *Elsinoë* spp. retrieved from the NCBI were utilized in this study to assess their variations in terms of putative effectors and other secondary metabolites, along with *in silico* descriptions of their potential biological function and evolutionary relationships.

MATERIALS AND METHODS

Retrieval of whole-genome sequences

Sequences of representative plant pathogenic *Elsinoë* spp. were downloaded from the NCBI GenBank (https://www.ncbi.nlm.nih.gov/genbank/) for the last 2021 (Table 1.). The following information was used to determine the background of each of the genomes: (a) origin of the isolates, (b) sequencing technology, (c) genome assembly statistics, and (d) year of isolation (Table 2).

Table 1: List of *Elsinoë* spp. used in this study obtained from NCBI GenBank.

Species and Isolate	Information	Accession Number	Year Collected	Reference
E. ampelina CECT 20119	Used in Phylogenetics	JAAEIW010000001-JAAEIW010000548	2020*	Haridas et al., (2020)
E. ampelina YL-1	Used in transcriptome study	SMYM01000001-SMYM01000013	2015	Li et al., (2021)
E. arachidis LNFT-H01	Used in elucidating metabolic pathway of elsinochrome	JAAPAX010000001-JAAPAX010000016	2020*	Jiao et al., (2021)
E. australis NL1	Draft whole genome	NHZQ01000001-NHZQ01000452	2015	Shanmugam et al., (2019)
E. australis Forbes_1	Draft whole genome	PTQO01000001-PTQO01000637	2005	-
E. batatas CRI-CJ2	Genome resource	JAESVG020000001-JAESVG020000013	2018	Zhang et al., (2022)
E. fawcettii 53147a	Used in genome mining for candidate effectors	SDJM01000001-SDJM01000286	2009	Jeffress et al., (2020)
E. fawcettii DAR-70024**	Used in genome mining for candidate effectors	SWCR01000001-SWCR01000053	1995	Jeffress et al., (2020)
E. murrayae CQ-2017a**	Draft Whole Genome	NKHZ01000001-NKHZ01000098	2015	-

*The collection date indicated inNCBI is unknown and only the submission date is available
**Not included in ITS phylogenetic analysis

Table 2: Genomic features of *Elsinoë* spp. based on NCBI (L50 and Assembly Method) and GenSAS.

Isolate	*E. ampelina* CECT 20119	*E. ampelina* YL-1	*E. arachidis* LNFT-H01	*E. australis NL1**	*E. australis* Forbes_1	*E. batatas* CRI-CJ2	*E. fawcettii* 53147a	*E. fawcettii* DAR-70024	*E. murrayae* CQ-2017a*
Geographic Location	-	China	China	China	Australia	China	Australia	Australia	China
Host of Origin	-	*Vitis vinifera* cv. Red Globe	-	*Populus tomentosa*	*Simmondsia chinensis*	*Ipomoea batatas*	*Citrus limon*	*Citrus* spp.	*Salix babylonica*
Sequencing technology	Illumina Hiseq	PacBio; Illumina HiSeq	PacBio RSII	Illumina	Illumina HiSeq	Oxford Nanopore	Illumina Miseq	Illumina	Illumina
Assembly Length (Mbp)	27.93	28.30	33.18	23.20	26.43	26.44	26.01	26.33	20.72

-Indicated as missing data from NCBI
*Draft whole genome assembly

Gene Prediction and Gene Ontology Analysis
Whole-genome sequences retrieved were uploaded to GenSAS 6.0 for annotation using standard genetic code (Humann et al. 2019). To address repeats in each of the sequences, the sequences were first masked using RepeatModeler or RepeatMasker (Flynn et al. 2020) using the following parameters: slow sensitivity and fungus as the DNA source, with all the remaining settings retaining the default. Transcript and protein alignments of the translated nucleotide sequences were subsequently conducted via tBLASTN v 2.12.0 (Altschul et al. 1990) and Diamond v2.0.6 (Buchfink et al. 2021), respectively, using NCBI RefSeq fungi as a reference database and all the remaining settings were maintained by default. Likewise, gene prediction was performed using GeneMark-ES (Besemer et al. 2001) using default settings and Glimmer M (Majoros et al. 2004) using the *Aspergillus* genome as the training dataset. A consensus gene was created using EVidenceModeler v1.1.1 using the predicted and aligned results from Diamond, GeneMark-ES, and Glimmer M with the default weight setting of 1 and tBLASTN with the default weight setting of 10 (Haas et al. 2008), followed by BUSCO v5.2.2 (Simão et al. 2015) for selection of the desired gene set. The chosen gene sets were analyzed via Companion (https://companion.gla.ac.uk/jobs/new) to determine the number of pseudogenes present in the dataset. InterProScan v 5.53-87.0 (Jones et al. 2014), Pfam (Finn et al. 2016), and Diamond were used for the prediction of protein families. Protein sequences from the desired annotated gene set were subsequently downloaded as fasta files and analyzed for gene ontology using the OMA browser (https://omabrowser.org/oma/functions/) to generate a gaf file, which was subsequently plotted using WEGO 2.0 (Ye et al. 2018), providing a general overview of the biological, molecular, and cellular components related to fungal pathogen virulence.

Prediction of Effector Proteins and Secondary Metabolites
The prediction of secreted effectors was performed on the *Elsinoë* spp. sequences using SignalP or Effector P (Almagro Armenteros et al. 2019; Sperschneider et al. 2016) and subsequently analyzed on the TMHMM server (Krogh et al., 2001). AntiSMASH v6.0, which is a web-based software program that detects existing and putative secondary metabolites, was also utilized (Blin et al. 2021). The parameters were set with strict detection and with the following tools: known cluster blast, MIBiG cluster comparison, ClusterBlast, Active Site Finder, Pfam-based GO term annotation, Cluster Pfam analysis, SubClusterBlast, REFinder, TIGRFam analysis, and CASSIS.

Determining Evolutionary Relationships among *Elsinoë* spp. Isolates from across Hosts and Geographic Locations
Conserved regions such as the ITS region, β-tubulin region, and elsinochrome biosynthetic genes were obtained via BLASTN against the representative genomes of *Elsinoë*. These additional sequences have GenBank accession nos. WLZB01, PTQQ01, VAAB01, WLYY01, and WLYZ01 (Supplementary Table 1). The tree based on the ITS region contains the outgroup *Myriangium duriaei* and was obtained from the NCBI GenBank under accession number MH855793 (*Myriangium duriaei* CBS 260.36); the outgroup for the β-tubulin tree was obtained via BLASTN via the whole genome of the same sample, JAAEIR01, of the same *M. duriaei* isolate with the ITS *region*. The sequences were aligned using clustalW in MEGA X (Kumar et al. 2018). Phylogenetic congruence was also assessed by means of an incongruence length test using PAUP (Wilgenbusch and Swofford 2003) with the following settings: maxtrees=1000, increase=auto, number of replications=1000. The aligned sequences were subjected to an optimal nucleotide model of substitution, and phylogenetic construction was performed by means of maximum likelihood using MEGA X (Kumar et al. 2018). Bayesian inference was performed using the BEAST package v2.7.1 (Suchard et al. 2018), which includes BEAUTi, BEAST, and Tree Annotator. The alignments were imported into BEAUTi to set the nucleotide model of substitution used, set the chain length to 10 million generations and generate the BEAST file. A Bayesian phylogenetic tree was constructed with BEAST software, and a summary tree was obtained with Tree Annotator with a burn-in percentage of 10%. Finally, the tree was graphically constructed using Fig Tree v1.4.4. (http://tree.bio.ed.ac.uk/software/figtree/). The number of mutations and nucleotide diversity of the elsinochrome genes were calculated using DnaSP with the default settings (Librado and Rozas 2009).

RESULTS

Genome statistics of the draft assemblies of *Elsinoë* spp.
The genome sizes of the different *Elsinoë* spp., based on the whole-genome assemblies obtained from NCBI, ranged from 20 Mb to 33 Mb (Table 2). Moreover, there were approximately 9,000 to 10,000 predicted genes based on GeneMark ES. Genome completeness assessed using BUSCO resulted in approximately 97-98% completeness (Table 3).

Table 3: Assembly method, GC content, and predicted genes among representative *Elsinoë* sequences.

Isolate	*E. ampelina* CECT 20119	*E. ampelina* YL-1	*E. arachidis* LNFT-H01	*E. australis NL1*	*E. australis* Forbes_1	*E. batatas* CRI-CJ2	*E. fawcettii* 53147a	*E. fawcettii* DAR-70024	*E. murrayae* CQ-2017a*
Assembly Method*	AllPaths LG	Soap	Canu	Soap	Velvet	FALCON	SPAdes	SPAdes	Soap
GC Content	49.66%	49.52 %	48.24%	52.80%	50.98%	50.80%	52.30%	52.29 %	54.50%
Predicted Gene									
GeneMark ES	9,709	9,731	10,402	9,280	9,419	9,464	10,434	10,464	8,045
Glimmer	7,702	7,843	9,058	7,455	7,798	7,644	7,859	7,844	6,289
Evidence Modeler	8,113	8,230	8,732	8,094	8,365	7,947	8,885	8,803	6,719
No. of Pseudogenes	243	241	261	242	255	258	279	281	185
Gene Completeness									
GeneMark ES	97.3%	98.2%	98.3%	98.2%	96.4%	98.0%	98.3%	98.2%	97.4%
Glimmer	52.7%	77.7%	51.2%	60.8%	64.1%	58.0%	57.4%	57.1%	53.4%
Evidence Modeler	76.6%	54.2%	76.3%	83.8%	81.4%	80.5%	83%	83.0%	76.5%

*Based on details indicated in the NCBI sequences.

Prediction of Effector Proteins in *Elsinoë* spp.

Figure 1 shows approximately 3,000 cytoplasmic effectors and approximately 280 apoplastic effectors across the genomes analyzed, indicating a high frequency of cytoplasmic effectors in comparison with apoplastic effectors. Among the different *Elsinoë* spp., the *E. fawcetti* strain 53147a contained 1,101 cytoplasmic effectors according to the P signal and 3,884 from the Effector P signal.

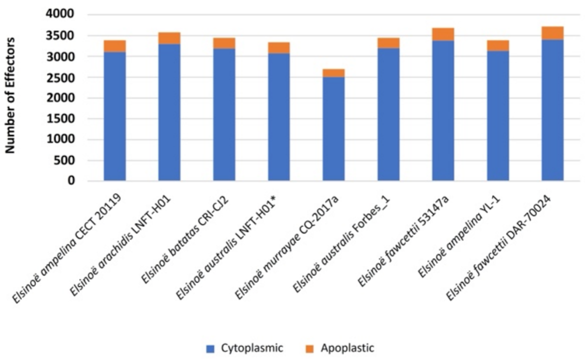

Figure 1: Number of cytoplasmic and apoplastic effectors present across the *Elsinoë* spp. genomes used in the study

Detection of Secondary Metabolites in *Elsinoë* spp.

Approximately 10-15 total secondary metabolite gene clusters were found among the representative genomes (Figure 2). Eight of them are known secondary metabolites from *Elsinoë* and other fungal species, with varying percentage similarity based on the reference gene database (Table 4). Elsinochrome C, a previously identified non-host selective phytotoxin, was detected in all of the representative *Elsinoë* spp.; however, this phytotoxin had low similarity with the reference compound (Figure 3 and Table 4). Moreover, melanin was detected in all the representative genomes with 100% similarity. Other secondary metabolites with known functions relevant to pathogenicity were detected in other plant pathogenic fungi, e.g., verticillin in *E. ampelina* representatives and abscisic acid in the *E. arachidis* strains LNFT-H01 and the *E. australis* strain NL1. Other secondary metabolites without known potential roles in fungal pathogenicity were detected, such as duclauxin in *E. ampelina, E. arcahidis,* and *E. batatas*; trypacidin in the *E. fawcettii* strain 53147a; clavaric acid in *E. amlelina* and *E. australis*; and Res-1214-2 in the *E. fawcettii* isolate DAR-70024.

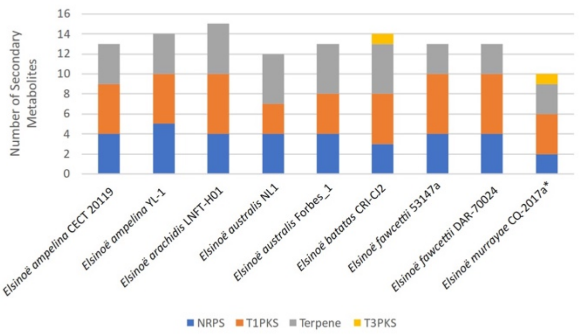

Figure 2: Total secondary metabolite gene clusters predicted in *Elsinoë* spp. using AntiSMASH

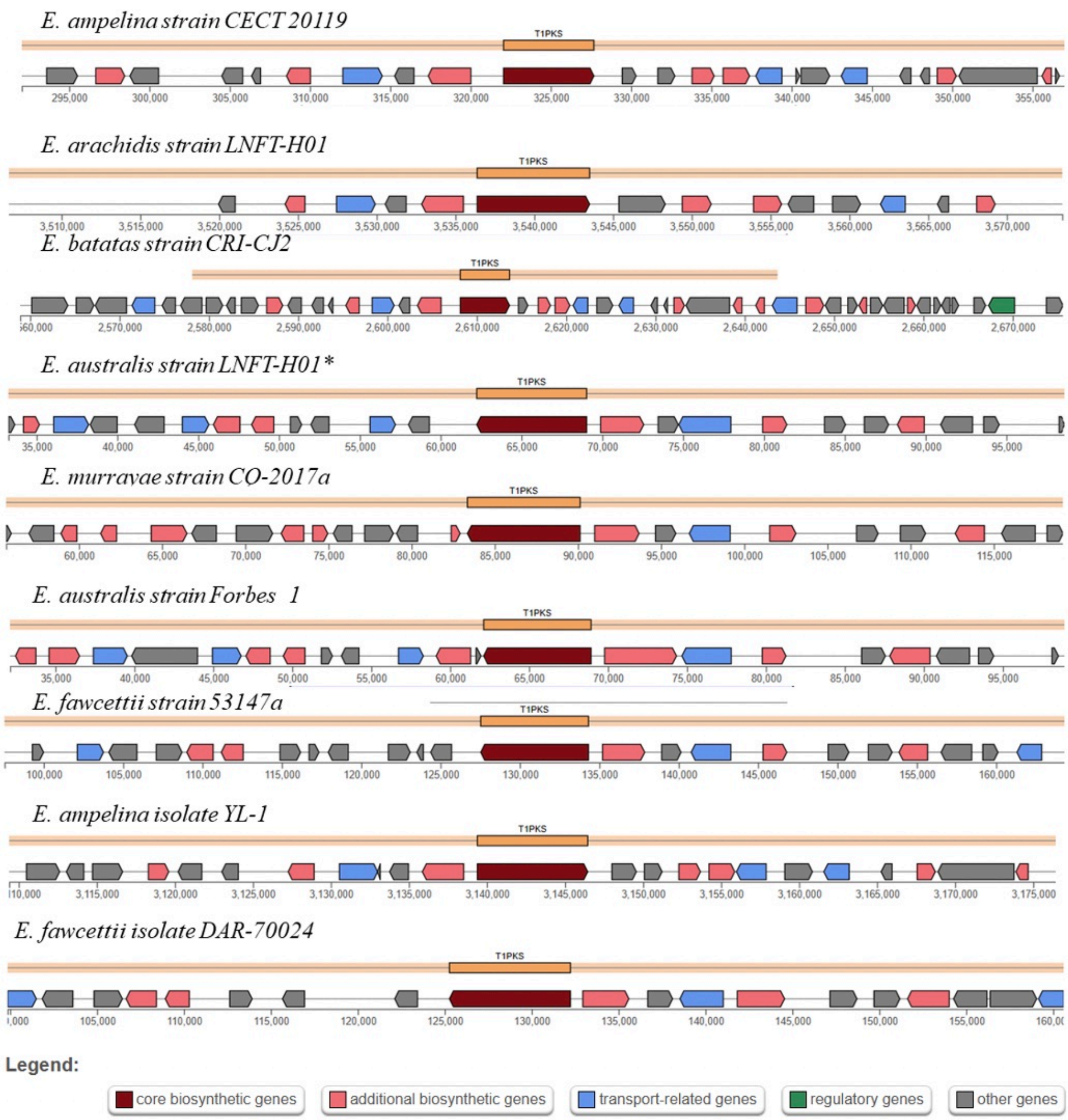

Figure 3: Genetic clusters of elsinochrome in representative *Elsinoë* spp. generated using AntiSMASH

Table 4: Known secondary metabolites detected in *Elsinoë* spp. with percent similarity based on Minimum Information about a Biosynthetic Gene cluster (MiBIG) database

Isolate	*E. ampelina* CECT 20119	*E. ampelina* YL-1	*E. arachidis* LNFT-H01	*E. australis* NL1	*E. australis* Forbes_1	*E. batatas* CRI-CJ2	*E. fawcettii* 53147a	*E. fawcettii* DAR-70024	*E. murrayae* CQ-2017a*
Clavaric Acid	100%	100%	-	100%	100%	-	-	-	100%
Melanin	100%	100%	100%	100%	100%	100%	100%	100%	100%
Duclauxin	28%	28%	28%			28%			
Elsinochrome C	25%	31%	25%	31%	31%	37%	31%	31%	31%
Verticilin	15%	15%	-	-	-	-	-	-	-
Abscisic Acid	-	-	50%	-	50%	-	-	-	-
Trypacidin	-	-	-	-	-	-	21%	-	-
Res-1214-2	-	-	-	-	-	-	-	18%	-

-Not detected with the data ran in AntiSMASH

Comparison of Gene Ontology

Gene Ontology analyses revealed approximately 7,000 to 9,000 genes as annotated by the OMA browser (Figure 4). The yielded GO terms were categorized into three broad categories: (i) 6,000-7,000 genes for biological processes; (ii) 5000-6000 for cellular components; and (iii) 6,000-7,000 for molecular function (Figure 5). Most of the gene functions identified were related to cellular components, catalytic activity, and the response to stimuli (Figure 6).

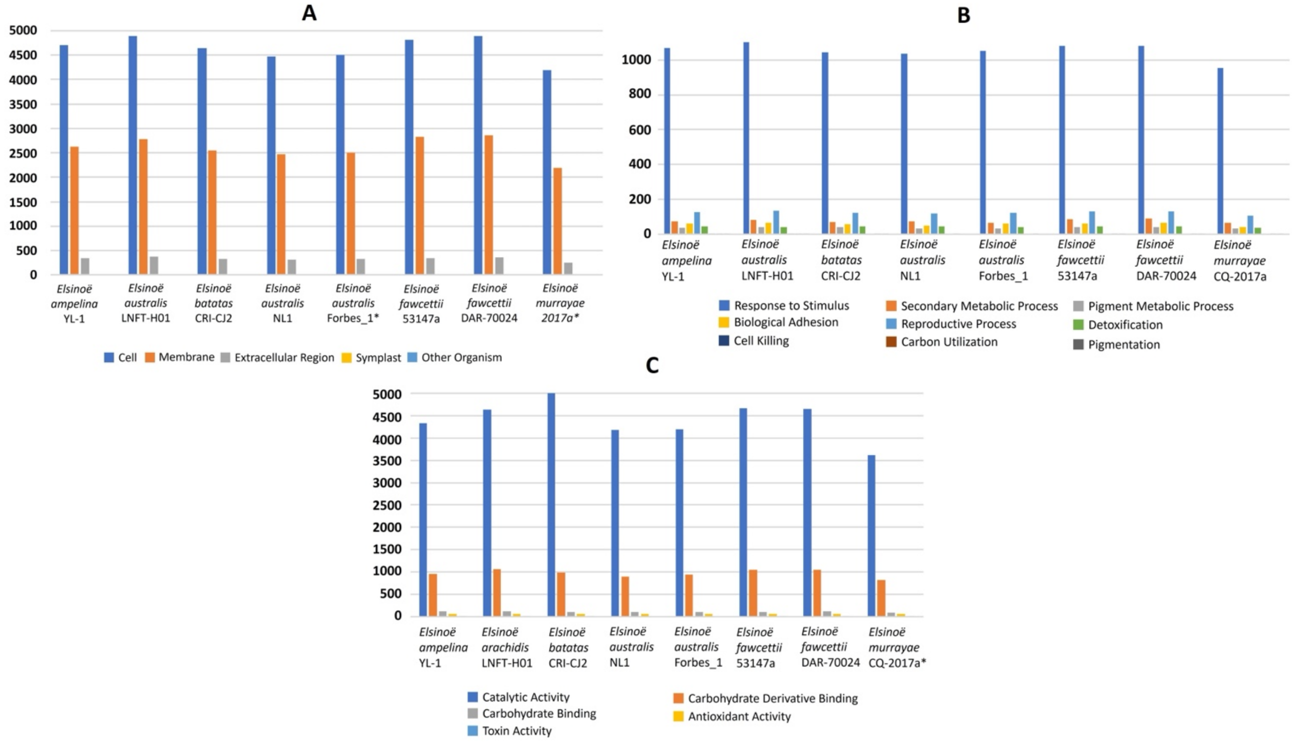

Figure 4: Distribution of gene function of *Elsinoë* spp.in terms of (A) cellular component, (B) molecular function, and (C) biological process

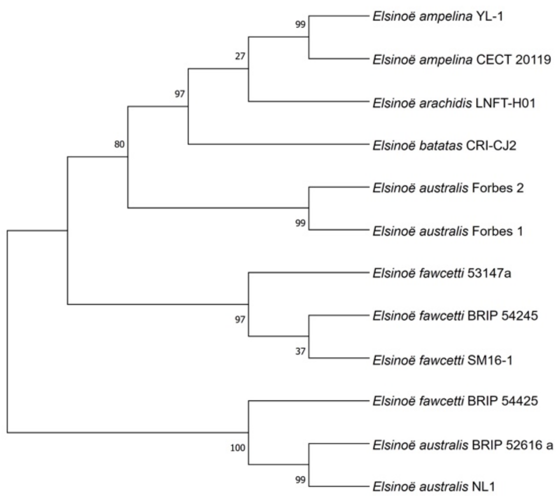

Figure 5: Maximum Likelihood Tree of *Elsinoë* spp. using 6196 nucleotides long core elsinochrome gene following Tamura-Nei + G of nucleotide model of substitution with 6 discrete gamma categories.

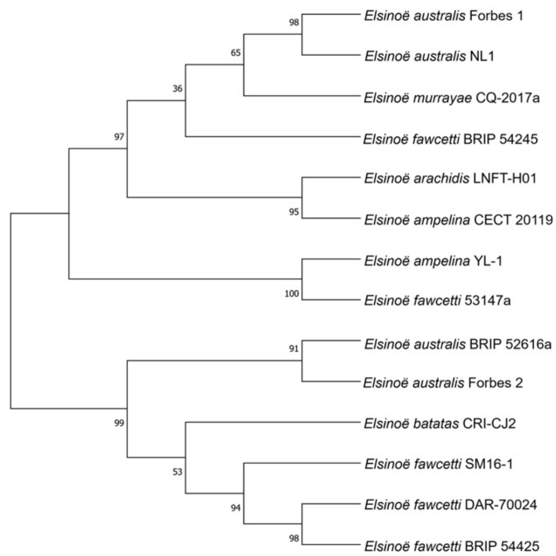

Figure 6: Maximum Likelihood tree of 542 nucleotide long ITS gene following Tamura-3 parameter + G of nucleotide model of substitution with 6 discrete gamma categories.

Phylogenetic Relationships among *Elsinoë* spp.

To determine the relationships among the plant pathogenic *Elsinoë* spp., *ITS, β-tubulin*, and a precursor gene for elsinochrome biosynthesis were used. With the initial test for incongruence of 0.001, the sets of genes were not combined, and phylogenetic analyses were performed separately (Cunningham, 1997). Similarly, the topologies of the trees generated based on the *ITS region, β-tubulin region*, and core elsinochrome sequence varied (Figures 5, 6, and 7). However, representative isolates clustered according to their source, particularly with the corresponding host from which they were isolated, regardless of the gene. For instance, all the citrus isolates, along with a few isolates from another host, clustered in one clade in the elsinochrome tree, except for the *E. australis* strain BRIP52616a, which differed from the rest of the isolates. The same was observed in the *ITS* tree except that *E. australis* BRIP5261a grouped with *E. australis* NL1 from *Populus tomentosa*.

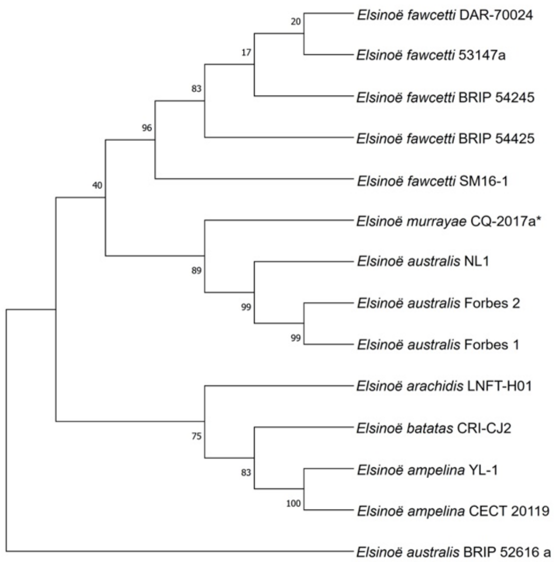

Figure 7: Maximum Likelihood tree of 351 nucleotide long β-tubulin following kimura 2 parameter of nucleotide model of substitution. The tree is rooted on *Myriangium duriaei* based on the study of Jeffress et al., (2020) Values on nodes represent percentage bootstrap support of 1000 replicates.

In addition to the observed distinct clusters, high nucleotide diversity and a high number of mutations in the core biosynthetic elsinochrome gene were observed using DNAsp, which had a number of mutations and segregating sites exceeding 3,000 (Table 5). However, further studies and samples are needed to evaluate the evolutionary relationships of *Elsinoë* spp. and to test the validity of the core elsinochrome gene as a potential candidate gene for determining host specialization in this genus.

Table 5: Genetic diversity of elsinochrome core biosynthetic genes.

Sequences type	Number of Sequences	Number of Mutations	Number of Segregating Sites	Nucleotide Diversity	Ave. Number of Nucleotide difference
Aligned and Trimmed	14	4923	3660	0.23251	1189.308
Unaligned and Untrimmed	14	14317	5372	0.68684	3691.769

DISCUSSION

The genus *Elsinoë* comprises known plant pathogenic species that affect a wide range of economically important hosts. However, studies on the biology of the obligate fungus within this genus are often limited and challenging. Whole-genome sequences are valuable resources for providing insights into the underlying basis of symptomatology and disease development in scab disease caused by *Elsinoë* spp. This study is among the first to detect candidate effectors and other secondary metabolites in *Elsinoë* spp. using whole-genome sequences and compare within the genus through *in silico* approaches. Furthermore, this study focused on the corresponding phylogeny and inferred the potential roles of these secreted effectors and secondary metabolites in host specialization, pathogenicity, and virulence.

Although the available genomes from the open-access database are smaller than the average size of Dothideomycetes in Ascomycota, similar genome assemblies have shown that those used in this study are within the range of *Elsinoë* spp. (Jeffress et al. 2020). Moreover, the study of Dothideomycetes by Haridas et al. (2020) suggested that the class of Dothideomycetes varies more than tenfold in size from less than 17 Mb to more than 117 Mb.

Across the different *Elsinoë* spp. used in this study, there were quite a few variations in the genome assemblies (Table 2). These included length and other genomic statistics, which may be attributed to certain factors, such as differences in sequencing technology. Most *Elsinoë* spp. were sequenced using Illumina sequencing technology, while recent genomes were generated using PacBio (Haridas et al. 2020). Some of these variations slightly affected the resulting number of predicted genes. For example, the *E. ampelina* YL isolate had 8,057 predicted genes according to Augustus 2.7 (Li et al. 2020); however, 9,731 predicted genes were obtained via GeneMarkES. The difference in the number of genes may be due to differences in the gene prediction programs used as well as differences in the probable criteria used with the various platforms used. Nevertheless, the predicted genes identified in this study are comparable to published data on *Elsinoë* spp.

For secondary metabolites, the *Elsinoë* spp. genomes used in this study were all shown to contain melanin as well as elsinochrome C (Table 4). Melanin has been associated with the fungal cell wall and virulence toward their hosts (Nosanchuk et al. 2015). Specifically, melanin plays a role in the penetration of plant cells by reducing the porosity of the appressorial wall; it also causes blackening and hardening at the site of infection (Jacobson 2000). Elsinochrome is generally a non-host selective toxin that is produced by a genus that exhibits toxicity via the accumulation of reactive oxygen species. However, there are other derivatives and amounts of elsinochrome produced that may be responsible for host specificity, although the cause is unclear (K.-R. Chung, 2011; Jiao et al., 2019; Liao & Chung, 2008a). There are currently four tautomers of elsinochrome that differ in their side chains, namely, elsinochrome A to D (Hu et al., 2019; Liao & Chung, 2008b; Yang & Chung, 2010). All the tested elsinochrome strains possessed the highest singlet oxygen yield, exhibiting ROS stress in response to the plant host. (Chung 2011; Liao and Chung, 2008). One of the core genes analyzed in the literature is the polyketide synthase-encoding gene *EfPKS1,* which is essential for elsinochrome biosynthesis and required for full virulence (Liao and Chung 2008). Interestingly, representative *E. ampelina* strains were found to harbor vertcillin, which is structurally related to chaetocin and interferes with the host chromatin remodeling machinery to enhance pathogenesis (Schenke et al. 2011; Wang et al. 2017). Verticillin may also be a secondary metabolite that contributes to anthracnose in grapes. Abscisic acid detected in the *E. arachidis* LNFT-H01 strain and in the *E. australis* NL1 strain (Izquierdo-Bueno et al. 2018) results in virulence by inhibiting the host immune response to the pathogen but has yet to be elucidated (Lievens et al. 2017).

Some of the predicted secondary metabolites, also found in other fungi, have no known roles in plant pathogenicity; these include duclauxin (Gao et al. 2018), Res-1214-2 fungal diphenyl ether (Xu et al. 2014), and trypacidin (Bignell et al. 2016). However, further studies are needed to validate the role of these secondary metabolites in relation to plant host invasion.

In terms of gene ontology, genes that might be related to pathogenicity, such as extracellular region (GO:0005576), antioxidant activity (GO:0016209), and toxin activity (GO:0090729), were identified and found to be similarly distributed across *Elsinoë* spp. (Supplementary Table 2) (Soliai et al. 2014). Approximately 100 genes were categorized as carbohydrate binding and carbohydrate derivative binding genes with functions related to the degradation of plant cell walls (Rodriguez-Moreno et al. 2018). Genes that were classified as antioxidant, detoxifying, or pigment-producing genes were also detected; these genes might aid in the defense against ROS generated by plant hosts as well as resistance to fungicides (Pedras and Abdoli 2017; Westrick et al. 2021). In the *E. ampelina* strains CECT 20119 and *E. arcahidis*, one gene confers necrosis (GO:0001906) (Chibucos et al. 2009).

Finally, the constructed phylogenetic tree, particularly of the core elsinochrome genes, showed groupings based on host, with the exception of *E. australis* BRIP 52616a; this was also observed in the ITS tree (Figure 7), with the exception of the *E. fawcettii* strain BRIP 54425, which was grouped with *E. australis* isolates. Fan et al. (2017) conducted a phylogenetic analysis of 75 *Elsinoë* spp. using ITS, LSU, rbp2, and Tef-1α. Their results included 26 new combinations of the genus described originally as *Sphaceloma,* and the trees seemed to be grouped according to host specialization. Phylogenetic analysis of *E. fawcettii* and other related *Elsinoë* spp. using *ITS* and *Tef-1α* sequences revealed that all the *E. fawcettii* isolates form a monophyletic group (Jeffress et al. 2020). The core elsinochrome biosynthetic genes may include the *ESCB1* gene, as observed in *E. arachidis* (Jiao et al. 2019); a polyketide synthase gene, which is in the same clade as EaPKS in *E. australis*; and EfPKS1, which is expressed under light conditions and is based on the genetic dissection of *E. fawcettii* (Liao and Chung 2008).

While the whole-genome resources available in open data sources for the plant pathogen *Elsinoë* vary, putative effectors and other secondary metabolites associated with pathogenicity and virulence, such as melanin and elsinochrome, are consistently detected. Moreover, according to elsinochrome C, Elsinoë spp. generally clustered according to the host. Overall, these results help to further elucidate the molecular basis of pathogenicity and virulence among *Elsinoë* spp., particularly those that are present in the country and affecting crops with economic importance, such as mango, citrus, avocado, and sweet potato.

AVAILABILITY OF DATA AND MATERIALS

All data analyzed and its supporting information are available in this paper. The NCBI accession number is available online (See methodology).

CONFLICT OF INTEREST

The authors declare no competing interests.

REFERENCES

Almagro Armenteros JJ, Tsirigos KD, Sønderby CK, Petersen TN, Winther O, Brunak S, von Heijne G, Nielsen H. SignalP 5.0 improves signal peptide predictions using deep neural networks. Nat Biotechnol 2019; 37:420-423

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol 1990; 215:403-410

Besemer J, Lomsadze, Borodovsky. GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding motifs in regulatory regions. Nucleic Acids Res 2001; 29:2607-2618

Bignell E, Cairns TC, Throckmorton K, Nierman WC, Keller NP. Secondary metabolite arsenal of an opportunistic pathogenic fungus. Philos Trans R Soc Lond B Biol Sci 2016; 371:1709

Blin K, Shaw S, Kloosterman AM, Charlop-Powers Z, Van Wezel GP, Medema MH, Weber T. AntiSMASH 6.0: Improving cluster detection and comparison capabilities. Nucleic Acids Res 2021; 49:W29-W35

Buchfink B, Reuter K, Drost HG. Sensitive protein alignments at tree-of-life scale using DIAMOND. Nat Methods 2021; 18:366-368

Cairns T and Meyer V. In silico prediction and characterization of secondary metabolite biosynthetic gene clusters in the wheat pathogen *Zymoseptoria tritici*. BMC Genom 2017; 18:1-16

Chibucos MC, Collmer CW, Torto-Alalibo T, Gwinn-Giglio M, Lindeberg M, Li D, Tyler BM. Programmed cell death in host-symbiont associations, viewed through the gene ontology. BMC Microbiol 2009; 9:S5

Chung KR. *Elsinoë fawcettii* and *Elsinoë australis*: the fungal pathogens causing citrus scab. Mol Plant Pathol 2011; 12:123-135

CropLife Philippines. Integrated pest management. In: Paredes JC and Cabaluna C, eds. Training manual on mango production in the Philippine. Manila: CropLife, 2018:27-68

Dacones LS, Kemerait RC, Brewer MT. Comparative genomics of host-specialized populations of *Corynespora cassiicola* causing target spot epidemics in the southeastern United States. Front Fungal Biol 2022; 3:1-11

De Jonge R, Ebert MK, Huitt-Roehl CR, … Bolton MD. Gene cluster conservation provides insights into cercosporin biosynthesis and extends production to the genus *Colletotrichum*. Proc Natl Acad Sci 2018; 115:E5459-E5466

Fan XL, Barreto RW, Groenewald JZ, Bezerra JDP, Pereira OL, Cheewangkoon R, Mostert L, Tian CM, Crous PW. Phylogeny and taxonomy of the scab and spot anthracnose fungus *Elsinoë* (Myriangiales, Dothideomycetes). Stud Mycol 2017; 87:1–41

Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M, Sangrador-Vegas A, Salazar GA, Tate J, Bateman A. The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res 2016; 4:D279-D285

Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, Smit AF. RepeatModeler2 for automated genomic discovery of transposable element families. Proc Natl Acad Sci 2020; 117:9451-9457

Gao SS, Zhang T, Garcia-Borràs M, Hung YS, Billingsley JM, Houk KN, Hu Y, Tang Y. Biosynthesis of heptacyclic duclauxins requires extensive redox modifications of the phenalenone aromatic polyketide. J Am Chem Soc 2018; 140:6991-6997

Haas BJ, Salzberg SL, Zhu W, Pertea M, Allen JE, Orvis J, White O, Buell CR, Wortman JR. Automated eukaryotic gene structure annotation using EVidenceModeler and the program to assemble spliced alignments. Genom Biol 2008; 9:1-22

Haridas S, Albert R, Binder M, Bloem J, LaButti K, Salamov A, Andreopoulos B, Baker SE, Barry K, Bills G, Bluhm BH, Cannon C, Castanera R, Culley DE, Daum C, Ezra D, González JB, Henrissat B, Kuo A, … Grigoriev IV. 101 Dothideomycetes genomes: a test case for predicting lifestyles and emergence of pathogens. Stud Mycol 2020; 96:141-153

Hu J, Sarrami F, Li H, Zhang G, Stubbs KA, Lacey E, Stewart SG, Karton A, Piggott AM, Chooi YH (2019). Heterologous biosynthesis of elsinochrome A sheds light on the formation of the photosensitive perylenequinone system. Chem Sci 2019; 10:1457-1465

Humann JL, Lee T, Ficklin S, Main D. Structural and functional annotation of eukaryotic genomes with GenSAS. Methods Mol Biol 2019; 1962:29-51

Izquierdo-Bueno I, González-Rodríguez VE, Simon A, Dalmais B, Pradier JM, Le Pêcheur P, Mercier A, Walker AS, Garrido C, Collado IG, Viaud M. Biosynthesis of abscisic acid in fungi: identification of a sesquiterpene cyclase as the key enzyme in *Botrytis cinerea*. Environ Microbiol 2018; 20:2469-2482

Jacobson, ES. Pathogenic roles for fungal melanins. Clin Microbiol Rev 2000; 13:708-717

Jeffress S, Arun-Chinnappa K, Stodart B, Vaghefi N, Tan YP, Ash G. Genome mining of the citrus pathogen *Elsinoë fawcettii*; prediction and prioritisation of candidate effectors, cell wall degrading enzymes and secondary metabolite gene clusters. PLoS ONE 2020; 5:e0227396.

Jiao W, Liu L, Zhou R, Xu M, Xiao D, Xue C. Elsinochrome phytotoxin production and pathogenicity of *Elsinoë arachidis* isolates in China. PLoS ONE 2019; 14: e0218391

Jones P, Binns D, Chang HY, Fraser M, Li WZ, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, Pesseat S, Quinn AF, Sangrador-Vegas A, Scheremetjew M, Yong SY, Lopez R, Hunter S. InterProScan 5: genome-scale protein function classification. Bioinformatics 2014; 30:1236-1240

Krogh A, Larsson B, von Heijne G, Sonnhammer ELL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol 2001; 305:567-580

Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: molecular evolutionary genetics analysis across computing platform. Mol Biol Evol 2018; 35:1547–1549

Li J, Cornelissen B, Rep M. Host-specificity factors in plant pathogenic fungi. Fungal Genet Biol 2020; 144:103447

Liao HL and Chung KR. Cellular toxicity of elsinochrome phytotoxins produced by the pathogenic fungus, *Elsinoë fawcettii* causing citrus scab. New Phytol. 2008; 177:239-250

Librado P, Rozas J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 2009; 25:1451-1452

Lievens L, Pollier J, Goossens A, Beyaert R, Staal J. Abscisic acid as pathogen effector and immune regulator. Front Plant Sci 2017; 8:587

Lopez D, Ribeiro S, Label P, Fumanal B … Pujade-Renaud V. Genome-wide analysis of *Corynespora cassiicola* leaf fall disease putative effectors. Front Microbiol 2018; 9:1-21

Majoros WH, Pertea M, Salzberg SL. TigrScan and GlimmerHMM: two open-source b initio eukaryotic gene-finders. Bioinformatics 2004; 20:2878-2879

Marais L. Avocado Diseases of Major Importance Worldwide and their Management. In Diseases of Fruits and Vegetables. Navqi SAMH ed. Springer, 2004:1-36

Nosanchuk JD, Stark RE, Casadevall A. (2015). Fungal melanin: what do we know about structure? Front Microbiol 2015; 6:1463

Pedras MSC, Abdoli A. Pathogen inactivation of cruciferous phytoalexins: detoxification reactions, enzymes and inhibitors. RSC Adv 2017; 7:23633–23646

Rodriguez-Moreno L, Ebert MK, Bolton MD, Thomma BPHJ. Tools of the crook- infection strategies of fungal plant pathogens. Plant J 2018; 93:664-674

Schenke D, Böttcher C, Lee J, Scheel D. Verticillin A is likely not produced by Verticillium sp. J Antibiot 2011; 64:523-524

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 2015; 31:3210-3212

Soliai MM, Meyer SE, Udall JA, Elzinga DE, Hermansen RA, Bodily PM, Hart AA, Coleman CE. De novo genome assembly of the fungal plant pathogen *Pyrenophora semeniperda*. PLoS ONE 2014; 9:e87045

Sperschneider J, Gardiner DM, Dodds PN, Tini F, Covarelli L, Singh KB, Manners JM, Taylor J M. EffectorP: Predicting fungal effector proteins from secretomes using machine learning. New Phytol 2016; 210(2), 743–761

Suchard MA, Lemey P, Baele G, Ayres DL, Drummond AJ, Rambaut A. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. Virus Evol 2018; 4:vey016

Świderska-Burek U, Daub ME, Thomas E, Jaszek M, Pawlik A, Janusz G. Phytopathogenic cercosporoid fungi – from taxonomy to modern biochemistry and molecular biology. Int J Mol Sci 2020; 21:8555

Wilgenbusch JC and Swofford D. Inferring evolutionary trees with PAUP*. Curr Protoc Bioinform 2003; 00:6.4.1-6.4.28

Van der Does HC and Rep M. Virulence genes and the evolution of host specificity in plant-pathogenic fungi. Mol Plant Microbe Interact 2007; 20:1175–1182

Wang Y, Hu P, Pan Y, Zhu Y, Liu X, Che Y, Liu G. Identification and characterization of the verticillin biosynthetic gene cluster in *Clonostachys rogersoniana*. Fungal Genet Biol 2017; 103:25-33

Westrick NM, Smith DL, Kabbage M. Disarming the host: detoxification of plant defense compounds during fungal necrotrophy. Front Plant Sci 2021; 12:684

Xu X, Liu L, Zhang F, Wang W, Li J, Guo L, Che Y, Liu G. Identification of the first diphenyl ether gene cluster for pestheic acid biosynthesis in plant endophyte *Pestalotiopsis fici*. ChemBioChem 2014; 15:284-292

Ye J, Zhang Y, Cui H, Liu J, Wu Y, Cheng Y, Xu H, Huang X, Li S, Zhou A, Zhang X, Bolund L, Chen Q, Wang J, Yang H, Fang L, Shi C. WEGO 2.0: a web tool for analyzing and plotting GO annotations, 2018 update. Nucleic Acids Res 2018; 46:W71–W75

SUPPLEMENTARY DATA

Supplementary Table 1: Additional sequences used in phylogenetic study

Name of Sequence	Gene Regions	Max Score	Query Cover	E-Value	Percent Identity
Elsinochrome					
E. australis BRIP 52616 a	74,247 to 79,633	4,429	96%	0.0	76.28%
E. australis Forbes_2	81,962 to 18,7348	4,494	99%	0.0	76.50%
E. fawcettii SM16-1	133,963 to 139,337	4,205	99%	0.0	80.33%
E. fawcettii BRIP 54245 a	118,227 to 123,601	5,404	99%	0.0	79.80%
E. fawcettii BRIP 54425 a	118,800 to 124,174	5,404	99%	0.0	79.82%
ITS					
E. australis BRIP 52616 a	2,685 to 3,284	673	100%	0.0	88.89%
E. australis Forbes_2	4,115 to 4,715	768	100%	0.0	88.71%
E. fawcettii SM16-1	1,805 to 2,113	670	95%	0.0	89.92%
E. fawcettii BRIP 54245 a	1,806 to 2,114	670	95%	0.0	89.92%
E. fawcettii BRIP 54425 a	740,823 to 741,131	670	95%	0.0	89.92%
Beta Tubulin					
E. australis BRIP 52616 a	1,106,926 to 1,107,329	214	100%	0.0	75.07%
E. australis Forbes_2	12,081 to 12,482	217	100%	2e-56	75.67%
E. fawcettii SM16-1	1,635,520 to 1,635,922	146	65%	5e-35	90.57%
E. fawcettii BRIP 54245 a	224,363 to 22,475	239	97%	7e-63	77.27%
E. fawcettii BRIP 54425 a	683,019 to 683,421	239	97%	7e-63	77.27%

Supplementary Table 2: Summary of the number of gene functions among the *Elsinoë* spp. used in this study.

Function/Isolate	*E. ampelina* CECT 20119	*E. ampelina* YL-1	*E. arachidis* LNFT-H01	*E. batatas* CRI-CJ2	*E. australis* NL1	*E. australis* Forbes_1	*E. fawcettii* 53147a	*E. fawcettii* DAR-70024	*E. murrayae* CQ-2017a
Cell	4645	4700	4888	4645	4473	4503	4811	4883	4188
Membrane	2591	2618	2782	2552	2472	2493	2826	2854	2185
Extracellular Region	290	345	382	323	312	326	350	356	253
Symplastic	1	2	2	2	3	2	2	2	2
Other Organism	0	1	1	1	0	0	1	1	2
Catalytic Activity	4283	4335	4638	5242	4178	4200	4672	4661	3623
Carbohydrate Derivative Binding	985	955	1063	985	896	934	1041	1041	818
Carbohydrate Binding	114	121	118	102	106	105	106	108	79
Antioxidant Activity	54	56	57	50	57	57	58	58	50
Toxin Activity	2	3	1	1	0	0	1	1	3
Response to Stimulus	1049	1067	1110	1045	1035	1054	1083	1081	955
Secondary Metabolic Process	76	72	82	66	70	65	84	87	64
Pigment Metabolic Process	40	35	40	41	33	32	38	38	31
Biological Adhesion	49	58	62	55	48	61	58	62	38
Reproductive Process	106	124	133	121	119	123	128	130	105
Detoxification	38	43	41	42	43	39	43	43	36
Cell Killing	1	1	1	0	0	0	0	0	0
Carbon Utilization	1	2	2	2	2	2	2	2	2
Pigmentation	0	1	1	1	0	0	1	1	1