GENOME MINING AND COMPARATIVE GENOMICS OF *ELSINOË* SPECIES,
CAUSING SCAB DISEASES ON CITRUS AND JOJOBA IN AUSTRALIA, FOR THE
PREDICTION OF VIRULENCE GENES.

A Thesis submitted by

Sarah Jeffress, BMedSc, BForensicBiotech, BSc(Hons)

For the award of

Doctor of Philosophy

2021

ABSTRACT

Elsinoë fawcettii and *E. australis* are necrotrophic fungal pathogens which cause citrus scab and sweet orange scab, respectively. *E. fawcettii* causes disease on a wide range of citrus varieties around the world, while *E. australis* causes disease on sweet orange in South America. Recently, novel pathotypes of *E. australis* have been identified in Australia, including the Jojoba Black Scab (JBS) pathotype which causes black scab disease on jojoba and the Finger Lime (FL) pathotype which causes disease on finger lime fruit. As citrus is an economically important crop in Australia, research into pathogens of citrus and closely related pathogens is needed. It is known that species of *Elsinoë* produce elsinochrome, a secondary metabolite which contributes towards virulence in a non-host selective manner. However, the pathogen-host processes occurring before the release of elsinochrome are not understood, yet host-specificity is well known among the species and pathotypes suggesting such processes occur. This study produced genome assemblies for two species of *Elsinoë*, including four *E. fawcettii* (25.8 – 26.4 Mb), four *E. australis* JBS pathotype (26.6 Mb) and one *E. australis* FL pathotype (23.8 Mb). Phylogenetic analyses were conducted indicating the *E. fawcettii* pathotypes were very closely related, while the *E. australis* pathotypes were distinct from one another. Genome mining for pathogenicity related genes indicated that *Elsinoë* have a repertoire of predicted candidates including five secondary metabolite gene clusters, in addition to elsinochrome, that are conserved among *Elsinoë*, as well as 186 - 219 cell wall degrading enzymes and 94 – 120 prioritised effectors which may be enabling their host-specific pathogenesis pathways. Comparative analyses with other *Elsinoë* and with other fungal plant pathogens (*Botrytis cinerea*, *Parastagonospora nodorum*, *Pyrenophora tritici-repentis*, *Sclerotinia sclerotiorum*, *Zymoseptoria tritici*, *Leptosphaeria maculans*, *Pyricularia oryzae* (syn. *Magnaporthe oryzae*), *Rhynchosporium commune*, *Verticillium dahliae* and *Ustilago maydis*) assisted with the shortlisting of candidate effectors. Genomic regions of potential plasticity were also identified, indicating predicted virulence related genes of some *Elsinoë*, such as candidate effectors, were more likely to be located in the close vicinity of transposable elements and adenine and thymine (AT)-rich regions. This provides an insight into the potential maintenance of pathogenicity-related genes of *Elsinoë*, in

addition it highlights regions of the genome which may hold further virulence factors not predicted by current methods. This study has established a comprehensive resource for future experimental validation of genes which *Elsinoë* utilise in order to infect plants in a host selective manner.

CERTIFICATION OF THESIS

This Thesis is entirely the work of Sarah Jeffress except where otherwise acknowledged, with the majority of the authorship of the paper presented undertaken by the student. The work is original and has not previously been submitted for any other award, except where acknowledged.

Principal Supervisor: Prof. Gavin Ash (USQ)

Associate Supervisor: Dr. Kiruba Arun Chinnappa (USQ)

Associate Supervisor: Dr. Ben Stodart (CSU)

Student and supervisors' signatures of endorsement are held at the University.

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

AA: amino acids

A: adenine

AGRF: Australian Genome Research Facility

AT: adenine and thymine

B. cinerea: *Botrytis cinerea*

bp: base pairs

C: cytosine

C. aurantium: *Citrus aurantium*

C. australasica: *Citrus australasica*

C. clementina: *Citrus clementina*

C. grandis: *Citrus grandis*

C. jambhiri: *Citrus jambhiri*

C. junos: *Citrus junos*

C. limon: *Citrus limon*

C. limonia: *Citrus limonia*

C. natsudaidai: *Citrus natsudaidai*

C. obovoidea: *Citrus obovoidea*

C. paradisi: *Citrus paradisi*

C. reshni: *Citrus reshni*

C. reticulata: *Citrus reticulata*

C. sunki: *Citrus sunki*

C. unshiu: *Citrus unshiu*

CAZy database: Carbohydrate Active Enzymes Database

CDS: coding DNA sequence

CE: candidate effectors

CRISPR: clustered regularly interspaced short palindromic repeats

CSU: Charles Sturt University

CWDE: cell wall-degrading enzymes

DAF: Department of Agriculture and Fisheries

DAWR: Department of Agriculture and Water Resources

E. ampelina: *Elsinoë ampelina*

E. australis: *Elsinoë australis*

E. batatas: *Elsinoë batatas*

E. citricola: *Elsinoë citricola*

E. fawcettii: *Elsinoë fawcettii*

E. genpiae-americanae: *Elsinoë genpiae-americanae*

E. piri: *Elsinoë piri*

E. punicae: *Elsinoë punicae*

E. randii: *Elsinoë randii*

E. tiliae: *Elsinoë tiliae*

EfPKS1*: Elsinoë fawcettii* polyketide synthase 1

FBHR: Florida Broad Host Range

FL: finger lime

FNHR: Florida Narrow Host Range

G: guanine

GC: guanine and cytosine

GPI: glycosylphosphatidylinositol

GTR: General Time Reversible

HIGS: host-induced gene silencing

HSP: high-scoring segment pair

IFR: intergenic flanking region

IGV: Integrated Genome Viewer

ITS: internal transcribed spacer

JBS: Jojoba Black Scab

Kb: kilobases

L. maculans: *Leptosphaeria maculans*

LAMP: Loop-Mediated Isothermal Amplification

LSU: ribosomal RNA large subunit

M. hispanicum: *Myriangium hispanicum*

Mb: megabases

ML: maximum likelihood

N: Unknown bases

N. crassa: *Neurospora crassa*

NCBI: National Center for Biotechnology Information

NRPS: non-ribosomal peptide synthase

NSW: New South Wales

PCE: prioritised candidate effectors

PCR: polymerase chain reaction

Pfam: Protein Family Database

PHI-base: Pathogen-Host Interaction Database

Pma1: plasma membrane ATPase

polyAA: polyamino acid

P oryzae: Pyricularia oryzae

R genes: resistance genes

rDNA: ribosomal DNA

RIP: repeat induced point mutation

RNAi: RNA interference

RPB2: DNA-directed RNA polymerase II second largest subunit

rRNA: ribosomal RNA

S. chinensis: *Simmondsia chinensis*

S. punctatus: *Spizellomyces punctatus*

S. sclerotiorum: *Sclerotinia sclerotiorum*

SM: secondary metabolites

SNP: single nucleotide polymorphisms

SO: sweet orange

SP: secreted proteins

SRA: Short Read Archive

SSR: short simple repeat

T: thymine

T1PKS: type I polyketide synthase

T3PKS: type 3 polyketide synthase

TE: transposable element

TEF1-α: translation elongation factor 1 alpha

TM: transmembrane

TN93: Tamura-Nei model

tRNA: transfer RNA

TSF1: transcription factor 1

U. maydis: *Ustilago maydis*

USQ: University of Southern Queensland

V. dahliae: *Verticillium dahliae*

WGS: whole genome shotgun

Z. tritici: *Zymoseptoria tritici*

β-tubulin: beta-tubulin

CHAPTER 1: INTRODUCTION

1.1 OVERVIEW

Species of *Elsinoë* are necrotrophic fungal pathogens within the Ascomycota phylum (subphylum Pezizomycotina, class Dothideomycetes, subclass Dothideomycetidae, order Myriangiales). *Elsinoë australis* Bitancourt & Jenkins and *E. fawcettii* Bitancourt & Jenkins are filamentous phytopathogens which cause necrotic disease, known as sweet orange scab and citrus scab, respectively, to many citrus crops worldwide (Bitancourt & Jenkins, 1936, 1937; Hyun et al., 2001; Hyun et al., 2009; Tan et al., 1996; Timmer et al., 1996; Whiteside, 1975, 1978; Xin et al., 2014). Necrotrophic pathogens obtain nutrients and energy from dead cells, which contrasts with biotrophic pathogens, which obtain nutrients and energy from living cells, and hemi-biotrophic pathogens, which live in a biotrophic manner before switching to a necrotrophic lifestyle (Tronsmo et al., 2020). *Elsinoë* are filamentous fungi commonly described as anamorphs, and their teleomorph has only been reported in Brazil (Bitancourt & Jenkins, 1936, 1937), suggesting these pathogens mainly reproduce asexually. Both pathogens cause scab diseases, blemishing fruit with necrotic lesions; while not poisonous the fresh market value is reduced. Annually, over 750,000 tonnes of oranges, mandarins, lemons, limes, grapefruit and tangelo are produced in Australia, of which in 2019, almost 273,000 tonnes were exported with a reported value over $500 million (Citrus Australia, 2019, 2020). Protecting these crops from disease is an important step in maintaining Australia's reliability in producing high quality fresh fruit. *E. fawcettii* infects citrus varieties worldwide, while *E. australis*, a major pathogen of sweet oranges (*Citrus sinensis*), is predominantly found in South America (Hyun et al., 2009). Recently two new putative pathotypes of *E. australis* were reported in Australia, one causing scab disease to jojoba crops in New South Wales (NSW) (Ash, Stodart, & Hyun, 2012) and a second on finger lime (*C. australasica*) fruit in forest regions of Queensland (Miles et al., 2015). Jojoba (*Simmondsia chinensis*) is cultivated for its unique oil which is used in pharmaceuticals, cosmetics, and lubricants, and as it thrives in arid and

semi-arid areas (reviewed by Sánchez et al. (2016)), Australia is an ideal location for this crop.

Three further species of *Elsinoë* which cause crop disease are *E. ampelina*, causing anthracnose of grapes (dos Santos et al., 2018), *E. batatas*, affecting sweet potato in Papua New Guinea (Kokoa, 2001; Mau, 2018) and *E. piri*, affecting organic apples in New Zealand (Scheper, Wood, & Fisher, 2013). Whole genome sequencing has recently been published for *E. fawcettii* and *E. australis* (Shanmugam et al., 2020), as well as *E. ampelina* (Li et al., 2020). Symptoms of diseases caused by *Elsinoë* are illustrated in Figure 1.
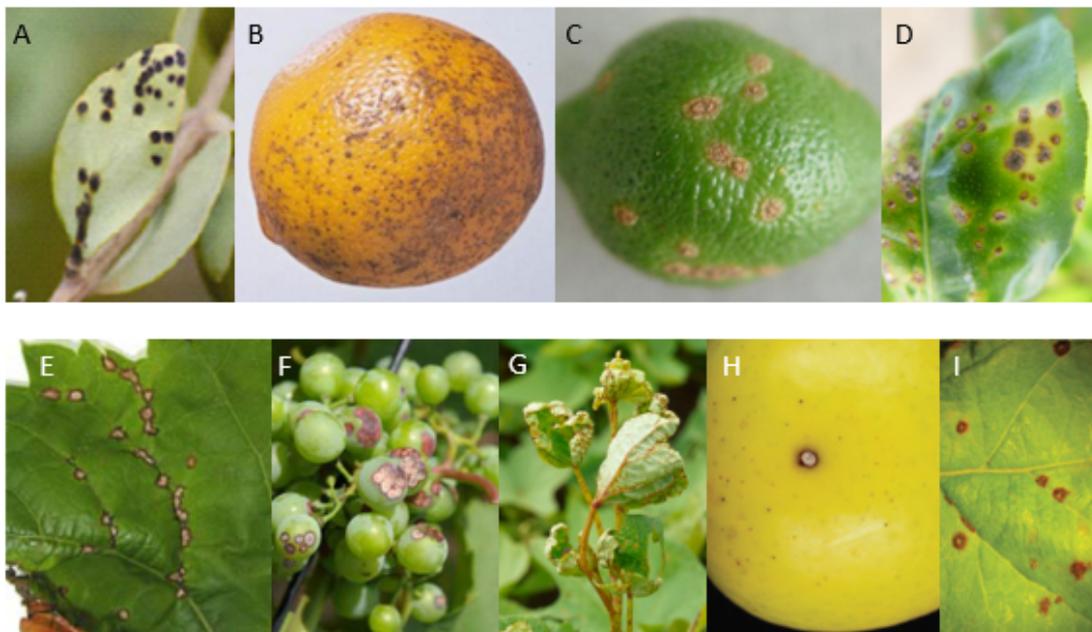
Figure 1. Diseases caused by *Elsinoë*. A. Black scab of jojoba (putative *E. australis* pathotype). B. Sweet orange scab (*E. australis*). C-D. Citrus scab on fruit and leaves (*E. fawcettii*). E-F. Anthracnose of grapes (*E. ampelina*). G. Distorted leaves of sweet potato (*E. batatas*). H-I. Diseased fruit and leaves of apple tree (*E. piri*). Sourced from Ash et al. (2012); Carisse and Lefebvre (2011); McTaggart (2007); O'Sullivan (2017); Scheper et al. (2013); Silveira et al. (2007).

1.1.1 *ELSINOË AUSTRALIS*

Currently, most research surrounding *Elsinoë australis* has focused on the sweet orange pathotype found only in South America (Chung, 2011; Hyun et al., 2009; Shanmugam, Jeon, & Hyun, 2020). In this study, genomes of the two putative *E. australis* pathotypes, collected from necrotic lesions of jojoba and finger lime in Australia, were investigated and compared with the genome of the sweet orange pathotype (Arg-1) from Argentina (Shanmugam et al., 2020).

The putative Jojoba Black Scab (JBS) pathotype which causes disease of jojoba (*Simmondsia chinensis*), produces necrotic lesions on leaves, calyces and stems, resulting in leaf drop and is found widespread in jojoba plantations in Eastern Australia (Ash et al., 2012; Miles et al., 2015). Initial sequencing of the internal transcribed spacer (ITS) regions of ribosomal DNA (rDNA) of fungal isolates from jojoba indicated 100% identity to *E. australis* the sweet orange pathogen, leading to the classification of the jojoba pathogen as a putative *E. australis* pathotype (Ash et al., 2012). Whole genome shotgun (WGS) sequencing of eight fungal isolates collected from Australian jojoba plants was completed by Charles Sturt University (CSU), generating a wealth of data for analysis in the current project, thus making further investigations into its identity and into its repertoire of virulence genes possible.

Recently, reclassification of some species of *Elsinoë* has relied on phylogenetic analysis of four genetic regions, namely, the ITS, translation elongation factor 1 alpha (TEF1-α), ribosomal RNA large subunit (LSU) and DNA-directed RNA polymerase II second largest subunit (RPB2). For 75 species of *Elsinoë*, it was demonstrated that many species shared identity in one or more of these four regions. For example, the ITS, LSU and TEF1-α regions do not differentiate *E. fici* from *E. randii*, however these species can be identified using the RPB2 region and by host range, *Ficus luschnathiana*

and *Carya* sp., respectively (Fan et al., 2017). Of particular interest to the current study, it was demonstrated that the ITS region alone did not distinguish *E. australis* isolates, as it was identical to *E. punicae* and *E. genipae-americanae* (Fan et al., 2017). Therefore, a multi-locus phylogenetic analysis of the *E. australis* JBS pathotype together with closely related species has the potential to further investigate the true identity of this pathogen. Furthermore, host and tissue specificity are characteristics that can assist in the identification of species within *Elsinoë*. Comparison of these traits between *E. australis* isolated from sweet orange, from this point on referred to as the sweet orange (SO) pathotype, and the *E. australis* JBS pathotype demonstrates significant differences. The *E. australis* SO pathotype is currently well known for infection of sweet orange (*Citrus sinensis*), satsuma mandarin (*C. unshiu*), tangerine (*C. reticulata*), grapefruit (*C. paradisi*) and natsudaidai (*C. natsudaidai*) (Hyun et al., 2001; Hyun et al., 2009; Kunta et al., 2013), whereas the *E. australis* JBS pathotype is known to infect jojoba and not sweet orange, lemon (*C. limon*), rough lemon (*C. jambhiri*), mandarin (*C. reticulata*) or grapefruit (Ash et al., 2012). Furthermore, tissue specificity is also often seen among these species. *E. australis* (SO pathotype) causes lesions predominantly to the skin of citrus fruit, and not the leaves (Timmer et al., 1996), whereas the JBS pathotype causes disease to the leaves and stems of jojoba, but not the fruit (Ash et al., 2012).

Additionally, the *E. australis* finger lime (FL) pathotype collected from scab lesions on only the fruit skin of finger lime (*C. australasica*), a fruit endemic to forest regions of coastal NSW and QLD in Australia, has been previously studied. The FL pathotype (BRIP 52616a) was identified using the ITS and TEF1-α in a phylogenetic analysis with both the SO and JBS pathotypes (Miles et al., 2015). This comparison indicated while all three pathotypes were grouped in the same clade, they were distinct from one another. Miles et al. (2015) further determined the FL pathotype was also pathogenic to *Citrus × aurantium* cv. Murcott tangor. The finger lime is a fruit found growing in the wild and more recently also as an agricultural crop on a small scale. The fruit has gained awareness around the world, known as The Australian

4

Caviar (Delort, 2018). The genome of one isolate of *E. australis* FL pathotype was sequenced and used as a comparative species in the current study.

1.1.2 *ELSINOË FAWCETTII*

Elsinoë fawcettii causes necrotic disease to a range of citrus crops including 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*) (Hyun et al., 2001; Hyun et al., 2009; Tan et al., 1996; Xin et al., 2014). Multiple pathotypes of *E. fawcettii* have been identified and can be distinguished from one another by host range. For example, the Florida Broad Host Range (FBHR), Florida Narrow Host Range (FNHR), Tyron's, Lemon, Jinguel, SRGC and Satsuma Mandarin, as well as several reported cryptic and novel pathotypes (Hyun et al., 2009; Wang et al., 2009; Xin et al., 2014). In Australia, the only documented pathotypes include the Tyron's pathotype (affecting Eureka lemon, Rough lemon, clementine, Rangpur lime and Cleopatra mandarin) and the Lemon pathotype (affecting Eureka lemon, Rough lemon and Rangpur lime) (Hyun et al., 2001; Hyun et al., 2009; Timmer et al., 1996). However, in Australia it has been reported that *E. fawcettii* was isolated from kumquat (*Fortunella* sp.) in 2009, tea plant (*Camellia sinensis*) in 2004 and mango (*Mangifera indica*) in 2005 (DAF Biological Collections, 2019), suggesting that additional pathotypes exist, or current pathotypes have a wider host range than currently known. In the current study, the genomes of four isolates of *E. fawcettii* (BRIP 53147a, BRIP 54245a, BRIP 54425a and BRIP 54434a) were sequenced and compared to the assemblies of two recently published isolates (DAR 70024 and SM16-1) (Shanmugam et al., 2020).

1.2 LIFECYCLE OF *ELSINOË*

The lifecycle cycle of *Elsinoë* begins once hyaline conidia are transported by rain or wind from necrotic scab lesions on mature diseased fruit to the fruit, twigs and leaves of young plants (Paudyal & Hyun, 2015; Timmer, Garnsey, & Graham, 2000). Young leaves are susceptible to infection from when shoots first appear through to when they are half expanded, while fruit is susceptible from after petal fall for 6 to 8 weeks (Timmer et al., 2000). The conidia penetrate the plant tissue by extending their germ tubes. Cuticle and mesophyll cells in the infected area degrade, resulting in a necrotic region which hyphae then colonise (Paudyal, Hyun & Hwang (2017). For infection of host plants, conidia require four hours of contact with water and temperatures from 23.5°C to 27°C. One to two days after inoculation of young plants, degradation occurs in the cuticle, epidermal cells and mesophyll tissue and hyphal generation begins. After three to four days, visible symptoms appear and necrotic lesions are produced on fruit, twigs and leaves (Chung, 2011; Paudyal & Hyun, 2015). Conidia are then produced from the pustules on the scab lesions, becoming available to inoculate further plants and generate wider spread. After five days the host cell walls lignify around infected areas, limiting localised pathogen spread (Paudyal & Hyun, 2015). The pathogen survives from season to season as conidia can generate from previously infected fruit, leaves or twigs which remain on the plant between seasons (Batuman & Ritenour, 2020). It has also been suggested that *E. australis* specifically may survive as an endophyte or epiphyte prior to causing disease, as a basis for how the pathogen survives between seasons (Kunta et al., 2013).

1.3 SECONDARY METABOLITE PRODUCTION IN *ELSINOË*

During infection, necrosis of plant tissue develops after release of elsinochrome, a necrotising secondary metabolite of the *Elsinoë* pathogens. Between pathogen and host, elsinochrome initiates the production of reactive oxygen species, in aerobic and light-activated conditions this toxic

environment is created in a non-host selective manner (Liao & Chung, 2008b). Currently, there are two genes within the secondary metabolite (SM) gene cluster, being *Elsinoë fawcettii* polyketide synthase 1 (EfPKS1) and transcription factor 1 (TSF1), that are known to be required for elsinochrome production (Chung & Liao, 2008; Liao & Chung, 2008a). However, simple production of the elsinochrome toxin does not fully explain the pathogenesis observed. For example, the amount of elsinochrome produced varies and does not correlate with virulence (Wang, Bau, & Chung, 2009). Secondly, species and pathotypes of *Elsinoë* cause host-specific disease, yet the elsinochrome toxin is non-host selective. To account for the observed host specificity, it is possible that, prior to the production of elsinochrome, host-specific proteins of the pathogen are targeted towards, and inhibit, host proteins involved in immune defence. These potential host-specific proteins may be effectors and/or cell wall degrading enzymes.

1.4 DIFFERENTIATION OF *E. AUSTRALIS* AND *E. FAWCETTII*

Elsinoë australis and *E. fawcettii* are not easily distinguished by morphology and cultural characteristics (Timmer et al., 1996). Comprehensive differentiation of these species and pathotypes has only been achieved through host-pathogenicity testing (Ash et al., 2012; Hyun et al., 2009; Timmer et al., 1996; Wang et al., 2009; Xin et al., 2014). While species-specific primers may differentiate the two species from one another (Hyun et al., 2007), distinguishing pathotypes proves challenging. Sequencing of the ITS and TEF regions allows differentiation of pathotypes within *E. australis*, but not those within *E. fawcettii* (Hyun et al., 2009; Kunta et al., 2013; Miles et al., 2015). As mentioned previously, the recent phylogenetic analysis of a wider species range of *Elsinoë* (Fan et al., 2017) revealed the ITS region of *E. australis* (isolated from citrus) was indistinguishable from that of *E. genipae-americanae* and *E. punicae*.

1.5 CURRENT MECHANISMS OF *ELSINOË* CONTROL AND POTENTIAL FUTURE STRATEGIES

Currently *Elsinoë* citrus scab diseases are managed by the application of protectant fungicides, such as copper, ferbam, thiram, difenoconazole and chlorothalonil, and systemic fungicides, such as benomyl and carbendazim (CABI, 2017). Due to the unwanted introduction of chemicals into the environment, new alternatives to fungicide application are preferred, such as those targeting pathogen-specific genes. For example, host-induced gene silencing (HIGS), an emerging technique which involves host plant RNA interference (RNAi) against a specific and essential gene of a pathogen (as reviewed by Koch and Kogel (2014)). This mechanism has been successful for the inhibition of *Fusarium* spp. in barley (Koch & Kogel, 2014), *Blumeria graminis* in barley and wheat (Nowara et al., 2010) and *Puccinia triticina* in wheat (Panwar, McCallum, & Bakkeren, 2013), indicating potential as a future crop protection strategy. Furthermore, an additional technique which can be used during the generation of pathogen-resistant plants is clustered regularly interspaced short palindromic repeats (CRISPR)/Cas9 (Jinek et al., 2012), a gene editing mechanism which has assisted in the production of numerous pest-resistant transgene-free plants. For example, the mutagenesis of a susceptibility gene in tomato (Nekrasov et al., 2017), grapevine and apple (Malnoy et al., 2016) to increase resistance to powdery mildew pathogens. Identifying candidate pathogenicity-related genes of *Elsinoë* may assist in the development of improved management methods. Current biosecurity methods to prevent the introduction of sweet orange scab to Australia include visual inspection of fruit, and brushing, washing and waxing of oranges imported from regions affected by *E. australis* (DAWR, 2013). While visual inspection may frequently identify highly diseased fruit, less well-established infections may be missed. Furthermore, if *E. australis* does survive as an endophyte or epiphyte prior to causing disease, as Kunta et al. (2013) suggest, visual inspection would be an unacceptable method of detection. As identification of *Elsinoë* by morphology is unreliable, and ITS sequencing is expensive and time consuming, molecular diagnostic tools such as multiplex conventional polymerase chain reaction (PCR) and

quadruplex real-time PCR have recently been developed as a rapid method to distinguish *E. fawcettii* from *E. australis* on citrus (Ahmed et al., 2020). In addition to these, further methods to identify the JBS pathotype, need to be developed. This could include Loop-Mediated Isothermal Amplification (LAMP), for example. The use of comparative genomics has assisted in primer design for LAMP diagnostics to identify the rice pathogen *Pseudomonas fuscovaginae* (Ash et al., 2014). Generating genomic resources and performing comparative analyses of multiple pathotypes will enable the identification of suitable genomic regions for use in these molecular tools.

1.6 SIGNIFICANT GENOMIC FEATURES OF FUNGAL PATHOGENS

1.6.1 GENE ORTHOLOGY CLASSIFICATION

Through comparative genomics it is possible to identify core, accessory and unique gene components between organisms of interest (Verma et al., 2016). Core genes are shared by all comparative species, accessory genes are shared by at least two species, but not all, and unique genes are found in only one species. Gene orthology classification may be associated with pathways involved in pathogenesis. For example, core genes, also predicted as virulence-associated genes, may be involved in broad host range elements of pathogenesis, while unique or accessory genes may be involved in host- or tissue-specific elements of pathogenesis. Accessory and unique genes that are functionally validated for their role in causing disease can be useful markers of pathogenicity (Coleman et al., 2009; Ma et al., 2010). The elucidation of genes involved in host specialisation would be a major step towards understanding the mechanisms of pathogenesis utilised by *Elsinoë* species. Host specialisation is known to occur through a variety of methods such as hybridisation (Menardo et al., 2016), transposable elements

(Yoshida etal., 2016), and chromosomal rearrangements (Hartmann et al., 2017).

1.6.2 FUNGAL EFFECTOR GENES

Fungal effectors are secreted proteins involved in pathogenesis, targeted to the host cytoplasm or apoplast. Effectors allow the pathogen to evade activities of the host's defence system, retaliating against pathogen-triggered defences of the host, and if successful, infection proceeds (Kamoun, 2006). For example, the Slp1 effector of *Pyricularia oryzae* (syn. *Magnaporthe oryzae*) binds to chitin in the fungal cell wall to inhibit its detection by the host plant (Mentlak et al., 2012), the Avr2 effector of *Cladosporium fulvum* inhibits multiple host proteases required by the host defence system (Rooney et al., 2005; van Esse et al., 2008) and Tin2 of *Ustilago maydis* binds a host protein kinase, thought to effect downstream biosynthesis of lignin, reducing the ability of the host to strengthen its cell walls in response to fungal invasion (Tanaka et al., 2014). In resistant hosts, however, effectors are recognised by plant resistance (R) genes that initiate effector-triggered immunity in the plant and in these cases, pathogenesis fails (Hogenhout et al., 2009; Kamoun, 2006). It has been assumed that necrotrophic fungal pathogens rely on carbohydrate-active enzymes or secondary metabolites (SM) to infect target hosts (Bolton, Thomma, & Nelson, 2006), however it now appears their use of protein effectors may provide an important element to their pathogenesis (Friesen et al., 2008; Lo Presti et al., 2015; Lyu et al., 2016; Rodriguez-Moreno et al., 2018; Wang et al., 2014; Yu et al., 2017). Genomic and proteomic features of effectors can include; an N-terminal signal peptide in the protein sequence, absence of glycosylphosphatidylinositol (GPI) anchors and transmembrane (TM) helices, small size, disulphide bonds, repetitive DNA sequences, gene duplication, absence of conserved protein domains, encoded in transposon-rich genomic locations, amino acid polymorphism, and absence in non-pathogenic strains (Chuma et al., 2011; Guyon et al., 2014; Manning et al., 2013; Saunders et

al., 2012; Sperschneider et al., 2014; Syme et al., 2013; Ve et al., 2013). Additionally, some protein effectors are species-specific, for example NIP1, NIP2 and NIP3 of *Rhynchosporium commune* (Kirsten et al., 2012), while others are coded by genes shared by multiple pathogens, or have similar protein domains, for example, Cmu1 (chorismate mutase effector) of *U. maydis* (Djamei et al., 2011) and MoCDIP1 (cell death-inducing effector) of *P. oryzae* (syn. *M. oryzae*) (Chen et al., 2013). Chorismate mutsases are found in a variety of plant parasites including bacteria, protozoans, and nematodes (Kim et al., 2020, Roberts et al., 1998 & Wang et al., 2018), while homologs of MOCDIP1 have been found in multiple species including *Phaeosphaeria nodorum*, *Aspergillus nidulans* and *Streptomyces griseoflavus* (Chen et al., 2013). While effectors do not contain all features listed, using them as a guide and analysing combinations of these features can shortlist candidates for experimental validation (Sperschneider et al., 2015). While it is known that the elsinochrome toxin produced by species of *Elsinoë* contributes to necrosis (Liao & Chung, 2008a), effectors are unknown, as are their target host proteins. In the current study, we take the initial step of identifying candidate fungal effectors, which in the future can be tested for interaction with host plant proteins and effects on pathogenesis. The ultimate goal is the inhibition of these interactions, to enable resistance or increase tolerance. As previously mentioned, potential mechanisms include host induced gene silencing against a pathogen effector, or the generation of genetically modified crops using gene editing methods, such as CRISPR, to integrate R genes into susceptible hosts or modify interacting plant proteins so they are not recognised and therefore the host remains unaffected.

1.6.3 CELL WALL DEGRADING ENZYMES

Cell wall-degrading enzymes (CWDE) are virulence associated proteins which take part in fungal pathogenesis. These secreted proteins include glycoside hydrolases, polysaccharide lyases and carbohydrate esterases

and enable the cleavage of plant cell wall components (Cantarel et al., 2009; King et al., 2011; Kubicek, Starr, & Glass, 2014). Plant cell wall components targeted by pathogens include cellulose, hemicelluloses (xyloglucan and arabinoxylan) and pectin (rhamnogalacturonan I, homogalacturonan, xylogalacturonan, arabinan and rhamnogalacturonan II) (Cosgrove, 2005). CWDE can enable pectin degradation, for example polygalacturonases, pectate lyases, and pectinesterases (Bravo Ruiz, Di Pietro, & Roncero, 2016; López-Pérez, Ballester, & González-Candelas, 2015; Rogers et al., 2000; Shieh et al., 1997; Valette-Collet Bravo Ruiz, 2003; Yakoby et al., 2001; Zuppini et al., 2005), cleavage of glucose molecules (e.g. glucanase or cellulase) (Fu et al., 2013) or breaking links within xyloglucan (e.g. xylanase) (Brito, Espino, & González, 2006; Nguyen et al., 2011; Noda, Brito, & Gonzalez, 2010). CWDE enable pathogens to obtain nutrients and cause physical degradation to the host.

1.6.4 TWO-SPEED GENOMES

Some fungal pathogens have a two-speed genome, also known as a compartmentalised or bipartite genome, containing slow (conserved) core genomic regions, as well as fast genomic regions showing higher rates of molecular evolution. This contrast in rate of variation is thought to enable some filamentous plant pathogens to protect essential genes while generating variation in regions of the genome coding for pathogenicity-related genes (Croll & McDonald, 2012; Raffaele & Kamoun, 2012). Two-speed genomes can show an uneven distribution of Single Nucleotide Polymorphisms (SNP), gene duplications, repetitive regions, positive selection, guanine and cytosine (GC) content and/or transposons (Coleman et al., 2009; Grandaubert et al., 2014; Stukenbrock et al., 2010; Wang et al., 2017), providing much needed genetic variation in virulence genes of asexual fungal pathogens. This contrasts with a one-speed genome when an uneven distribution has not been identified (Stam et al., 2018; Frantzeskakis et al., 2018). Presumably, higher rates of variation in pathogenicity-related

genes would allow a pathogen to maintain virulence in the face of host immune defences and provide greater potential to adapt to environmental changes and opportunities to colonize new hosts. Hence, the genomic location of virulence genes, within some species, may potentially be an identifying marker. An example of this is seen in *Leptosphaeria maculans*, where pathogenicity-related genes are found more frequently in adenine and thymine (AT)-rich regions, compared to GC-equilibrated regions, of the genome (Rouxel et al., 2011). It is thought that effector proteins and the host proteins that they target, co-evolve (Raffaele & Kamoun, 2012), making signs of variability, within the pathogen's genome, a potential indicator of regions coding for effector genes. A more comprehensive analysis of whole genome sequencing of multiple species and isolates may assist in identifying specific regions of high variation, potentially highlighting genes coding for proteins in direct contact with host proteins, and under pressure to evolve, providing a shortlisted set of genes for further investigation.

1.7 *DE NOVO* GENOME ASSEMBLY

WGS sequencing can provide a near complete view of a pathogen's coding instructions without requiring prior knowledge of genes, making it a superior method over sequencing only select genes. WGS sequencing of multiple organisms allows the application of comparative genomics to identify regions of interest, for example identification of candidate effectors in *Stagonospora nodorum* (Syme et al., 2013) and *Puccinia graminis* (Sperschneider et al., 2014), putative pathogenicity-related genes of *Colletotrichum* spp. (Buiate et al., 2017) and *Fusarium* spp. (Walkowiak et al., 2016), and evidence of two-speed genomes in *Zymoseptoria tritici* (syn. *Mycosphaerella graminicola*) (Stukenbrock et al., 2010) and *Nectria haematococca* (Coleman et al., 2009). The success of fungal genomics is well illustrated by projects such as Mycocosm (Grigoriev et al., 2014), which endeavour to sequence and analyse many fungi for public use. However, while it has proven to be an excellent method, challenges exist in the use of WGS sequencing.

Obtaining a good quality genome assembly can prove difficult from short reads. Repetitive regions, low coverage and loss during fragmentation can result in unresolved regions. The use of paired end and mate pair data can resolve some assembly issues by providing additional information on the orientation and distance between reads, while the use of long reads can help greatly in joining contigs into scaffolds (Ekblom & Wolf, 2014). However, due to the associated cost, many projects use only paired ends reads, which can provide a draft assembly suitable for gene prediction and small variant identification, but not for identification of structural variation. This limitation has been acknowledged, leading this study to focus on predicted genes and small variants, rather than structural variation. A limitation of the current study is that highly repetitive genomic regions may not be resolved during assembly, therefore pathogenicity-related genes contained in these regions will not be identified. The use of long read sequencing technology would likely allow better resolution of these repetitive regions.

An alternative to investigating an entire genome is sequencing only the transcriptome, however the benefits of a smaller scale of data need to be weighed against the knowledge that only genes being actively expressed at the time of sampling will be included. Genomic analyses not only have the potential to include most genes, but also intergenic regions allowing investigation of regulatory regions. As transcriptomic and genomic analyses complement each other, sequencing of the *Elsinoë* transcriptome is a likely future step.

1.8 THESIS AIMS

This study aimed to generate comprehensive genomic resources for multiple isolates of *Elsinoë* collected from Australia, with subsequent discovery and analysis of pathogenicity-associated genes, such as candidate effectors, SM and CWDE. As computational prediction of these pathogenicity-associated genes can produce an overabundance of candidates, genomic features and similarity to known virulence factors was used to prioritise candidates. The

work produced by this study provides the foundations for future experimental investigation of virulence genes for *Elsinoë* and mechanisms of pathogenesis. Additionally, candidate regions for use in differentiating the species and pathotypes were identified. This study focused on four isolates of *E. fawcettii*, eight isolates of *E. australis* JBS pathotype and one isolate of *E. australis* FL pathotype. Publicly available genomes used for comparative analysis included representatives of *E. fawcettii*, *E. australis*, *E. ampelina* and 10 fungal phytopathogens with known effectors.

Research objectives:

I.   To generate genomic resources for *E. fawcettii* and both the JBS and FL pathotypes of *E. australis*. A draft assembly for each isolate will be generated from Illumina WGS sequencing data, with subsequent gene prediction and annotation for downstream analyses. Genome assemblies will be submitted to GenBank, providing publicly available genomic resources.

II.   To identify candidate virulence genes including protein effectors, secondary metabolite gene clusters and cell wall degrading enzymes. Candidate virulence genes will be compared to known virulence genes of fungal pathogens for the purpose of prioritisation for future experimental validation.

III. To make genomic analyses and comparisons between species of *Elsinoë* and other fungal pathogen genomes. Core, accessory, unique and lineage-specific genes will be determined and analysed for potential association with host- or tissue-specificity. Candidate regions for use in future differentiation of species within *Elsinoë* will be proposed.

IV.   To investigate, within the species of *Elsinoë*, the potential for rapid molecular evolution in genomic regions associated with pathogenesis. Regions identified as having evidence of potential genome plasticity

will be cross referenced with the location of candidate pathogenicity-related genes.

CHAPTER 2: MATERIALS AND METHODS

2.1 MATERIALS

2.1.1 RAW SEQUENCE READS FOR *ELSINOË* SPECIES

Eight isolates of the *Elsinoë australis* JBS pathotype, collected from *Simmondsia chinensis* (jojoba) plantations in Hillston, Wagga Wagga and Forbes, NSW, Australia in 2005, were sequenced using Illumina paired end libraries at the Australian Genome Research Facility (AGRF) in 2017. Raw sequence reads for four of these isolates were available for the current study (Forbes-1, Hillston-2, Wagga-2 and Forbes-2), while for the remaining four isolates (Hillston-1, Wagga-1, Wagga-3 and Hillston-3) assembled genomes were available (detailed in section 2.1.2).

The first isolate of *E. fawcettii*, BRIP 53147a, isolated from *Citrus limon* (L.) Burm.f. in Queensland, Australia, was obtained from the Department of Agriculture and Fisheries (DAF) Biological Collections for the preparation fungal cultures. Paired end Illumina libraries were prepared and whole genome shotgun sequencing was performed at the Centre for Crop Health, University of Southern Queensland (USQ) in 2018.

Three additional isolates of *E. fawcettii*, BRIP 54245a, BRIP 54425a and BRIP 54434a, collected from *C. limon* 'Meyer', *C. reticulata* 'Clementine' and *C. limon* 'Rough' respectively, and one *E. australis* FL pathotype, BRIP 52616a, collected from *C. australasica*, were obtained from DAF Biological Collections. Fungal cultures and DNA extractions were prepared at the Centre for Crop Health, USQ. Paired end whole genome Illumina sequencing

was performed at the AGRF in 2019. Methods used for fungal culture and sequencing are detailed in Appendix A.

2.1.2 ADDITIONAL SEQUENCE INFORMATION

Four of the eight isolates of *Elsinoë australis* jojoba pathotype (Hillston-1, Wagga-1, Wagga-3 and Hillston-3) (section 2.1.1) were assembled by Dr. Ben Stodart at CSU; WGS sequencing projects available from GenBank (accessions: QGIJ00000000 (no. of scaffolds = 1,378), QGII00000000 (no. of scaffolds = 465), QGIH00000000 (no. of scaffolds = 438) and QGIG00000000 (no. of scaffolds = 603)).

Genome assemblies for *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* Arg-1 (accession GCA_007556505.1, no. of scaffolds = 21) (Shanmugam et al., 2020) and *E. ampelina* (accession GCA_005959805.1, no. of scaffolds = 13) (Li et al., 2020), were obtained from GenBank. Gene prediction was performed on these isolates as described in section 2.2.1.3.

Genome assemblies of further fungal species, and their predicted proteomes, were obtained from GenBank, including *Ustilago maydis* (accession GCF_000328475.2, no. of scaffolds = 27) (Kämper et al., 2006), *Leptosphaeria maculans* (accession GCF_000230375.1, no. of scaffolds = 76) (Rouxel et al., 2011), *Pyricularia oryzae* (syn. *Magnaporthe oryzae*) (accession GCF_000002495.2, no. of scaffolds = 53) (Dean et al., 2005), *Rhynchosporium commune* (accession GCA_900074885.1, no. of scaffolds = 164) (Penselin et al., 2016), *Verticillium dahliae* (accession GCF_000150675.1, no. of scaffolds = 55) (Klosterman et al., 2010), *Botrytis cinerea* (accession GCF_000143535.2, no. of scaffolds = 18) (Van Kan et al., 2017), *Parastagonospora nodorum* (accession GCF_000146915.1, no. of scaffolds = 108) (Hane et al., 2007), *Pyrenophora tritici-repentis* (accession

GCA_003231415.1, no. of scaffolds = 3964) (Moolhuijzen et al., 2018), *Sclerotinia sclerotiorum* (accession GCF_000146945.2, no. of scaffolds = 37) (Amselem et al., 2011) and *Zymoseptoria tritici* (accession GCA_900184115.1, no. of scaffolds = 20) (Plissonneau, Hartmann, & Croll, 2018). Predicted genes which overlapped TE regions were identified in each proteome and removed (sections 2.2.1.2 - 2.2.1.3) prior to analysis.

Experimentally validated effector protein sequences were retrieved from EffectorP 2.0 (Sperschneider et al., 2018).

2.2 METHODS

2.2.1 GENOMIC ANALYSES

2.2.1.1 *DE NOVO* GENOME ASSEMBLY

Data pre-processing, assembly and quality assessment were performed for *Elsinoë australis* JBS (Forbes-1, Hillston-2, Wagga-2 and Forbes-2) and *E. fawcettii* (BRIP 53147a) isolates on the Galaxy-Melbourne/GVL 4.0.0 webserver (Afgan et al., 2015) using the first version of tools mentioned. The same methods were performed on *E. fawcettii* (BRIP 54245a, BRIP 54425a and BRIP 54434a) and the *E. australis* FL pathotype (BRIP 52616a) using the later versions on the Galaxy-Australia webserver (version 19.05) (Afgan et al., 2018).

2.2.1.1.1 QUALITY CONTROL AND DATA PRE-PROCESSING

FastQC (version 0.11.5/0.11.8) (Andrews, 2010) was first used to check the quality of raw Illumina read sequences. Raw reads were subsequently trimmed using Trimmomatic (version 0.36) (Bolger, Lohse, & Usadel, 2014) for: Illumina adaptor sequences (TruSeq3); average base quality >Q20

across a sliding window with a width of 4 bases; leading and trailing base quality >Q20 and; a minimum read length after trimming of 31 base pairs (bp). The BRIP 52616a and BRIP 54425a isolates were reduced to a minimum read length of 101 bp and 115 bp, respectively to obtain forward and reverse read files of <10G each. Singleton reads were removed from paired-end datasets and retained. FastQC was used a second time to check the quality of processed reads.

2.2.1.1.2 ASSEMBLY

For the four *Elsinoë australis* JBS genomes (Forbes-1, Hillston-2, Wagga-2 and Forbes-2), *de novo* assembly was performed using Velvet (version 1.2.10) (Zerbino & Birney, 2008) and VelvetOptimiser (version 2.2.5) (Gladman, 2012) using paired-end and singleton reads. The k-mer length was tested between 71 and 97, in 2 bp increments and optimized for N50. Coverage cut-off was set between 0 and 0.8 and optimised for total number of base pairs in large contigs. Contigs > 200 bp in length were retained.

For *E. fawcettii* (BRIP 53147a), *de novo* assembly was performed in two steps. Firstly Velvet (v1.2.10) (Zerbino & Birney, 2008) and VelvetOptimiser (v2.2.5) (Gladman, 2012) were used, with input k-mer length range of 81-101, with a step size of 2, on the trimmed paired-end and singleton reads. Secondly, SPAdes (v3.11.1) (Bankevich et al., 2012) was run on the trimmed paired-end reads only with the following parameters: read error correction, careful correction, automatic k-mer values, automatic coverage cut-off and Velvet contigs (> 500 bp in length), from the previous step, included as trusted contigs. Contigs > 500 bp in length were retained.

The most recently sequenced *Elsinoë* genomes (BRIP 52616a, BRIP 54245a, BRIP 54425a and BRIP 54434a) were assembled *de novo* using SPAdes (version 3.12.0) (Bankevich et al., 2012), with the same parameters

mentioned above, however without the input of trusted contigs. Contigs > 500 bp in length were retained.

Velvet and VelvetOptimiser were tested at lower and higher k-mer length ranges, in addition SPAdes was tested for *de novo* assembly without "trusted" contigs. Optimal methods were chosen based on number of contigs and N50 values of assemblies.

2.2.1.1.3 QUALITY OF ASSEMBLIES

The genome assemblies were checked for completeness by using BUSCO (v2.0/v.3.02) (Simão et al., 2015) to search against the Ascomycota orthoDB (v9) dataset (Kriventseva et al., 2019), using an e-value of 0.01, to check the number of conserved single copy genes appearing within the assemblies. To assess coverage reads were mapped back to the assemblies using Bowtie2 (v2.2.4/v2.3.4.1) (Langmead & Salzberg, 2012) with default parameters. BAM files were sorted using Samtools (v0.1.18/v1.0) (Li et al., 2009). Duplicate reads were discarded using the MarkDuplicates tool, with stringency set to lenient, and coverage obtained using the CollectWGSMetrics tool, both from The Picard Toolkit (v2.7.1/v2.18.2.1) (http://broadinstitute.github.io/picard). The assemblies and mapped reads were visualised using the Integrated Genome Viewer (IGV) (v2.3.92) (Thorvaldsdóttir, Robinson, & Mesirov, 2013).

2.2.1.2 TRANSPOSABLE ELEMENTS

TE prediction, for all genomes, was performed on the GenSAS (v6.0) webserver (Lee et al., 2011). TE regions were predicted using RepeatMasker (v4.0.7) (Smit, Hubley, & Green, 2015) and Repbase (release 18.02) (Bao, Kojima, & Kohany, 2015), using the National Center for Biotechnology Information (NCBI) search engine and slow speed sensitivity.

2.2.1.3 GENE PREDICTION AND ANNOTATION

Gene prediction was performed at different times, with the genomes representing *Elsinoë australis* JBS pathotype being predicted first and hence the *E. australis* JBS pathotype Hillston-2 isolate submitted to GenBank was annotated using earlier software versions. Predicted genes used in all downstream analyses in this project, including the Hillston-2 isolate, were obtained using the more recent software versions detailed below.

Ab initio gene prediction was performed on unmasked genome assemblies on the GenSAS (v5.1/v6.0) webserver (Lee et al., 2011) using GeneMark-ES (v2.3e/v4.33) (Lomsadze et al., 2005). Incomplete genes and genes containing exons <15 bp in length were discarded. Genes which overlapped a TE region were removed from the predicted proteomes. 18s and 28s rRNA sequences were predicted using RNAmmer (v1.2) (Lagesen et al., 2007). tRNAscan-SE (v1.3.1/v2) (Lowe & Eddy, 1997) was used to predict transfer RNA (tRNA) sequences.

Annotation for *E. australis* JBS pathotype Hillston-2 and *E. fawcettii* BRIP 53147a was performed using BlastP (v2.7.1+) (Altschul et al., 1997) to query the predicted proteomes against the Swiss-Prot Ascomycota database (release 2017_11 and 2018_08) (The Uniprot Consortium, 2018) with an e-value of 1e-06, word size of 3 and minimum high-scoring segment pair (HSP) query coverage of 50%. Blast results were loaded into Blast2GO Basic (v4.1.9 and v5.2.1) (Conesa et al., 2005), InterProScan, mapping was performed to link blast hits to terms in the Gene Ontology database and annotation performed with default settings, except HSP-hit coverage cut off was set to 50% to increase stringency during annotation. Further annotation was achieved using HmmScan in HMMER (HMMerWeb v2.20.0 and command line v3.2.1) (Johnson, Eddy, & Portugaly, 2010) to query the

predicted proteome against the Protein Family Database (Pfam) (release 31 and 32) (Finn et al., 2016), using the gathering threshold cut off.

Annotations for the *E. australis* FL pathotype genome (BRIP 52616a) were transferred from the *E. australis* JBS pathotype Hillston-2 genome, while those for *E. fawcettii* (BRIP 54245a, BRIP 54425a and BRIP 54434a) were transferred from *E. fawcettii* BRIP 54317a, using Blast Reciprocal Best Hit (v0.1.11) with a minimum identity of 70% and minimum query coverage of 50%. Additionally, results from HMMER HmmScan (v3.2.1) search of all proteins from each genome against Pfam (release 32) was used to generate additional annotations.

Feature tables were created for the annotated genome assemblies (*E. australis* JBS pathotype Hillston-2 isolate and *E. fawcettii* BRIP 53147a), formatted as sequin files using the NCBI tbl2asn tool and submitted to GenBank.

All predicted proteins of each genome were checked for completeness using BUSCO (v3.0) (Simão et al., 2015).

2.2.1.4 PHYLOGENETIC ANALYSES

2.2.1.4.1 PHYLOGENETIC TREES I TO IV

ITS and partial TEF1-α, LSU and RPB2 sequences of all isolates of *Elsinoë* were obtained by searching the assemblies using blastn (Altschul et al., 1997) megablast (expect value of 1e-6), with those regions of 75 accessions of *Elsinoë* (GenBank accessions: KX886831-KX887304) (Fan et al., 2017) used as query sequences. Four phylogenetic analyses were conducted to show either species and/or pathotypes of *Elsinoë*; (I) found in Australia and New Zealand; (II) closely related to the *E. australis* JBS pathotype and *E. australis* FL pathotype; (III) closely related to isolates of *E. fawcettii*; and (IV) in relation to distantly related fungal pathogens included in the

comparative analysis. *Myriangium hispanicum* was included as an outgroup in trees I-III and *Spizellomyces punctatus* in tree IV. As all four genetic regions were not available for some species, trees I and III included the ITS and TEF1-α regions, tree II included ITS, TEF1-α, LSU and RPB2 regions and tree IV utilised partial TEF1-α and RPB2 regions. GenBank accessions for sequences included in each analysis are listed in Appendix B Tables 1-4. For each phylogenetic tree, sequences for each region were aligned using MUSCLE (Edgar, 2004) and manually trimmed to remove gaps. Alignments were concatenated and subjected to the "find the best model" tool in MEGA7 (Kumar, Stecher, & Tamura, 2016) to determine which maximum likelihood (ML) method was most suitable for each alignment.

The multi-locus alignments were used to perform maximum likelihood analyses in MEGA7 (Kumar et al., 2016). Trees I and III used the General Time Reversible (GRT+G+I) model (Nei, 2000) with partial deletion of 90% and 1000 bootstrap replicates. The initial trees for the maximum likelihood analysis were 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; tree I parameter = 0.5377; tree III parameter = 0.5132) was used and the rate variation model allowed some sites to be invariable (+*I*; tree I = 47.6797% sites; tree III = 45.6316% sites). For tree II, the multi-locus alignment was used to perform maximum likelihood analysis in MEGA7 (Kumar et al., 2016) using the Tamura-Nei (TN93+G+I) model (Tamura & Nei, 1993), with four discrete gamma categories (+G parameter = 0.4025), rate variation model allowing evolutionary invariability (+*I* = 35.8926% sites), partial deletion of 90% and 1000 bootstrap replicates. The initial tree for the maximum likelihood analysis was automatically selected using Neighbor-Join and BioNJ on the matrix of pairwise distances estimated using the Maximum Composite Likelihood method. For tree IV, the multi-locus alignment was used to perform maximum likelihood analysis in MEGA7 (Kumar et al., 2016) using the General Time Reversible (GTR+G) model (Nei, 2000) with partial deletion of 90% and 1000 bootstrap replicates. The initial tree for the 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 = 0.3965).

2.2.1.4.2 PHYLOGENETIC TREES V TO VII

The beta-tubulin (β-tubulin), Histone 4 and plasma membrane ATPase (*Pma1*) partial genes of all the isolates of *Elsinoë* were analysed in the current study and were utilised to differentiate the species and pathotypes. Sequences were obtained by searching the *Elsinoë* assemblies with a blastn (traditional blastn) search using query sequences obtained from GenBank (EU437545.1 *Elsinoë fawcettii* β-tubulin gene, X01611.1 *Neurospora crassa* Histone 4 and XM_951793.3 *N. crassa Pma1*). Three separate trees, one for each partial gene, were generated to illustrate the effectiveness of each individual region for identification purposes. For each analysis, sequences were aligned using MUSCLE (Edgar, 2004). Each alignment was used to perform maximum likelihood analyses in MEGA7 (Kumar et al., 2016) using the Tamura-Nei (TN93+G) model (Tamura & Nei, 1993) with complete deletion and 1000 bootstrap replicates, and with four discrete gamma categories (+G parameter = 0.1905 [β-tubulin, Tree V], 0.2157 [Histone 4, Tree VI] and 0.1869 [*Pma1*, Tree VII]). The initial tree for the maximum likelihood analysis was automatically selected using Neighbor-Join and BioNJ on the matrix of pairwise distances estimated using the Maximum Composite Likelihood method.

2.2.1.5 COMPARATIVE GENOMICS

Genomic features from all isolates of *Elsinoë* were identified and compared to those of the 10 comparative fungal genomes, as described below.

2.2.1.5.1 GENERAL GENOMIC FEATURES

Genome assembly size, number of predicted genes, median protein length and number of scaffolds were determined using CLC Sequence Viewer (version 7.8.1) (digitalinsights.qiagen.com).

2.2.1.5.2 GENE ORTHOLOGY ANALYSIS

Core, accessory, *Elsinoë*-specific and unique genes were determined using orthoMCL (Fischer et al., 2011) and ProteinOrtho (v6.0.14) (Steiner et al., 2011). Initially, proteins for each organism were grouped into orthologous groups using the orthoMCL algorithm run on the Eukaryotic Pathogen Database Resources (EuPathDB) webserver (v5.4) (Aurrecoechea et al., 2017). This allocated an orthoMCL group ID to proteins which were orthologous to proteins listed in the OrthoMCL DB (Release 5, 23 July 2015). Next, proteins from each isolate/species which did not obtain an orthoMCL ID were grouped together and run through orthoMCL for a second time to capture orthologs among the analysed species, but which did not obtain orthoMCL group ID's (labelled as "paralog groups" by the orthoMCL algorithm). Finally, any unclassified proteins, and those assigned to "no groups" by orthoMCL, were passed through ProteinOrtho (using the diamond algorithm for all-versus-all reciprocal best hit comparison, an e-value of 0.001 and a minimal algebraic connectivity of 0.1 for clustering) to establish any further potential orthologous relationships. Core genes were those shared by all species included in the comparison (and which appeared in at least seven of the eight *E. australis* JBS pathotype genomes), accessory genes were shared by at least two species, but not all, and unique genes were found in only one species. Additionally, the method described above was used to determine isolate-specific genes within the *E. fawcettii*, *E. australis* JBS pathotype and the *E. australis* pathotypes (SO, JBS and FL) examined.

2.2.1.5.3 GC CONTENT OF CODING DNA SEQUENCE

The GC% content of coding DNA sequence (CDS) of each gene was calculated using Bedtools nucBed (v2.27.1) (Quinlan, 2014). Mean, median, Q_1 and Q_3 values for each species were determined using R (R Core Team, 2018).

2.2.1.5.4 TRANSPOSABLE ELEMENTS

Results of the TE identification, described in section 2.2.1.2, were used to determine the distance between each gene and the closest TE region, using Bedtools closestBed (v2.27.1) (Quinlan, 2014).

2.2.1.5.5 REPETITIVE DNA

All genome assemblies utilised in this study were subjected to a search for Short Simple Repeats (SSR) using the Microsatellite Identification tool (MISA) (Beier, Thiel, Münch, Scholz, & Mascher, 2017). The SSR motif minimum length parameters were 10 for mono, 6 for di, and 5 for tri, tetra, penta and hexa motifs. This identified short repetitive regions within genes and in intergenic regions. Additionally, polyamino acid (polyAA) repeats of at least five consecutive amino acid residues were searched for in the predicted proteomes, using the FIMO motif search tool (Grant, Bailey, & Noble, 2011) within the Meme suite (v5.0.2) (Bailey et al., 2009).

2.2.1.5.6 PROTEIN FAMILY DATABASE ANALYSIS

The HMMER program (v3.2.1) (Johnson et al., 2010) was used to search all protein sequences against Pfam (release 32) (Finn et al., 2016). Firstly, a profile database of the PFam-A.seed file was generated, using hmmbuild,

and secondly binary indexes for the database were generated using hmmpress. Finally an hmmscan was used to perform the search against the PFam database, using the gathering threshold. The top hit was retained from the results for analysis.

2.2.1.5.7 AT-RICH REGIONS

The location of AT-rich regions in each genome assembly was determined using the OcculterCut program (v1.1) (Testa, Oliver, & Hane, 2016), developed specifically to investigate differences in GC% throughout fungal genomes. Firstly, mitochondrial contigs were removed from the assemblies, as these regions typically have a higher AT content. This was achieved using similarity to fungal mitochondrial genomes from NCBI (February 2015), for identification of mitochondrial contigs. Next, OcculterCut was run, using default parameters.

2.2.2 VIRULENCE GENE PREDICTION

2.2.2.1 PREDICTION OF SECONDARY METABOLITE GENE CLUSTERS

SM gene clusters were predicted from the genomes for all isolates of *Elsinoë* and comparative species by searching genome assembly files in conjunction with gene annotation files using antiSMASH Fungal using the Known Cluster Blast setting. For SM gene cluster prediction as a primary focus (section 3.2.1) the detection parameter was set to "strict" using version 5.1.2 (Blin et al., 2019), whereas when used during the known effector analysis (3.2.3.2) and candidate effector prioritisation (3.2.3.3) the detection parameter was set to "relaxed" using version 4.2.0 (Blin et al., 2017). This genome mining tool was used to predict SM biosynthetic gene clusters, SM type (e.g. type I polyketide synthase (T1PKS), terpene), potential function of some genes (e.g. core biosynthetic gene, transport-related gene, regulatory gene) and similarity to known SM gene clusters. Predicted SM gene clusters shared by

the species of *Elsinoë* was determined by analysing antiSMASH results of five *Elsinoë* isolates. The predicted SM genes of *E. fawcettii* (BRIP 53147a), the three *E. australis* pathotypes (SO, FL and JBS [Hillston-2 isolate]) and *E. ampelina* were analysed in an all-against-all BlastP search, using an evalue of 1e-6 and similarity >70%. Further identification of the elsinochrome gene cluster was performed using a BlastP search (evalue = 1e-6) of the predicted SM gene clusters against genes of the known elsinochrome gene cluster available on GenBank (Chung & Liao, 2008a).

2.2.2.2 SECRETOME PREDICTION

Secretome prediction was conducted on the proteomes of all genomes analysed in the current study. Proteins predicted as secreted by either SignalP (v4.1) (Petersen et al., 2011), Phobius (Käll, Krogh, & Sonnhammer, 2004) or ProtComp-AN (v6) (Softberry Inc, 2018) were collected for each species. These sets were run through both TMHMM (v2.0) (Krogh et al., 2001) and PredGPI (Pierleoni, Martelli, & Casadio, 2008) to identify proteins with likely TM helices and GPI-anchors, respectively. Proteins with more than one TM helix or with one TM helix after the first 60 amino acids, were discarded. Those proteins with one predicted TM helix within the first 60 amino acids were retained, as a signal peptide at the N-terminal can be falsely predicted as a TM helix by TMHMM. Those protein predicted to have "highly probable" or "probable" GPI anchors were also removed. All remaining proteins were retained as the predicted secretome.

2.2.2.3 EFFECTOME PREDICTION

2.2.2.3.1 EFFECTOR PREDICTION USING EFFECTORP

Candidate effector prediction was performed on the predicted secretomes using both versions of EffectorP (v1.0 and v2.0) (Sperschneider et al., 2018;

Sperschneider et al., 2016), a machine learning tool which has been trained on known fungal effector proteins.

2.2.2.3.2 KNOWN EFFECTOR ANALYSIS

The sequences of 42 experimentally validated effector proteins, obtained from EffectorP 2.0 (Sperschneider et al., 2018), were utilised in the known effector analysis. The existence of each of the 42 known effectors within their corresponding genome was validated by searching the known effector sequences against the available predicted proteomes of each species, using a cut off of >98% similarity. The 42 known effectors were also selected as they were correctly predicted as both secreted and as candidate effectors. The comparative species were chosen due to the presence of their experimentally validated effectors and the availability of whole genome sequencing. Analyses, detailed further below, included gene density, genomic location, GC% content, polyAA repeats, gene orthology classification, protein length and predicted involvement in SM gene clusters. A comparative analysis between the results of the 42 known effectors and the results of all proteins from the 10 comparative species, was performed. This known effector analysis method was published as part of the current project (Jeffress et al., 2020).

2.2.2.3.2.1 GENE DENSITY

The intergenic flanking region (IFR) was calculated both upstream and downstream of each predicted gene, as the number of nucleotides to the next adjacent gene. Genes in gene-dense locations were identified as those with IFR < 1,500 bp both upstream and downstream. Those genes on the edge of a contig were excluded from this group. Genes in gene-sparse locations were identified as those with IFR > 1,500 bp both upstream and downstream.

2.2.2.3.2.2 GENOMIC LOCATION

Genomic location was analysed by determining the distance in base pairs between genes and nearby genomic features, including closest TE region, AT-rich region and SSR using Bedtools closestBed (v2.27.1) (Quinlan, 2014).

2.2.2.3.2.3 ADDITIONAL FEATURES

Additional features in the known effector analysis identified were determined in previous sections, including GC% content of coding DNA sequence (CDS) (section 2.2.1.5.3), involvement in predictable SM gene clusters (section 2.2.2.1), gene orthology classification (section 2.2.1.5.2), protein length (section 2.2.1.5.1), and polyAA-containing proteins (section 2.2.1.5.5).

2.2.2.3.3 CANDIDATE EFFECTOR PRIORITISATION

Prioritisation of candidate effectors was performed using a points-based system optimised during the known effector analysis. All candidate effectors had the opportunity to be allocated at least four points. One point was allocated to a candidate effector for each of the following factors: if it was not in a gene-dense location, if it was not involved in any predicted SM gene clusters, if it was either species-specific or obtained a hit to the same orthoMCL ID as a known effector, and lastly if the GC% of the CDS was either $<Q_1$ and $>Q_3$ for the specific species. An extra point was allocated to species with genome assemblies containing >2% TE coverage and another point for those with >25% AT-rich region coverage. This provided the means to score all candidate effectors out of n (four, five or six) points. Finally,

prioritised candidate effectors were those which were allocated a total of *n* or *n*-1 points.

2.2.2.3.4 ANALYSIS OF PRIORITISED CANDIDATE EFFECTORS

Prioritised candidate effectors were further analysed for the following features. Cysteine content was analysed used CLC Viewer 7.8.1 (digitalinsights.qiagen.com). Candidate effectors targeted to the host chloroplast were predicted using TargetP 2.0 server (Armenteros et al., 2019), using the plant parameter. Candidate effectors were queried against the Pathogen Host Interactions Database (PHI-base) (v4.7) (Urban et al., 2017) in a BlastP (v2.7.1) (Altschul et al., 1997) search, using an e-value cut off of 1e-06, a query coverage HSP of 70% and a minimum of 40% similarity.

2.2.2.4 PREDICTION OF CELL WALL DEGRADING ENZYMES

Prediction of CWDE was performed on the predicted proteomes using the dbCAN2 meta server (v9.0) (Zhang et al., 2018), three tools were selected including HMMER to search the dbCAN HMM database (Yin et al., 2012), Diamond (Buchfink, Xie, & Huson, 2014) search against the Carbohydrate-Active enZYmes (CAZy) database (Lombard et al., 2014) and Hotpep query against the Peptide Pattern Recognition library (Busk et al., 2017). Proteins with a positive result for at least two of the three tools were labelled as predicted CWDE. The predicted proteomes were also searched against PHI-base (v4.7) (Urban et al., 2017), to identify potential pathogenesis-related proteins, using BlastP (v2.7.1) (Altschul et al., 1997) analyses with an e-value of 1e-06 and a query coverage HSP of 70%, those results with >40% similarity were retained. Predicted CWDE which; (I) were predicted as secreted; (II) obtained hits to virulence-related genes of plant pathogens in PHI-base which showed evidence of reduced virulence or loss of pathogenicity in knockout or mutant experiments; and (III) were orthologs to

other candidate CWDE among all five species of *Elsinoë* (*E. australis* JBS pathotype Hillston-2, *E. australis* FL pathotype, *E. australis* SO pathotype, *E. fawcettii* BRIP 53147a and *E. ampelina*) were labelled as prioritised CWDE.

2.2.3 SEARCH FOR EVIDENCE OF A TWO-SPEED GENOME WITHIN *ELSINOË* SPP.

In order to search for potential two-speed genomes among the *Elsinoë* spp., genomic features associated with genome plasticity were determined and their location analysed in respect to the location of predicted virulence genes.

2.2.3.1 ANALYSIS OF FEATURES ASSOCIATED WITH GENOME PLASTICITY

2.2.3.1.1 GENE-SPARSE REGIONS AND REGIONS DEVOID OF FUNGAL CORE GENES

Genes located in gene-sparse locations were determined as described in section 2.2.2.3.1.1. Additionally, results from the BUSCO (Simão et al., 2015) analysis in section 2.2.1.3, were used to identify contigs that did not contain orthologs to genes in the OrthoDB fungal database (release 9), denoting fragments of the genome potentially devoid of fungal core genes.

2.2.3.1.2 TRANSPOSABLE ELEMENT-ASSOCIATED REGIONS

The genomic location of TE regions was predicted using RepeatMasker (Smit et al., 2015), as described in section 2.2.1.2. Distance from gene models to the closest TE region was determined using Bedtools closestBed (v2.27.1) (Quinlan, 2014).

2.2.3.1.3 SHORT SIMPLE REPEATS

Repetitive regions, specifically SSR throughout the genome and polyAA repeats within CDS, were identified, as described in section 2.2.1.5.5.

2.2.3.1.4 AT-RICH REGIONS

AT-rich regions were identified as described in section 2.2.1.5.7 and labelled as repeat-induced-point-mutation-like (RIP-like) regions,

2.2.3.1.5 FRAGMENTED GENOMIC REGIONS

Scaffold length was determined using Bedtools2 (Quinlan, 2014), those shorter than 20 kilobases (Kb) were labelled as highly fragmented.

2.2.3.1.6 SNP DETECTION

In the case of *Elsinoë fawcettii*, SNP detection was determined by mapping reads of isolates BRIP 54245a, BRIP 54425a and BRIP 54434a back to BRIP 53147a. While for *E. australis* JBS pathotype, reads of other JBS pathotype isolates (Forbes-1, Wagga-2 and Forbes-2) were mapped back to the Hillston-2 isolate. Mapping was performed on the Galaxy-Australia webserver (Afgan et al., 2018) using Bowtie2 (v2.3.4.3) (Langmead & Salzberg, 2012) and SAMtools (v1.9) (Li et al., 2009) to align trimmed reads to the appropriate assemblies. A BAM file was produced for each mapped isolate and alignments corrected using Freebayes BamLeftAlign tool (v1.1.0.46) (Garrison & Marth, 2012). BAM files were checked for coverage using The Picard Toolkit CollectWgsMetrics tool (http://broadinstitute.github.io/picard). Variant calling was performed using

the Freebayes Bayesian genetic variant detector (v1.1.0.46) (Garrison &
Marth, 2012), using population model options to set ploidy to 1, allelic scope
options to search for only SNP and indels, and setting both the minimum
mapping quality and the minimum base quality to 20. All other options were
set as default. The variant calling file was filtered using VCFfilter (v1.7)
(Garrison, 2015), to retain variants with a depth of >10.

2.2.3.2 CANDIDATE VIRULENCE GENE LOCATION

Genomic location of candidate virulence genes (predicted CWDE, SM gene
clusters, secreted proteins, candidate effectors and *Elsinoë*-specific genes)
in relation to the location of genomic features potentially associated with
plasticity (gene-sparse regions, TE regions, AT-rich regions, SSR, polyAA
repeats, scaffold edges, short scaffolds and SNP) were analysed by
determining the distance in base pairs between these features using
Bedtools closestBed (v2.27.1) (Quinlan, 2014).

CHAPTER 3: RESULTS

3.1 GENOMIC ANALYSES

3.1.1 ASSEMBLY AND GENE PREDICTION

In this study, genomes were assembled for the JBS (Forbes-1, Hillston-2, Wagga-2 and Forbes-2) and FL pathotypes (BRIP 52616a) of *E. australis*, and four isolates of *E. fawcetii* (BRIP 53147a, BRIP 54245a, BRIP 54425a and BRIP 54434a). The assembly size of the four JBS pathotypes were 26.6 megabases (Mb), the FL pathotype obtained the smallest assembly size at 23.9 Mb, while for *E. fawcettii* assemblies ranged from 25.9 to 26.5 Mb (Table 1). The assembled genomes for the JBS pathotypes (Forbes-1, Hillston-2, Wagga-2 and Forbes-2) consisted of 637, 484, 466 and 495 scaffolds (or 1,593, 1,483, 1,425 and 1,453 contigs), respectively, greater than 200 bp in length. These genomes were each sequenced to an average of 165x coverage, the N50 of each assembly ranged between 347,413 and 444,387 bp and all had an overall GC content of 50.9%. Unknown bases (N's) were allowed in these earlier assemblies, however in all further assemblies a more conservative approach was taken with no unknown bases. The FL pathotype (BRIP 52616a) contained 61 contigs greater than 500 bp in length, had a coverage of 241x, an N50 of 947,746 bp and an overall GC content of 52.3%. The assemblies of the genomes for *E. fawcettii* (BRIP 53147a, BRIP 54245a, BRIP 54425a and BRIP 54434a) consisted of 286, 475, 474 and 473 contigs, respectively, greater than 500 bp in length, with coverage between 193x and 237x. The N50 ranged from 608,276 to 694,004 bp and all four assemblies had an overall GC content of 52.3%. Table 1 indicates the assemblies had a reasonably low degree of fragmentation. All nine genomes of *Elsinoë* showed low coverage of transposable elements, ranging between 0.32% and 1.35%.

Table 1. General genome features of isolates of *Elsinoë* sequenced in the current study.

Species	*E. australis*					*E. fawcettii*			
Isolate	Jojoba Black Scab (JBS) Forbes-1	JBS Hillston-2	JBS Wagga-2	JBS Forbes-2	Finger Lime BRIP 52616a	BRIP 53147a	BRIP 54245a	BRIP 54425a	BRIP 54434a
GenBank accession	GCA_005382415.1	GCA_005382405.1	GCA_005382395.1	GCA_005382375.1	GCA_016625675.1	GCA_012977835.1	GCA_016625505.1	GCA_016625515.1	GCA_016625525.1
Short Read Archive (SRA) accession	SRX8327457	SRX8334527	SRX8334860	SRX8340264	SRX7939848	SRX5307824	SRX7945230	SRX7950446	SRX7956662
Assembly length (bp)	26,621,968	26,626,433	26,604,401	26,617,985	23,866,517	26,011,141	26,483,858	25,855,480	25,873,468
Number of N's in assembly (bp)	76,743	82,087	53,790	71,592	0	0	0	0	0
Coverage	165x	165x	163x	169x	241x	193x	218x	235x	237x
Number of scaffolds	637	484	466	495	61	286	475	474	473
Number of contigs	1,593	1,483	1,425	1,453	61	286	475	474	473
Mean GC content (%)	50.9	50.9	50.9	50.9	52.3	52.3	52.3	52.3	52.3
N50 (bp)	406,158	444,387	347,413	428,165	947,746	694,004	668,319	646,472	608,276

Mean contig length (bp)	41,792	55,013	57,090	53,773	391,254	90,948	55,755	54,547	54,700
Minimum contig length (bp)	200	200	200	200	509	501	501	501	501
Maximum contig length (bp)	1,109,786	1,062,546	1,099,042	1,182,266	2,452,879	2,345,732	2,540,363	1,609,497	1,980,929
Transposable elements (bp)	342,744	355,201	359,987	350,690	102,110	95,654	96,658	82,437	82,565
Transposable element coverage (%)	1.29	1.33	1.35	1.32	0.43	0.37	0.36	0.32	0.32
# of predicted genes (not including genes overlapping predicted TE regions)	9,465	9,533	9,482	9,488	9,215	10,080	10,164	10,014	10,016
Complete and single copy BUSCO's out of a possible 290 (%)	282 (97.2%)	283 (97.6%)	282 (97.2%)	283 (97.6%)	281 (96.9%)	286 (98.6%)	278 (95.9%)	278 (95.9%)	278 (95.9%)

Gene prediction for each genome was performed using GeneMark-ES (Lomsadze et al., 2005) on all *Elsinoë* isolates. All results reported here were obtained using the latest software versions, while those submitted to GenBank for *E. australis* JBS pathotype Hillston-2 were obtained using previous software versions. Genes which overlapped a TE region were removed. The predicted proteomes of the *E. australis* JBS pathotypes each contained between 9,465 and 9,533 proteins (Table 1). A total of 9,215 proteins were predicted for the FL pathotype, while a higher number of predicted proteins were found among the genomes of *E. fawcettii*, which ranged from 10,014 to 10,164 proteins. All assemblies showed a high degree of completeness within the CDS, with 95.9% - 98.6% of complete single copy Ascomycota genes appearing within the assemblies.

For *E. australis* JBS pathotype Hillston-2 and *E. fawcettii* BRIP 53147a, a total of 5,557 and 5,636 genes were annotated, respectively, using the Swiss-Prot Ascomycota database and the Pfam database as references. A total of nine draft assemblies for *Elsinoë*, two of which were annotated, were submitted to GenBank and all raw reads submitted to the Short Read Archive (SRA) (accessions detailed in Table 1).

3.1.2 PHYLOGENETIC ANALYSIS

3.1.2.1 AUSTRALIAN AND NEW ZEALAND SPECIES OF *ELSINOË*

ITS and TEF1-α regions, of all newly sequenced species of *Elsinoë* and isolates originating from Australia and New Zealand (GenBank accessions detailed in Appendix B Table 1), were compared by phylogenetic analysis. *Myriangium hispanicum* was included as the outgroup. The two regions concatenated formed a total of 877 positions. Over the two regions, the eight *E. australis* JBS pathotypes were identical to each other, aside from two substitutions in the TEF1-α region, specifically cytosine (C) - thymine (T) at site 233 and adenine (A) – cytosine (C) at site 239, in the Hillston-3 isolate. The *E. australis* FL pathotype differed slightly from the JBS pathotypes with

one guanine (G) insertion at 119bp and one C deletion at 414bp in the ITS region and three T-C substitutions at sites 5, 71 and 134. These two pathotypes formed a clade with *E. piri*. The isolates of *E. fawcettii*, with 100% identity in these two regions, sat in a separate clade to the *E. australis* pathotypes, with *E. tiliae* as their closest relative. All Australian isolates of *E. fawcettii* showed 90.75 % identity to the *E. australis* JBS pathotype Hillston-2 isolate over the ITS and TEF1-α regions.
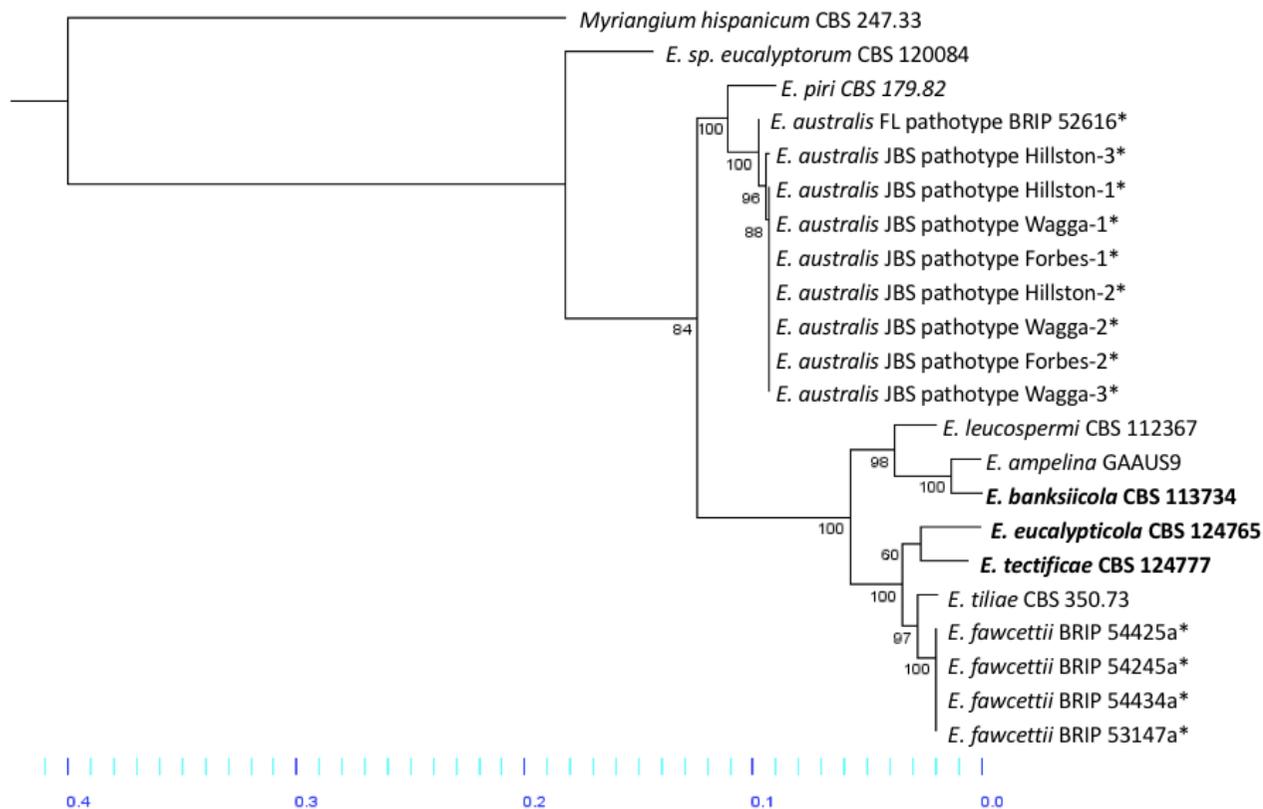
Figure 2. Tree I: Maximum likelihood analysis of the concatenated ITS and partial TEF1-α regions of all newly sequenced isolates with *Elsinoë* spp. collected in Australia and New Zealand. *Myriangium hispanicum* was included as an outgroup. The branch length is the number of nucleotide substitutions per site, bootstrap values are included, type strains are in bold and new isolates described in the current study denoted with an asterisk (*).

3.1.2.2 *E. AUSTRALIS* PATHOTYPES WITH CLOSELY RELATED SPECIES

Sequences for the eight isolates of *Elsinoë australis* JBS pathotype, the *E. australis* FL pathotype and closely related *Elsinoë* species (>98% similarity in the ITS, TEF1-α, LSU or RPB2 region) were selected for this analysis (GenBank sequences available in Appendix B Table 2). *Myriangium hispanicum* was selected as the outgroup. The maximum likelihood phylogeny displayed (Figure 3) depicts a total of 2,364 positions. The ITS, TEF-1α, LSU and RPB2 sequences for seven of the eight isolates (Hillston-1, Hillston-2, Wagga-1, Wagga-2, Wagga-3, Forbes-1 and Forbes-2) showed 100% identity to one another. The eighth isolate (Hillston-3) showed two base changes in the TEF1-α region, described above, and a further five base changes in the RPB2 region, specifically A-G (site 74), C-T (site 230), T-C (sites 525 and 666) and G-A (site 731).

This analysis indicated the *E. australis* JBS pathotype is more closely related, but not identical, to *E. punicae* (4-7/2364 base changes) and *E. genipae-americanae* (21-22/2364 base changes), than to the *E. australis* SO pathotype (23-48/2364 base changes). The *E. australis* FL pathotype showed 15-16/2364 base changes when compared to the JBS pathotype.
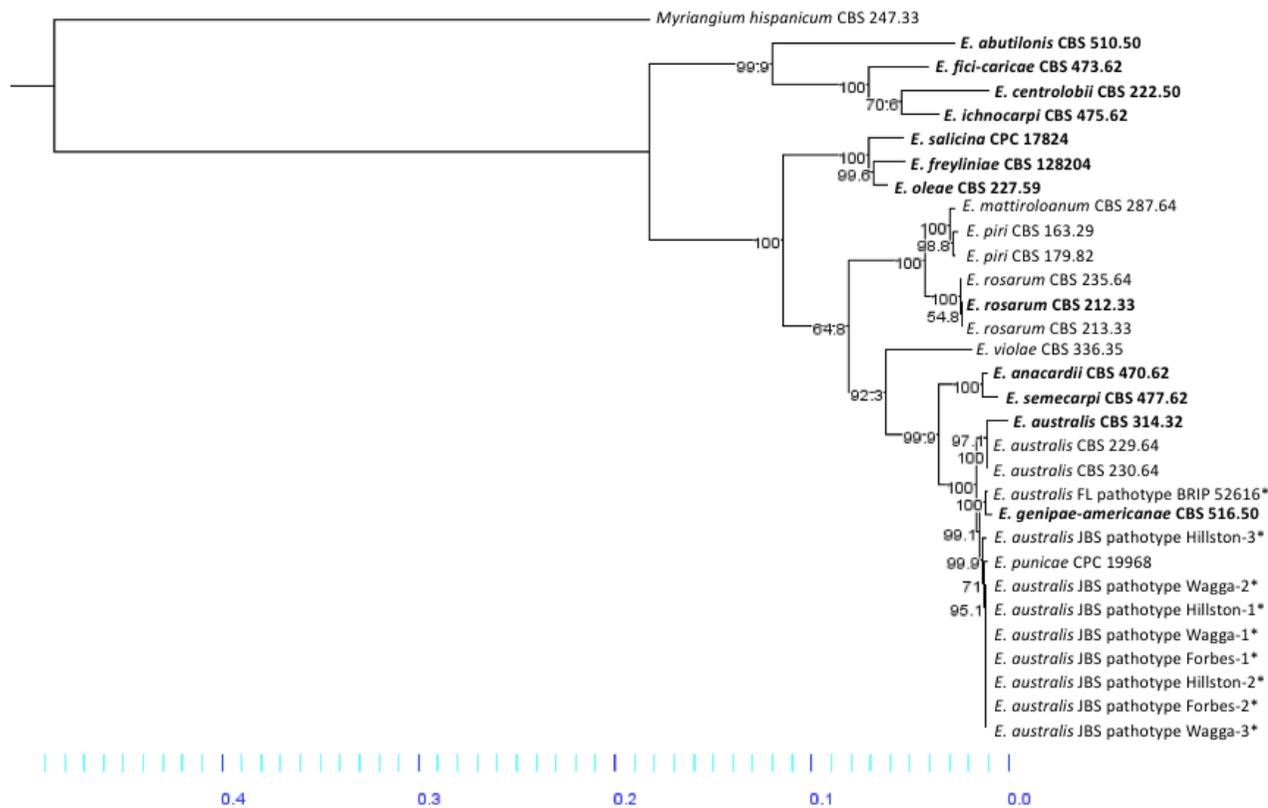
Figure 3. Tree II: Maximum likelihood phylogenetic tree of _Elsinoë_ spp. closely related to the _E. australis_ JBS and FL pathotypes. The phylogenetic tree was inferred from a concatenated dataset including ITS and partial TEF1-α, LSU and RPB2 regions. _Myriangium hispanicum_ was included as an outgroup. The branch length is the number of nucleotide substitutions per site, bootstrap values are included, new isolates described in the current study denoted with an asterisk (*) and type strains are in bold.

3.1.2.3 *E. FAWCETTII* PATHOTYPES WITH CLOSELY RELATED SPECIES

Phylogenetic analysis of ITS and partial TEF1-α regions of the newly sequenced isolates of *Elsinoë fawcettii* (BRIP 53147a, BRIP 54245a, BRIP 54425a and BRIP 54434a) in comparison with other *E. fawcettii* and closely related *Elsinoë* species (Figure 4) was generated (GenBank sequences available in Appendix B Table 3). The analysis indicated both sequence regions derived from all *E. fawcettii* analysed, aside from the Jingeul pathotype isolates, were identical. *E. citricola* was the most closely related species to *E. fawcettii*, with identical ITS sequences and only two base changes in the TEF1-α region.
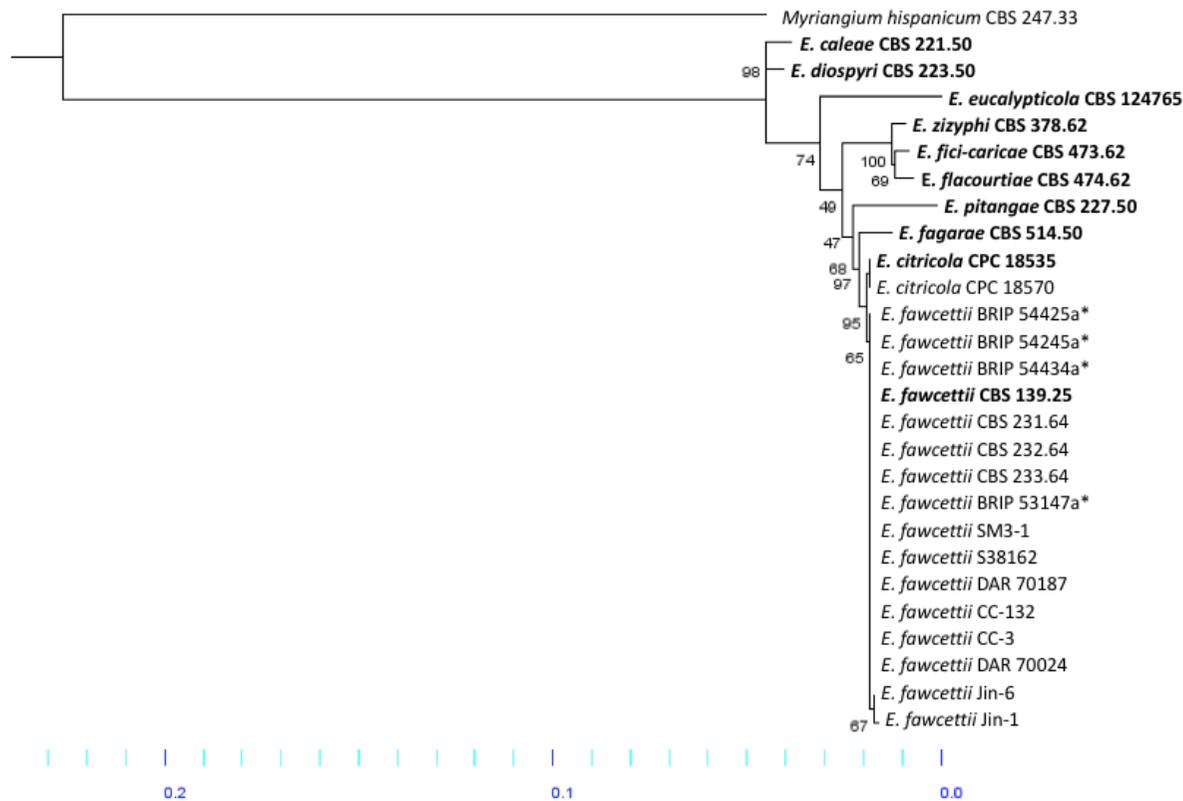
Figure 4. Tree III: Maximum likelihood phylogenetic tree of *E. fawcettii* and other *Elsinoë* spp. A concatenated dataset, including the ITS and partial TEF1-α regions was utilised in the analysis. *Myriangium hispanicum* was the outgroup. The branch length illustrates the number of nucleotide substitutions per site, bootstrap values are indicated, type strains are in bold and isolates sequenced in the current study are indicated with an asterisk (*).

3.1.2.4 *ELSINOË* SPECIES WITH TEN COMPARATIVE FUNGAL PATHOGENS

Maximum likelihood analysis of partial TEF1α and RPB2 regions of newly sequenced *Elsinoë fawcettii* BRIP 53147a, *E. australis* FL pathotype BRIP 52616a and *E. australis* JBS pathotype Hillston-2 in comparison with recently published *E. fawcettii*, *E. australis* SO pathotype and *E. ampelina*, and distant fungal pathogens included in the comparative analyses (Figure 5) was generated. GenBank sequences are detailed in Appendix B Table 4. The analysis indicated a distinct clade capturing all species of *Elsinoë*, consisting of three distinguishable groups, namely the *E. fawcettii* group, the *E. australis* group and *E. ampelina* alone. The closest relative to the *Elsinoë* clade, among the compared species, was *Zymoseptoria tritici*. While *Ustilago maydis*, being the only Basidiomycetes, among the pool of Ascomycetes, sat in a clade of its own, away from all comparative pathogens. *Spizellomyces punctatus* was included as the outgroup.
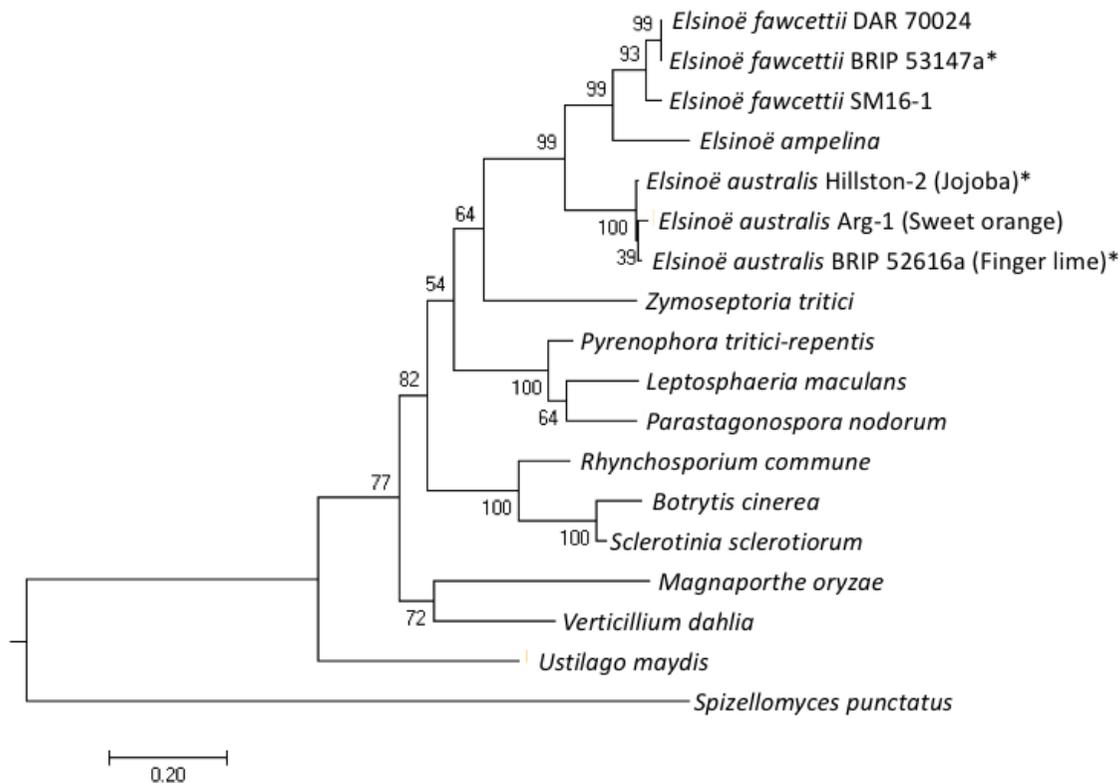
Figure 5. Tree IV: Maximum likelihood phylogenetic tree of *Elsinoë* spp. with other fungal pathogen species included in the comparative study. The analysis utilised a concatenated dataset including partial TEF1α and RPB2 regions. *Spizellomyces punctatus* was the outgroup. The branch length illustrates the number of nucleotide substitutions per site, bootstrap values are shown, and isolates sequenced in the current study are denoted with an asterisk (*).

3.1.2.5 INDIVIDUAL REGIONS FOR DIFFERENTIATION AMONG SPECIES AND PATHOTYPES OF *ELSINOË*

Three maximum likelihood phylogenetic analyses of species of *Elsinoë*, which have sequenced genomes available, were conducted, highlighting individual partial gene regions which have potential to be used as identifiers to differentiate species and pathotypes. Both the β-tubulin (405 positions) (Figure 6) and Histone 4 (299 positions) (Figure 7) regions showed identity within the two groups with multiple isolates; being 100% identity among those of *E. fawcettii*; and, 100% identity among those of the *E. australis* JBS pathotype. While differences were seen separating the three *E. australis* pathotypes (JBS, FL and SO) and again, separating the three species (*E. australis*, *E. fawcettii* and *E. ampelina*). This was a similar case in the third partial gene region, plasma membrane ATPase (*Pma1*) (1486 positions) (Figure 8), except for two base substitutions in *E. fawcettii* SM16-1 compared to all other isolates of *E. fawcettii*. *Pma1* was retained and illustrated as a potential genomic region for differentiation, as the isolate SM16-1 was the only *E. fawcettii* isolate in the study identified as the FBHR pathotype from Korea, while all other isolates of *E. fawcettii* were collected in Australia. This highlights the *Pma1* region as a good candidate for further investigation as an identifying region, not only of *Elsinoë* spp., but also potentially the different *E. fawcettii* pathotypes. *Alternaria alternata* MLL3 was included as an outgroup in all three analyses. GenBank sequences for regions utilised in Figures 6, 7 and 8 are detailed in Appendix B Table 5.
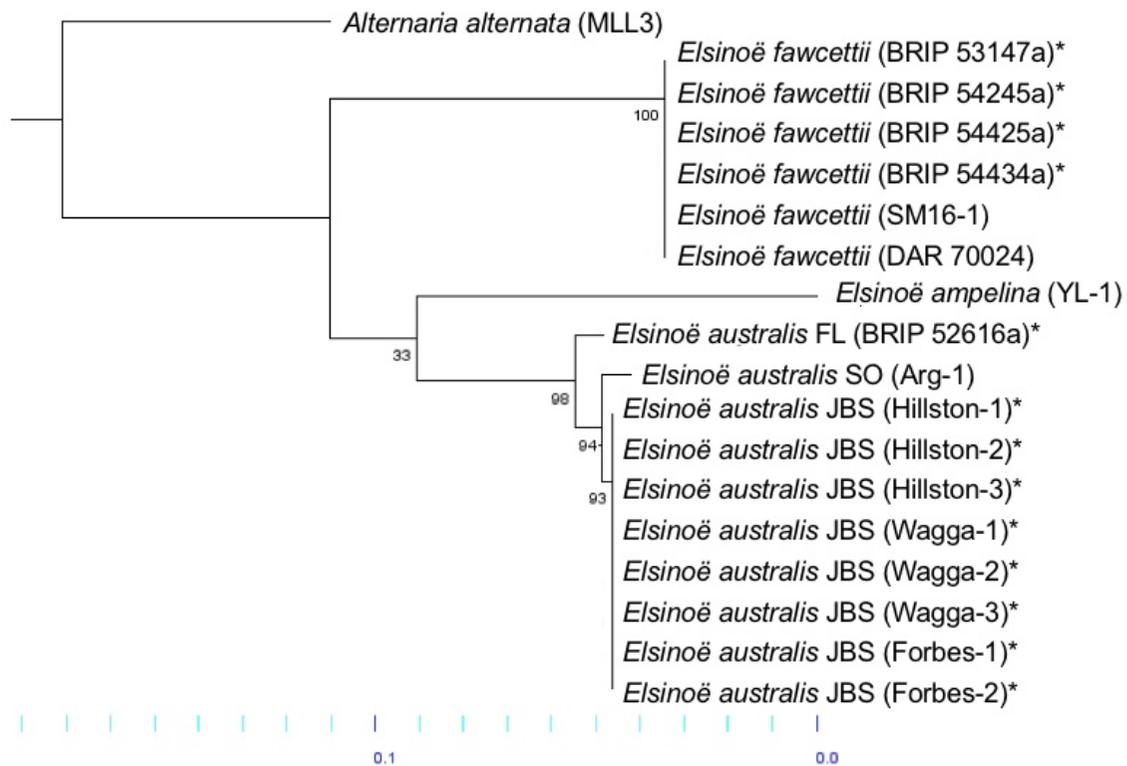
Figure 6. Tree V: Maximum likelihood phylogenetic tree for isolates of *Elsinoë*, which have available genome sequencing, using the partial β-tubulin gene region. *Alternaria alternata* was the outgroup. The branch length indicates the number of nucleotide substitutions per site and isolates sequenced in the current study are denoted with an asterisk (*).
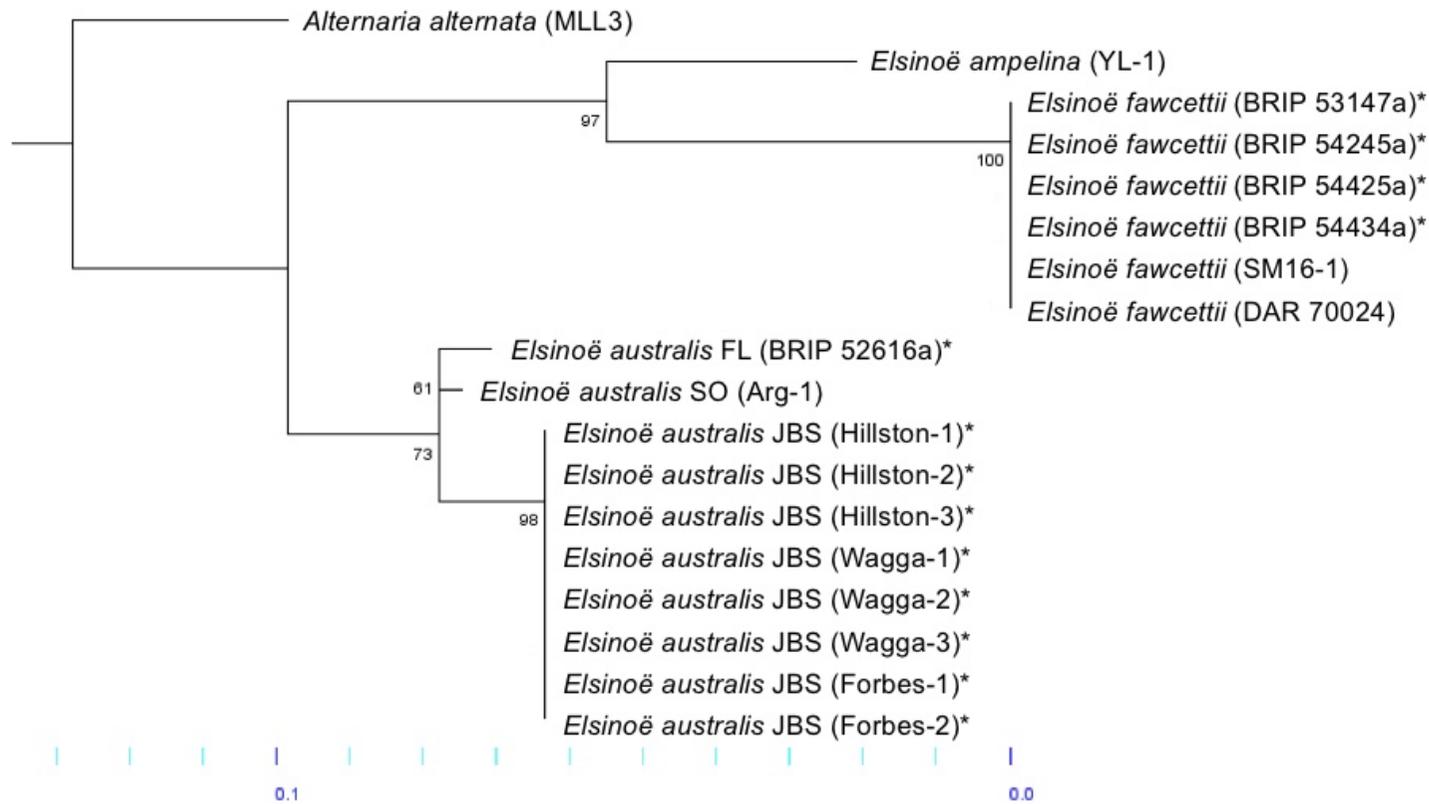
Figure 7. Tree VI: Maximum likelihood phylogenetic tree for isolates of *Elsinoë*, which have available genome sequencing, using the partial Histone 4 gene region. *Alternaria alternata* was the outgroup. The branch length indicates the number of nucleotide substitutions per site and isolates sequenced in the current study are denoted with an asterisk (*).
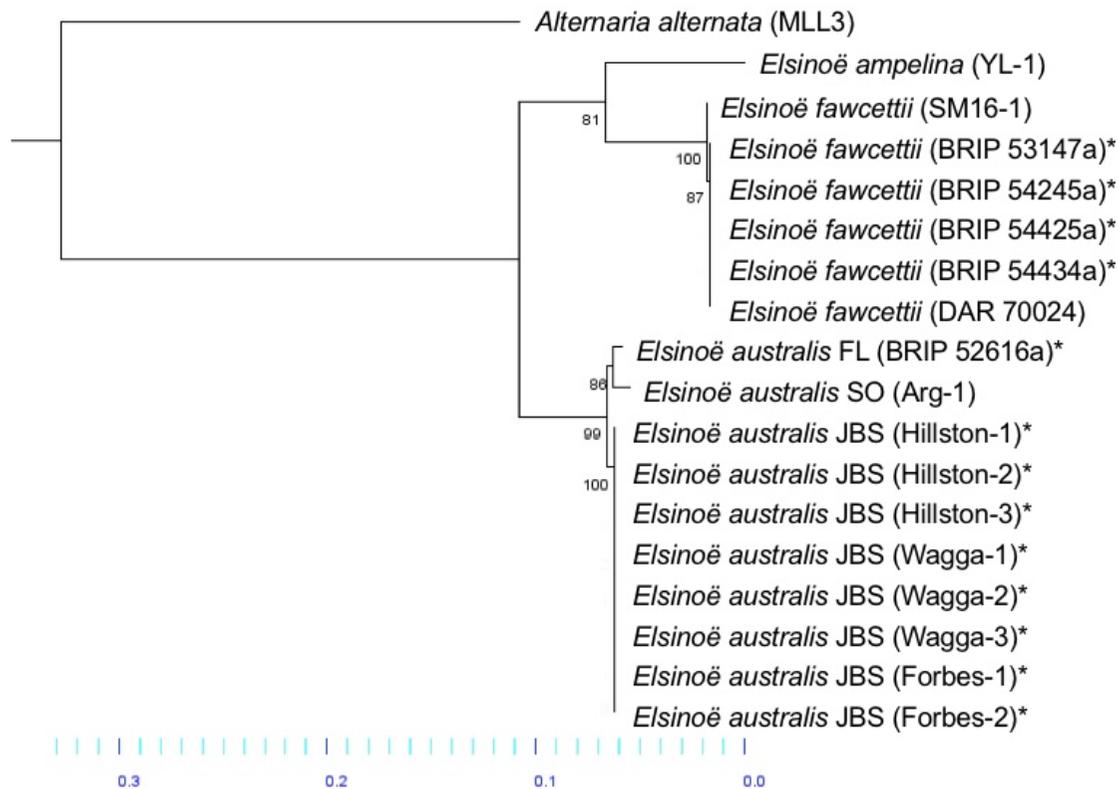
Figure 8. Tree VII: Maximum likelihood phylogenetic tree for isolates of *Elsinoë*, which have available genome sequencing, using the partial plasma membrane ATPase (*Pma1*) gene region. *Alternaria alternata* was the outgroup. The branch length indicates the number of substitutions per site and isolates sequenced in the current study are denoted with an asterisk (*).

3.1.3 COMPARATIVE GENOMICS: *ELSINOË* SPP. AND ISOLATES IN COMPARISON WITH TEN FUNGAL PATHOGENS

3.1.3.1 GENERAL GENOMIC FEATURES

General genomic features for the *Elsinoë fawcettii*, *E. australis* FL pathotype and *E. australis* JBS pathotype, sequenced in the current study, were compared to the *E. australis* SO pathotype, two isolates of *E. fawcettii* (DAR 70024 and SM16-1) and ten additional genomes of fungal pathogens including *Botrytis cinerea*, *Leptosphaeria maculans*, *Pyricularia oryzae* (syn. *Magnaporthe oryzae*), *Parastagonospora nodorum*, *Pyrenophora tritici-repentis*, *Rhynchosporium commune*, *Verticillium dahliae*, *Sclerotinia sclerotiorum*, *Ustilago maydis* and *Zymoseptoria tritici*. The assembly sizes of the species of *Elsinoë*, sequenced in the current study, ranged from 23.8 Mb (*E. australis* FL pathotype) to 27.1 Mb (*E. australis* JBS pathotype Hillston-3) and were comparable in size to the published assemblies of *E. australis* SO pathotype (Arg-1) (23.8 Mb), *E. ampelina* (28.3 Mb) and two isolates of *E. fawcettii* (DAR 70024 and SM16-1, 26.3 Mb and 26.7 Mb, respectively) (Figure 9). The assemblies for the species of *Elsinoë* were smaller than each of the compared species, except for *U. maydis* at 19.7 Mb. Similarly, *U. maydis* had the smallest number of predicted proteins (6,692), while the species and isolates of *Elsinoë* ranged from 8,962 (*E. australis* JBS pathotype Hillston-1) to 10,519 (*E. fawcettii* SM16-1) predicted proteins (Figure 10). Thus, the species with the most comparable proteome sizes to *Elsinoë* were *V. dahliae*, with 10,441 predicted proteins and secondly, *Pyrenophora tritici-repentis* with 10,772. The isolates of *Elsinoë* had a mean protein length ranging between 404 amino acids (AA) (*E. australis* JBS pathotype Hillston-1) and 419 AA (*E. australis* FL pathotype), which most closely aligned with *V. dahliae* (395 AA) and *B. cinerea* (405 AA) (Figure 11). Analysing the degree of genome fragmentation (Figure 12) indicated that the assemblies of the *Elsinoë* genomes, sequenced in the current study, are fairly fragmented (ranging between 61 and 1,378 scaffolds) in comparison to the majority of the comparative species. However, assemblies of *Pyrenophora tritici-repentis*

and *L. maculans* both contained high numbers of scaffolds; 3,964 and 1,717, respectively. Full data for the generation of Figures 9-12 is available in Appendix C Table 1.
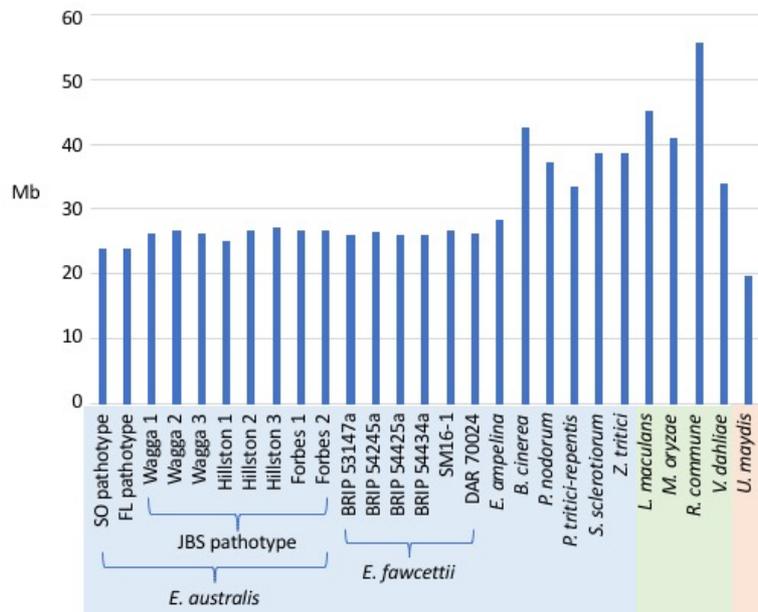
Figure 9. Genome assembly size comparison (Mb) among newly sequenced *Elsinoë* genomes, additional *Elsinoë* spp. and further fungal pathogens. Isolates/species on x-axis are grouped into necrotrophic (blue), hemi-biotrophic (green) and biotrophic (red) pathogens.
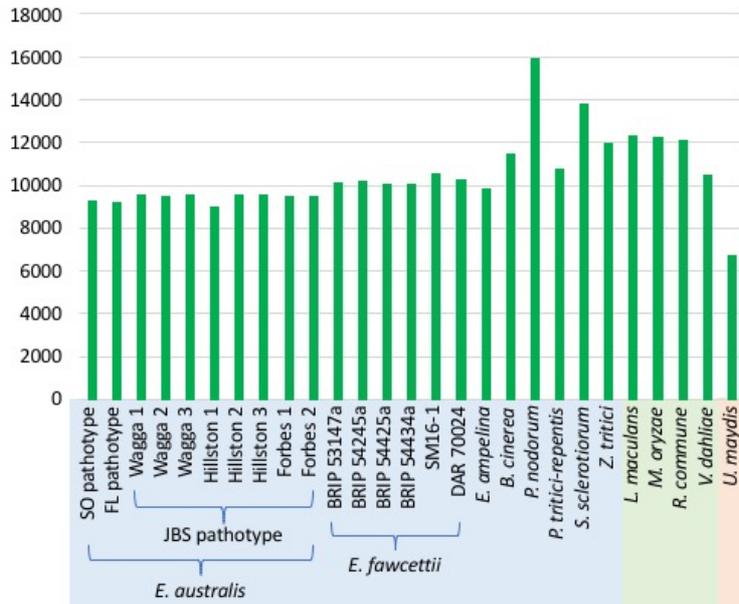
Figure 10. Comparison of number of predicted genes (not including genes overlapping predicted TE regions). Isolates/species on x-axis are grouped into necrotrophic (blue), hemi-biotrophic (green) and biotrophic (red) pathogens.
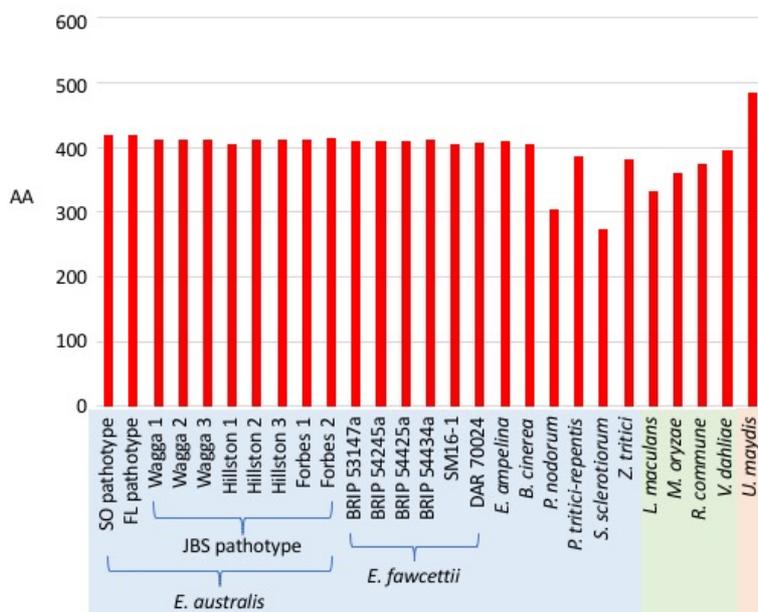
Figure 11. Comparison of median protein length in amino acids (AA). Isolates/species on x-axis are grouped into necrotrophic (blue), hemi-biotrophic (green) and biotrophic (red) pathogens.
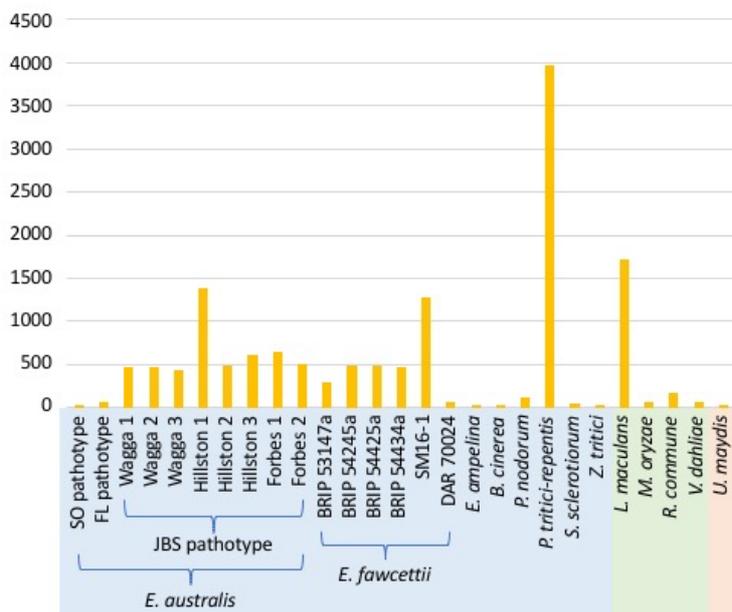
Figure 12. Comparison of number of scaffolds contained within genome assemblies. Isolates/species on x-axis are grouped into necrotrophic (blue), hemi-biotrophic (green) and biotrophic (red) pathogens

3.1.3.2 TRANSPOSABLE ELEMENTS AND REPETITIVE DNA

TE regions were identified through an analysis against Repbase (release 18.02) (Bao et al., 2015) and showed a coverage which ranged from 0.2% (*Elsinoë australis* SO pathotype) to 10.7% (*Leptosphaeria maculans*) over the 27 isolates analysed (Table 2). All *Elsinoë* spp. contained low TE coverage; isolates of *E. fawcettii* contained < 0.4%, the *E. australis* pathotypes contained between 0.2% and 1.36%, while *E. ampelina* had the highest TE coverage at 1.59%. These were comparable to *Parastagonospora nodorum* (0.64%), *Ustilago maydis* (1.0%), *Verticillium dahliea* (1.49%) and *Pyrenophora tritici-repentis* (1.63%). The remaining species including *Zymoseptoria tritici*, *Botrytis cinerea*, *Sclerotinia sclerotiorum*, *Rhynchosporium commune*, *Pyricularia oryzae* and *L. maculans* contained > 2% TE coverage. All species contained very low

proportions of SSR, all *Elsinoë* species had < 0.06% SSR coverage, while the remaining comparative species contained between 0.07% (*Parastagonospora nodorum*) and 0.55% (*B. cinerea*) (Table 2). The proportion of predicted proteins containing polyAA repeats ranged between 7.51% (*Parastagonospora nodorum*) and 22.98% (*U. maydis*) over all species, with a median of 11.87%. All *Elsinoë* spp. contained between 10.57% (*E. fawcettii* BRIP 54245a) and 12.37% (*E. ampelina*), with the exception of *E. fawcettii* SM16-1 which contained only 7.91%.

Table 2. Transposable element and repetitive sequence coverage of *Elsinoë* and comparative genome assemblies.

Isolates	Assembly size (bp)	TE coverage		SSR coverage		Proteins (not including those overlapping predicted TE regions)	Proteins containing at least one polyAA repeat	
		bp	%	bp	%		Number	%
Elsinoe australis JBS pathotype Hillston-1*	24,982,974	89,569	0.36	11,595	0.05	8,967	1,018	11.35
E. australis JBS pathotype Wagga-1*	26,279,129	297,453	1.13	15,451	0.06	9,547	1,135	11.89
E. australis JBS pathotype Forbes-1*	26,621,968	342,744	1.29	15,471	0.06	9,465	1,122	11.85
E. australis JBS pathotype Hillston-2*	26,626,433	355,201	1.33	15,460	0.06	9,533	1,128	11.83
E. australis JBS pathotype Wagga-2*	26,604,401	359,987	1.35	15,969	0.06	9,482	1,120	11.81
E. australis JBS pathotype Forbes-2*	26,617,985	350,690	1.32	15,379	0.06	9,488	1,127	11.88
E. australis JBS pathotype Wagga-3*	26,289,625	312,797	1.19	15,258	0.06	9,521	1,127	11.84
E. australis JBS pathotype Hillston-3*	27,122,845	367,733	1.36	16,544	0.06	9,570	1,121	11.71
E. australis FL pathotype	23,866,517	102,110	0.43	10,181	0.04	9,215	1,094	11.87

BRIP 52616a*								
E. australis SO pathotype **Arg-1**	23,793,094	48,291	0.20	8,354	0.04	9,253	1,093	12.04
E. fawcettii BRIP 54245a*	26,483,858	96,658	0.36	7,497	0.03	10,164	1,074	10.57
E. fawcettii BRIP 54425a*	25,855,480	82,437	0.32	6,543	0.03	10,014	1,068	10.67
E. fawcettii BRIP 54434a*	25,873,468	82,565	0.32	6,813	0.03	10,016	1,068	10.66
E. fawcettii BRIP 53147a*	26,011,141	95,654	0.37	6,868	0.03	10,080	1,073	10.64
E. fawcettii SM16-1	26,655,928	60,231	0.23	6,424	0.02	10,519	1,091	10.37
E. fawcettii DAR 70024	26,325,864	91,027	0.35	7,310	0.03	10,223	1,083	10.59
E. ampelina YL-1	28,296,786	449,125	1.59	19,124	0.07	9,804	1,190	12.14
Botrytis cinerea B05.10	42,630,066	1,175,590	2.76	235,976	0.55	11,481	1,657	14.43
Parastagonospora nodorum SN15	37,213,987	236996	0.64	24474	0.07	15878	1193	7.51
Pyrenophora tritici-repentis ptrDW5	33,413,674	546089	1.63	42164	0.13	10772	842	7.82

Sclerotinia sclerotiorum **1980 UF-70**	38,459,164	1343860	3.49	110869	0.29	13770	1193	8.66
Zymoseptoria tritici **ST99CH_1E4**	38,628,401	950403	2.46	51222	0.13	11936	1439	12.06
Ustilago maydis **521**	19,664,356	197573	1.00	52713	0.27	6692	1538	22.98
Leptosphaeria maculans **JN3**	45,124,619	4829371	10.70	185421	0.41	12337	1571	12.73
Pyricularia oryzae (syn. *Magnaporthe oryzae*) **70-15**	40,979,121	3715140	9.07	193960	0.47	12236	2002	16.36
Rhynchosporium commune **UK7**	55,639,110	3481411	6.26	150452	0.27	12100	1188	9.82
Verticillium dahliae **VdLs.17**	33,900,324	504619	1.49	106896	0.32	10441	1704	16.32

*Isolate sequenced in the current study

3.1.3.3 GC CONTENT

Overall GC content of assemblies and GC content of CDS were compared among all isolates of *Elsinoë* sequenced in the current study, recently published *Elsinoë* sequences (*Elsinoë australis* SO pathotype Arg-1, *E. ampelina*, and *E. fawcettii* isolates DAR 70024 and SM16-1) and ten additional fungal pathogen species. Results of individual isolates are detailed in Table 3. The overall GC content of assemblies ranged between 41.8 % (*Sclerotinia sclerotiorum*) and 55.8% (*Verticillium dahliae*), with a mean of 50.5% among the 27 assemblies. The *Elsinoë* spp. were found close to the mean, ranging from 49.5% (*E. ampelina*) to 52.6% (*E. australis* SO pathotype). The median GC content of coding DNA sequence ranged from 45.7% (*S. sclerotiorum*) to 59.4% (*V. dahliae*) and was higher in comparison to the corresponding overall GC content of the assembly for all isolates. All *E. australis* pathotypes and isolates ranged from 54.7 – 54.9%, *E. fawcettii* ranged from 54.1% – 54.4%, and *E. ampelina* was 54.9%. Additionally, searching the assemblies for the occurrence of AT-rich regions indicated levels ranged between 0% (multiple isolates of *E. fawcettii*, *Pyrenophora tritici-repentis, S. sclerotiorum*, *Ustilago maydis* and *Pyricularia oryzae*) and 37% (*Leptosphaeria maculans*) (Table 3). Seven of the eight isolates of the *E. australis* JBS pathotype contained 8.1% - 9.7% AT rich regions, with the exception of Hillston-1 at 3.4%. These AT-rich regions had a mean GC content ranging between 28.0% and 29.9%. The *E. australis* FL and SO pathotypes indicated lower proportions of AT-rich regions compared to the JBS pathotype, at 2.5% and 0.9%, respectively, however with similar mean GC content (24.9% and 26.3%, respectively). The isolates of *E. fawcettii* had the lowest proportion of AT-rich regions among the *Elsinoë* spp., with 0% for all isolates except for BRIP 53147a (1.0%) and DAR 70024 (0.9%). The GC content of these AT-rich regions was also less extreme at 33.8% and 35.3%, respectively. *E. ampelina* showed the greatest degree of AT-rich regions, with 17.9% of the assembly have a mean GC content of 30.5%. Additional comparative species which showed lower proportions of AT-rich regions included *Botrytis cinerea* (4.9%), *Parastagonospora nodorum* (6.6%) and

V. dahlia (1.5%), while those with higher proportions included *Zymoseptoria tritici* (17.3%), *Rhynchosporium commune* (29.5%) and *L. maculans* (37%).

Table 3. GC content and AT-rich region coverage of *Elsinoë* and comparative genome assemblies.

Isolates	Overall GC content of assembly (%)	Median GC content of coding DNA sequence [Q_1, Q_3] (%)	AT-rich regions		GC equilibrated regions	
			Proportion of assembly (%)	Mean GC content (%)	Proportion of assembly (%)	Mean GC content (%)
Elsinoe australis JBS pathotype Hillston-1	52.0	54.7 [52.8, 57.4]	3.4	28.0	96.6	53.0
E. australis JBS pathotype Wagga-1*	51.1	54.9 [52.8, 57.6]	8.3	29.9	91.7	53.1
E. australis JBS pathotype Forbes-1*	50.9	54.9 [52.8, 57.6]	9.1	29.7	90.9	53.1
E. australis JBS pathotype Hillston-2*	50.9	54.9 [52.8, 57.6]	9.1	29.7	90.9	53.1
E. australis JBS pathotype Wagga-2*	50.9	54.8 [52.8, 57.5]	9.2	29.7	90.8	53.1
E. australis JBS pathotype Forbes-2*	50.9	54.8 [52.8, 57.6]	9.3	29.8	90.7	53.2
E. australis JBS pathotype Wagga-3*	51.1	54.9 [52.8, 57.6]	8.1	29.5	91.9	53.1
E. australis JBS pathotype Hillston-3*	50.8	54.9 [52.8, 57.6]	9.7	29.5	90.3	53.2
E. australis FL pathotype BRIP 52616a*	52.3	54.9 [52.8, 57.6]	2.5	24.9	97.5	53.2
E. australis SO pathotype Arg-1	52.6	54.8 [52.8, 57.5]	0.9	26.3	99.1	53.0
E. fawcettii BRIP 54245a*	52.3	54.2 [52.3, 56.6]	-	-	-	-
E. fawcettii BRIP 54425a*	52.3	54.2 [52.4, 56.7]	-	-	-	-
E. fawcettii BRIP 54434a*	52.3	54.2 [52.4, 56.7]	-	-	-	-
E. fawcettii BRIP 53147a*	52.3	54.2 [52.4, 56.6]	1.0	33.8	99.0	52.8
E. fawcettii	52.8	54.4 [52.5, 57.2]	-	-	-	-

SM16-1						
E. fawcettii DAR 70024	52.3	54.1 [52.3, 56.6]	0.9	35.3	99.1	52.8
E. ampelina YL-1	49.5	54.9 [53.0, 57.6]	17.9	30.5	82.1	53.8
Botrytis cinerea B05.10	42.0	46.2 [44.7, 48.1]	4.9	18.2	95.1	43.4
Parastagonospora nodorum SN15	50.4	54.3 [51.9, 56.9]	6.6	26.8	93.4	52.2
Pyrenophora tritici-repentis ptrDW5	50.9	53.6 [51.7, 55.5]	-	-	-	-
Sclerotinia sclerotiorum 1980 UF-70	41.8	45.7 [43.9, 47.6]	-	-	-	-
Zymoseptoria tritici ST99CH_1E4	52.3	55.7 [54.4, 57.3]	17.3	44.7	82.7	54.2
Ustilago maydis 521	54.0	55.9 [54.4, 57.6]	-	-	-	-
Leptosphaeria maculans JN3	45.2	54.2 [51.7, 56.7]	37	33.8	63	52.2
Pyricularia oryzae (syn. *Magnaporthe oryzae*) 70-15	51.6	57.4 [54.5, 61.1]	-	-	-	-
Rhynchosporium commune UK7	42.6	49.9 [48.5, 51.5]	29.5	32.2	70.5	47.3
Verticillium dahliae VdLs.17	55.8	59.4 [56.4, 63.0]	1.5	30.9	98.5	56.6

*Isolate sequenced in the current study

3.1.3.4 GENE ORTHOLOGY CLASSIFICATION

3.1.3.4.1 *ELSINOË AUSTRALIS* JBS PATHOTYPE

Analysis of gene ortholog classification among the eight isolates of the
Elsinoë australis JBS pathotype indicated that a large majority of genes were
shared, finding hits through OrthoMCL or ProteinOrtho among all eight
isolates (Figure 13). Core genes among the eight isolates of *E. australis* JBS
pathotype, were determined as those which appeared in at least seven of the
eight isolates, accessory genes were those which appeared in two to six
isolates, while unique genes were found in only one isolate. A similar number
of core genes were found for seven of the eight isolates, within the range of
9,228 (Hillston-3) and 9,310 (Wagga-1). The Hillston-1 isolate, however,
could be singled out with only 8,792 core genes identified. Similarly,
accessory genes for the same seven isolates ranged between 157 (Forbes-
1) and 221 (Wagga-1), while Hillston-1 had 132. The number of unique
genes for the eight isolates varied widely between 3 (Forbes-2 and 147
(Hillston-3) genes. Data for the generation of Figure 13 is detailed in
Appendix C Table 2.
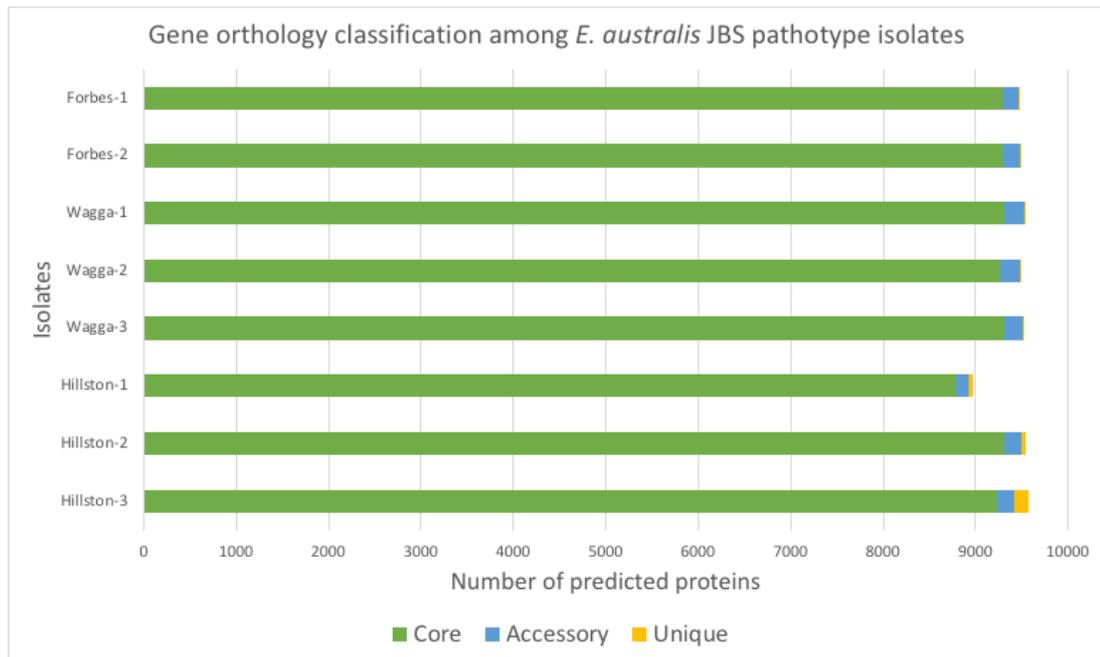
Figure 13. Comparison of gene classifications among the proteomes of eight *E. australis* JBS pathotypes. Genes were considered; (I) core if they were shared by at least seven of the eight isolates (green); (II) accessory if they were shared by between two and six isolates (blue); (III) unique if they were found in only one of the eight isolates (orange).

3.1.3.4.2 *ELSINOË AUSTRALIS*: JBS, FL AND SO PATHOTYPES

The Hillston-2 isolate of *Elsinoë australis* JBS pathotype was utilised for comparison with other *E. australis* pathotypes in this section, and with other *Elsinoë* spp. and fungal pathogens in the following sections. The Hillston-2 isolate was selected as a reference for the *E. australis* JBS pathotypes as it contained the highest number of predicted proteins and obtained the highest number of complete and since copy BUSCO's (Table 1). The three *E. australis* pathotypes (JBS, FL and SO) were subjected to the same gene ortholog classification analysis as above, with the exception that core genes were determined as those which appeared in all three isolates and accessory genes as those which appeared in two of the pathotypes. There were between 8,430 (JBS pathotype) and 8,460 (SO pathotype) core genes, making up the majority of predicted genes of each of the three pathogens

(88.4%, 91.7% and 91.4% of the JBS, FL and SO pathotypes, respectively) (Figure 14). The number of accessory genes ranged between 485 (SO pathotype) and 532 (FL pathotype), providing only 5.4%, 5.8% and 5.2% of JBS, FL and SO pathotype accessory genes, respectively. Finally, there were 584 (JBS pathotype), 237 (FL pathotype) and 308 (SO pathotype) unique genes, making up only a small proportion of total genes; 6.9%, 2.8% and 3.6%, respectively. Data for the generation of Figure 14 is detailed in Appendix C Table 3.
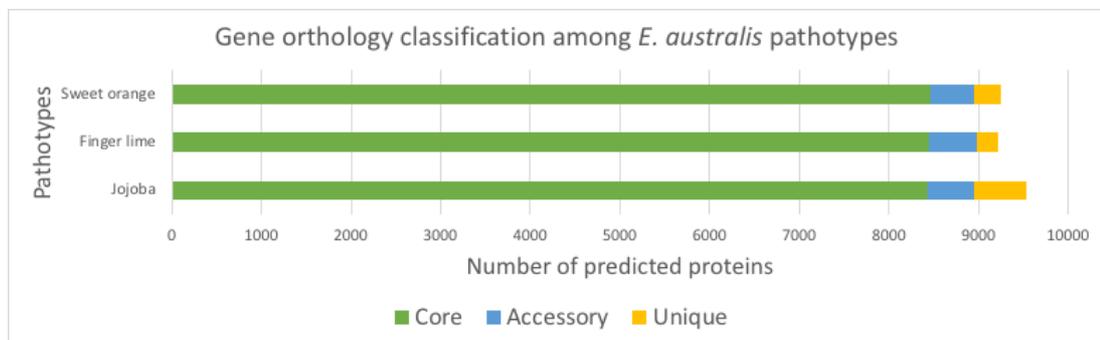
Figure 14. Comparison of gene classifications among the proteomes of three *E. australis* pathotypes. Genes were considered; (I) core if they were shared by all three pathotypes (green); (II) accessory if they were shared by at least two pathotypes, but not all (blue); (III) unique if they were found in only one of the three pathotypes (orange).

3.1.3.4.3 *ELSINOË FAWCETTII*

Gene ortholog classification was performed on six isolates of *Elsinoë fawcettii* in relation to one another, this included the four isolates sequenced in the current study (BRIP 53147a, BRIP 54245a, BRIP 54425a and BRIP 54434a) and a further two currently available from GenBank (SM16-1 and DAR 70024). As illustrated in Figure 15, the number of core genes appearing in all six isolates were closely similar, ranging from 9,496 (BRIP 54425a) to 9,592 (SM16-1). The number of accessory and unique genes, however, were

only close among five of the six isolates, while isolate SM16-1 sat apart from the others. SM16-1 contained only 223 accessory genes, which was lower in comparison to the other five isolates which ranged from 513 (BRIP 54425a) to 619 (DAR 70024). Furthermore, SM16-1 contained 704 unique genes, while the other five isolates ranged from 5 (BRIP 54425a) to 38 (DAR 70024) unique genes each. The overrepresentation of unique genes in SM16-1, seen in Figure 15, was investigated further. Of the 704 genes classified as unique to SM16-1, 485 (68.9%) had a high GC content (greater than the Q_3 value). Additionally, 308 (43.8%) genes obtained an orthoMCL group ID, of which 209 belonged to an orthoMCL group which contained bacterial species with no other fungal species. Of these 209, 201 (96.2%) also had a GC content greater than the Q_3 value. This suggested a portion of the SM16-1-specific genes may be attributed to contamination by bacterial sequences. While some of these SM16-1-specific genes were scattered throughout the assembly, others were grouped closely together and accounted for the majority of genes on a contig. Twenty contigs were found to contain either only SM16-1-specific genes or genes labelled as core, but which had paralogs elsewhere within the assembly, and which contained at least 85% of genes with a high CDS GC (greater than the Q_3 value) content. This led to the labelling of 516 genes over 20 contigs, of which 96.7% had a CDS GC content greater than the Q_3 value, as potential bacterial contamination of the assembly. When removing these potentially contaminating sequences from the predicted gene pool, the SM16-1 assembly contained a total of 10,003 predicted genes, of which 9,513 were core amongst all six isolates of *E. fawcettii*, 222 were accessory and 268 were unique to the SM16-1 isolate. In all downstream analyses from this point, the 516 potential bacterial contamination gene models were removed from *E. fawcettii* SM16-1.
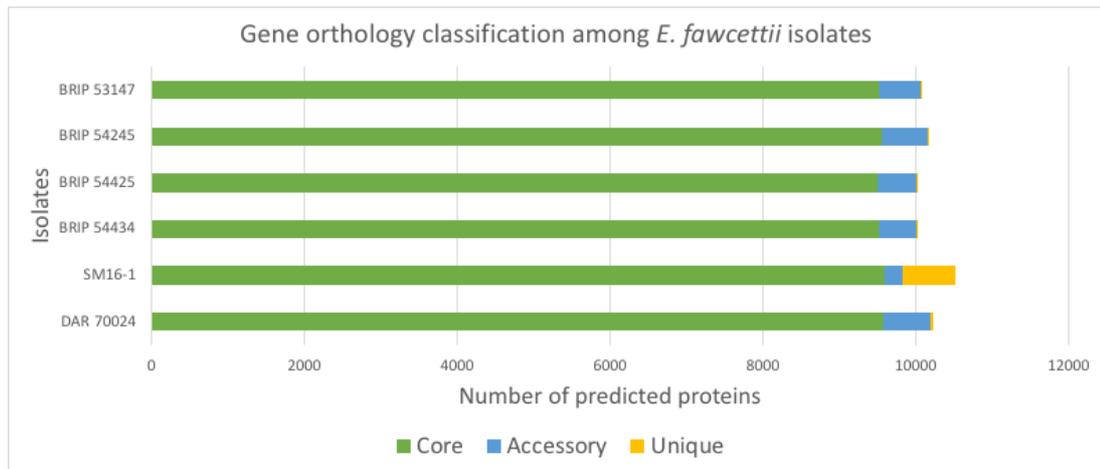
Figure 15. Comparison of gene classifications among the proteomes of six genomes representing *E. fawcettii*, before the labelling of potential bacterial contamination in isolate SM16-1. Genes were considered; (I) core if they were shared by all six isolates (green); (II) accessory if they were shared by at least two isolates, but not all (blue); (III) unique if they were found in only one of the six isolates (orange). Full data is detailed in Appendix C Table 4.

3.1.3.4.4 *ELSINOË* SPP. COMPARED WITH ADDITIONAL FUNGAL PATHOGENS

Five *Elsinoë* assembles, including *Elsinoë fawcettii* (BRIP 53147a), *E. australis* JBS pathotype (Hillston-2), *E. australis* FL pathotype (BRIP 52616a), *E. australis* SO pathotype (Arg-1) and *E. ampelina*, were analysed for gene orthology with 10 pathogenic fungal species. Core genes shared by all 15 genomes ranged from 2,710 (*Ustilago maydis*) to 3,217 (*Parastagonospora nodorum*) (Figure 16). Further genes, found in all 14 Ascomycota isolates, were labelled as "Ascomycota core" and ranged between 1,045 (*Leptosphaeria maculans*) and 1,148 (*Parastagonospora nodorum*). Necrotrophic-specific genes, defined as those found in at least two necrotrophic species (*Elsinoë* species, *Zymoseptoria tritici*, *Sclerotinia sclerotiorum*, *Pyrenophora tritici-repentis*, *Parastagonospora nodorum* and *Botrytis cinerea*), ranged between 333 (*E. australis* FL pathotype) and 1,393 (*S. sclerotiorum*). There were 22 core *Elsinoë*-specific

genes, while *Elsinoë*-specific accessory genes (those shared by at least two *Elsinoë* genomes but not all) ranged between 1,285 (*E. fawcettii*) and 1,591 (*E. australis* JBS pathotype). Of the *Elsinoë*-specific accessory genes, between 66 (*E. ampelina*) and 98 (*E. australis* SO pathotype) obtained orthoMCL groups, indicating potential conservation within a wider group of organisms. The number of accessory genes among all 15 pathogens, being those shared between at least two isolates, but not all, and which were not already classified as Ascomycota core, necrotrophic-specific or *Elsinoë*-specific, ranged between 1,550 (*U. maydis*) and 5,666 (*Parastagonospora nodorum*). Within this range, the five *Elsinoë* genomes contained between 3,189 (*E. australis* FL pathotype) and 3,582 (*E. fawcettii*) accessory genes. Lastly, the number of species-specific genes ranged between 104 (*E. australis* FL pathotype) and 5,031 (*Parastagonospora nodorum*). A subset of each group of species-specific genes (ranging from seven for the *E. australis* JBS pathotype to 515 for *U. maydis*) obtained a hit to an orthoMCL ID, indicating while they were unique within the group of 15 fungal pathogens analysed, orthologs likely existed in a larger dataset. Consequently, genes which obtained a hit to an orthoMCL ID, were not considered unique in downstream analyses. The remaining unique genes presented a subset of interest. Among the *Elsinoë* pathogens, between 90 (*E. australis* FL pathotype) and 786 (*E. ampelina*) genes were presumed to be specific to individual species of *Elsinoë* or their pathotypes. Data for species- and pathotype-specific genes, with and without orthoMCL group ID's, as well as data for the generation of Figure 16, are detailed in Appendix C Table 5.
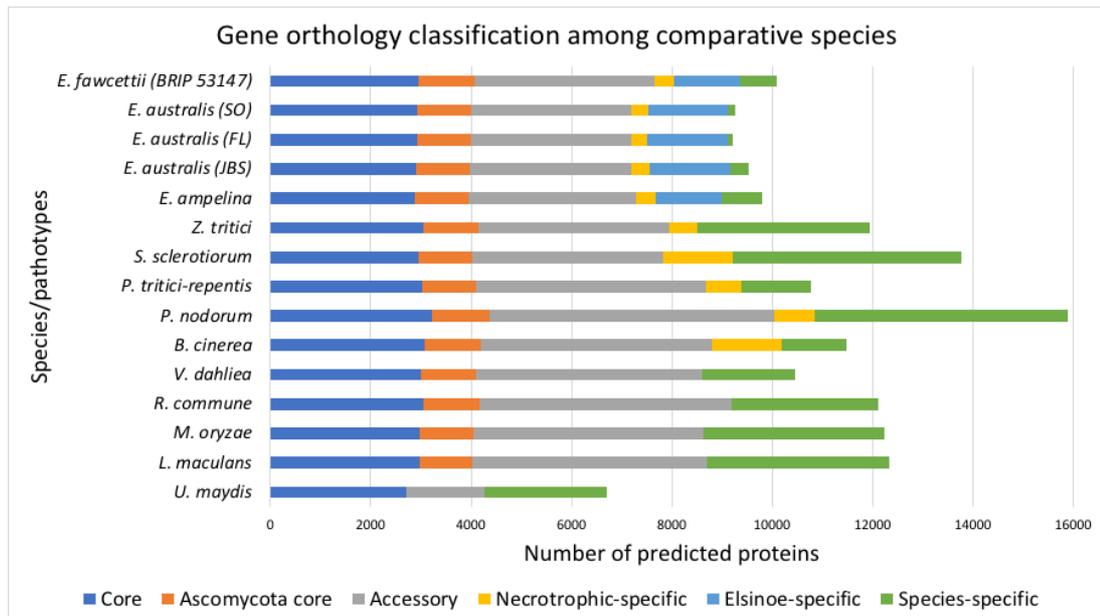
Figure 16. Comparison of gene classifications among the proteomes of 15 fungal pathogens. Genes were considered; (I) core if they were shared by all 15 species/pathotypes (dark blue); (II) Ascomycota core if they were shared by all pathogens except for *Ustilago maydis* (orange); (III) accessory if they were shared by at least two species, but not all (grey); (IV) necrotrophic-specific if they were shared by at least two necrotrophic pathogens, but no hemi-biotrophic or biotrophic species (yellow); (V) *Elsinoë*-specific if they were found only among isolates of the *Elsinoë* genus (light blue); (VI) unique if they were found in only one of the 15 pathogens (green).

3.1.3.5 PROTEIN FAMILY DATABASE ANALYSIS

An Hmmscan (Johnson et al., 2010) search of the predicted proteomes of the 17 isolates of *Elsinoë* against the Pfam database (Finn et al., 2016) indicated the majority of genes, between 68.5% and 73.0%, obtained at least one hit to a Pfam model (Table 4). Results for the proteomes of the ten comparative species showed between 48.6% (*Sclerotinia sclerotiorum*), and 74.9% (*Ustilago maydis*) of proteins obtained at least one hit. A mean of 67.2%, for all species and isolates analysed, was obtained. While the 27 proteomes had a large range in size, from 6,692 (*U. maydis*) to 15,878 (*Parastagonospora nodorum*) proteins, the range in number of proteins with

Pfam hits was comparatively smaller, from 5,011 (*U. maydis*) to 8,101 (*Parastagonospora nodorum*) (Table 4).

Table 4. Predicted proteins and numbers of Pfam hits, ordered by proportion of Pfam hits, smallest to largest.

Species	Isolate	Predicted proteins	Proteins with at least one Pfam hit	
			Number	%
Sclerotinia sclerotiorum	1980 UF-70	13,770	6,696	48.63%
Parastagonospora nodorum	SN15	15,878	8,101	51.02%
Leptosphaeria maculans	JN3	12,337	6,591	53.42%
Pyricularia oryzae	70-15	12,236	7,258	59.32%
Zymoseptoria tritici	ST99CH_1E4	11,936	7,392	61.93%
Rhynchosporium commune	UK7	12,100	7,511	62.07%
Pyrenophora tritici-repentis	ptrDW5	10,772	7,103	65.94%
Botrytis cinerea	B05.10	11,481	7,603	66.22%
Elsinoë ampelina	YL-1	9804	6717	68.51%
E. fawcettii	DAR 70024	10223	7120	69.65%
E. fawcettii	BRIP 54245a*	10164	7093	69.79%
E. fawcettii	BRIP 53147a*	10,080	7,069	70.13%
Verticillium dahlia	VdLs.17	10,441	7,326	70.17%
E. australis JBS pathotype	Hillston-3*	9570	6719	70.21%
E. fawcettii	BRIP 54425a*	10014	7037	70.27%

E. fawcettii	BRIP 54434a*	10016	7049	70.38%
E. australis **JBS pathotype**	Wagga-1*	9547	6740	70.60%
E. australis **JBS pathotype**	Hillston-2*	9533	6733	70.63%
E. australis **JBS pathotype**	Wagga-3*	9521	6731	70.70%
E. australis **JBS pathotype**	Wagga-2*	9482	6710	70.77%
E. australis **JBS pathotype**	Forbes-1*	9465	6706	70.85%
E. australis **JBS pathotype**	Hillston-1*	8967	6359	70.92%
E. australis **JBS pathotype**	Forbes-2*	9488	6737	71.01%
E. fawcettii	SM16-1	10003	7082	70.80%
E. australis **SO pathotype**	Arg-1	9253	6712	72.54%
E. australis **FL pathotype**	BRIP 52616a*	9215	6727	73.00%
Ustilago maydis	521	6,692	5,011	74.88%

*Isolate sequenced in the current study

3.2 PREDICTION OF VIRULENCE GENES

3.2.1 PREDICTION OF SECONDARY METABOLITE GENE CLUSTERS

SM gene clusters were predicted in all assemblies of all *Elsinoë* spp., pathotypes and isolates sequenced in the current study, as well as in the four *Elsinoë* assemblies currently available from GenBank. For the *Elsinoë australis* JBS pathotype Hillston-2, a total of 14 SM gene clusters were predicted, including three T1PKS, one type 3 polyketide synthase (T3PKS), six terpene, three non-ribosomal peptide synthase (NRPS) and one "other".

Each predicted cluster contained between four and eighteen genes. These numbers were comparable to the seven other isolates of *E. australis* JBS pathotype, which contained between 11 (Hillston-1) and 16 (Wagga-1) predicted SM clusters (Table 5). Similarly, the *E. australis* SO pathotype had 15 predicted SM clusters, including three T1PKS, one T3PKS, five terpene and six NRPS. The *E. australis* FL pathotype had the highest number of predicted SM clusters, with 17 in total, including six T1PKS, one T3PKS, six terpene and four NRPS. While there was a small variation in prediction of SM clusters among the pathotypes and isolates of *E. australis*, those for *E. fawcettii* showed no variation, all six isolates contained 15 predicted SM gene clusters, each of which included six T1PKS, one T3PKS, four terpene and four NRPS gene clusters. Lastly, *E. ampelina* had 16 predicted SM gene clusters, which included four T1PKS, one T3PKS, five terpene and six NRPS gene clusters.

Table 5. Predicted secondary metabolite gene clusters of the *Elsinoë* spp.

Species / Isolate	Total predicted clusters	Total genes involved in clusters	SM Type				
			T1PKS	T3PKS	Terpene	NRPS	Other
E. australis JBS pathotype							
Hillston-1*	11	114	4	0	4	3	0
Hillston-2*	14	145	3	1	6	3	1
Hillston-3*	14	144	3	1	6	3	1
Forbes-1*	13	144	4	1	5	3	0
Forbes-2*	15	164	4	1	6	4	0
Wagga-1*	16	176	4	1	6	4	1
Wagga-2*	14	146	3	1	6	3	1
Wagga-3*	15	166	4	1	6	4	0
E. australis FL pathotype							
BRIP 52616a*	17	201	6	1	6	4	0
E. australis SO pathotype							

Arg-1	15	200	3	1	5	6	0
E. fawcettii							
BRIP 53147a*	15	233	6	1	4	4	0
BRIP 54245a*	15	233	6	1	4	4	0
BRIP 54425a*	15	231	6	1	4	4	0
BRIP 54434a*	15	234	6	1	4	4	0
SM16-1	15	232	6	1	4	4	0
DAR 70024	15	233	6	1	4	4	0
E. ampelina							
YL-1	16	219	4	1	5	6	0

*Isolate sequenced in the current study

The degree to which these predicted SM gene clusters may be shared among the different species and pathotypes of *Elsinoë* was investigated. An all-vs-all BlastP search of predicted SM genes from *E. fawcettii* (BRIP 53147a), the three *E. australis* pathotypes (SO, FL and JBS) and *E. ampelina* indicated that predicted core biosynthetic genes for six candidate SM gene clusters were shared by all five species/pathotypes (Table 6). For three of these gene clusters additional biosynthetic genes were also shared. The Know Cluster Blast setting of the AntiSMASH program, which searches the MIBiG (Minimum Information about a Biosynthetic Gene cluster) database (Kautsar et al., 2020), identified three gene clusters with similarity to known SM gene clusters of *Bipolaris oryzae*, *Hypholoma sublateritium* and *Parastagonospora nodorum*, which produce melanin, clavaric acid and elsinochrome, respectively. Additionally, the known elsinochrome gene cluster (Chung & Liao, 2008a) was also queried in a BlastP against the predicted SM genes and retrieved a match to the first set of gene clusters in Table 6 with a very high similarity of 98.6%. The elsinochrome gene cluster of *E. fawcettii* (BRIP 53147a) is illustrated in Figure 17, illustrating the two essential biosynthetic genes, four other biosynthetic genes that are currently known and nine additional gene models

newly predicted as potentially being part of the gene cluster. Pfam hits obtained for the additional genes included the THUMP domain, peptidase M3, Apolipoprotein O, Gar1/Naf1 RNA binding region and Endonuclease/Exonuclease/phosphatase family.

Table 6. Candidate Secondary Metabolite (SM) gene clusters shared by species and pathotypes of *Elsinoë*. Predicted core biosynthetic genes shown have >80% similarity among the five *Elsinoë* isolates. Similarity of *E. fawcettii* (BRIP 53147a) genes to core biosynthetic genes of known SM gene clusters is also shown.

Gene Cluster #	SM class	GenBank accession, genomic location, (number of genes in cluster), predicted core biosynthetic gene					Similarity to known SM biosynthetic gene clusters	
		E. fawcettii BRIP 53147a	*E. australis* JBS pathotype Hillston-2	*E. australis* FL pathotype BRIP 52616a	*E. australis* SO pathotype Arg-1	*E. ampelina* YL-1	Known SM cluster (*organism*), core biosynthetic gene	Similarity to *E. fawcettii* BRIP 53147a
1	T1PKS	SDJM01000001 641093:686753 (15 genes) KAF4556316	PTQR01000039.1 681055:729151 (13 genes) Ee.00g045890	WLZB01000005.1 884771:932879 (13 genes) Ee.00g045120	SWCS01000013.1 165884:210802 (12 genes) Ea.00g045300	SMYM01000002.1 2232305:2269397 (13 genes) Ea.00g095530	Elsinochrome (*E. fawcettii*) ABU63483.1 & Melanin (*Bipolaris oryzae*) BAD22832.1	98.6% 64%
2	Terpene	SDJM01000003.1 907309:929586 (10 genes) KAF4555079	PTQR01000011.1 410170:432436 (11 genes) Ee.00g012120	WLZB01000029.1 84860:107126 (11 genes) Ee.00g074680	SWCS01000002.1 1604289:1626555 (12 genes) Ea.00g055240	SMYM01000006.1 1814970:1837241 (13 genes) Ea.00g085120	Clavaric acid (*Hypholoma sublateritium*) ACF70484.1	48%
3	T1PKS	SDJM01000028.1 107528:154316 (15 genes) KAF4548432	PTQR01000024.1 42113:88909 (17 genes) Ee.00g049710	WLZB01000005.1 55,324 - 99,012 (14 genes) Ee.00g041900	SWCS01000013.1 976526:1023321 (16 genes) Ea.00g048490	SMYM01000002.1 3119385:3166179 (15 genes) Ea.00g098490	Elsinochrome C (*Parastagonospora nodorum*) SNOG_08614	57%
4	Terpene	SDJM01000020.1 268920:290448 (8 genes) KAF4549728	PTQR01000082.1 232864:254400 (8 genes) Ee.00g090410	WLZB01000015.1 341252:362788 (6 genes) Ee.00g054860*	SWCS01000004.1 1683410:1704946 (7 genes) Ea.00g043010	SMYM01000008.1 286408:307935 (7 genes) Ea.00g002380	NA	NA
5	Terpene	SDJM01000007.1 22061:44026 (9 genes) KAF4553201	PTQR01000051.1 79428:101374 (4 genes) Ee.00g048810	WLZB01000004.1 375048:396994 (7 genes) Ee.00g023460*	SWCS01000010.1 970205:992150 (7genes) Ea.00g069660	SMYM01000010.1 323724:345694 (9 genes) Ea.00g052500	NA	NA
6	Terpene	SDJM01000003.1 566624:584089 (8 genes) KAF4554938	PTQR01000131.1 69587:90735 (9 genes) Ee.00g026310	WLZB01000024.1 204519:225667 (9 genes) Ee.00g083330	SWCS01000002.1 1296181:1314305 (9 genes) Ea.00g054000	SMYM01000006.1 1340329:1352359 (6 genes) Ea.00g083580	NA	NA
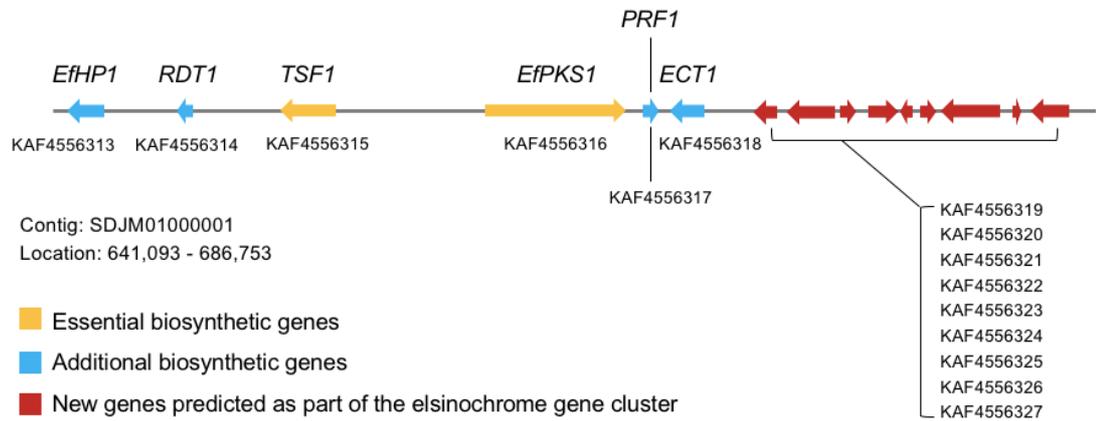
Figure 17. Elsinochrome gene cluster from *Elsinoë fawcettii* (BRIP 53147a). Two essential biosynthetic genes (yellow), four other biosynthetic genes currently known (blue) and nine additional gene models predicted as being potentially part of the gene cluster (maroon).

When comparing the total numbers of predicted SM gene clusters identified within the *Elsinoë* genomes with other fungal pathogens, *Elsinoë* are seen to have smaller overall numbers compared to most species, with *Ustilago maydis* being the only exception (Figure 18). While the results were relatively similar among the sequenced genomes of *Elsinoë*, numbers of predicted SM gene clusters varied greatly among the comparative species, with no similarities seen among either the necrotrophic or hemi-biotrophic species.
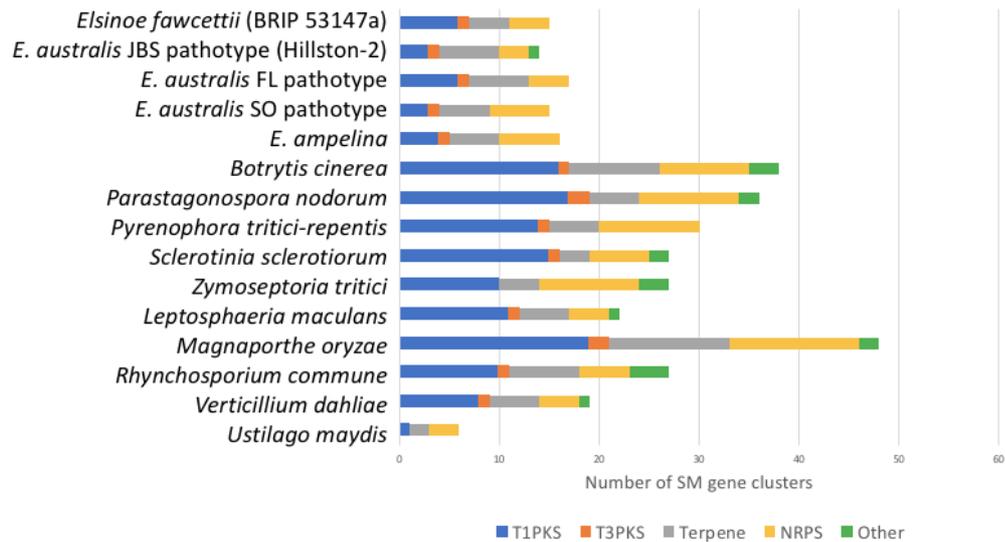
Figure 18. Comparison of number of predicted SM gene cluster types among *Elsinoë* spp. and other fungal pathogens. The proportion of SM gene cluster types is shown for T1PKS (blue), T3PKS (orange), terpene (grey), NRPS (yellow) and "other" (green).

3.2.2 SECRETOME PREDICTION

The secretome of each isolate of *Elsinoë* was predicted using the workflow illustrated in Figure 19, as was that of each species from section 3.1.3.1 for comparison. SignalP and/or Phobius programs predicted classically secreted proteins with a predictable signal peptide at the N-terminal, while ProtComp further identified potential non-classically secreted proteins (detailed in section 2.2.2.2). Any proteins with detectable transmembrane helices or GPI anchors were identified and removed; while these proteins may be targeted for secretion, they likely function from a position located in the cell membrane. For the *Elsinoë australis* JBS pathotype isolates, between 1,039 (Hillston-1) and 1,108 (Wagga-1) proteins were predicted to be secreted (Table 7), being 11.5% - 11.6% of total proteins. These numbers were comparable to the *E. australis* FL and SO pathotypes and *E. ampelina*, being 1,057 (11.5%), 1,091 (11.8%) and 1,167 (11.9%) secreted proteins, respectively. For *E. fawcettii*, these were only slightly higher, ranging between 1,264 (BRIP 54434a) and 1,296 secreted proteins (SM16-1), being

12.6% - 13.0% of total proteins. When comparing the *Elsinoë* spp. to other fungal pathogens, similar proportions of secreted proteins out of total proteins were seen. For example, the necrotrophic species ranged from 11.3% (*Botrytis cinerea*) to 13.9% (*Parastagonospora nodorum*). Similar proportions were seen amongst the hemi-biotrophic and biotrophic species, except for *Leptosphaeria maculans* and *Pyricularia oryzae* which had 15.3% and 18.5%, respectively.
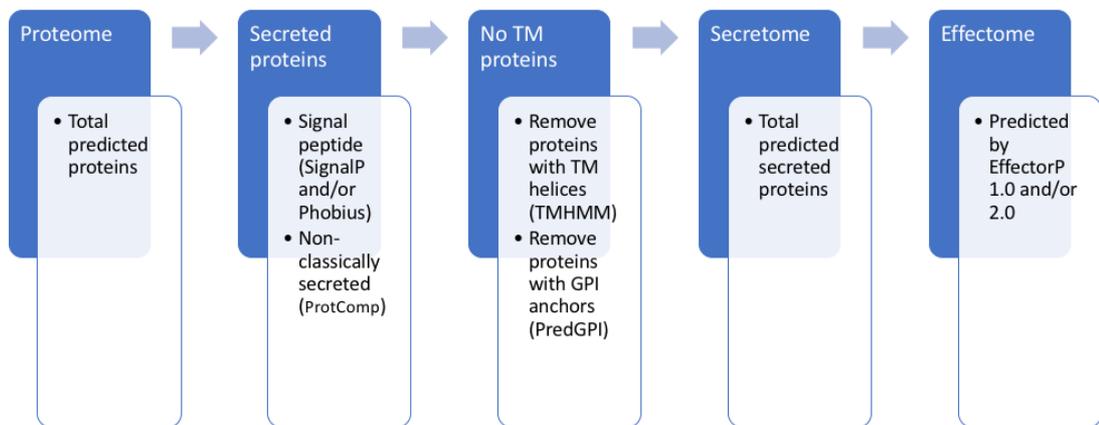
Figure 19. Pathway for the prediction of the secretomes and effectomes. Secretome prediction involved subjecting all predicted proteins of a species (not including those overlapping predicted TE regions) to three different prediction programs. Proteins which were predicted as secreted by at least one program and, additionally, did not contain any predicted transmembrane (TM) helices or GPI-anchors formed the predicted secretomes. From here, candidate effectors were identified.

Table 7. Predicted secreted proteins of *Elsinoë* and comparative species.

Species / isolate	Total proteins**	Total secreted proteins predicted by SignalP, Phobius and/or ProtComp	Total secreted proteins with no TM helices or GPI anchors (% of total proteins)
Elsinoë species:			
E. australis JBS pathotype:			
Hillston-1*	8,967	1,292	1,039 (11.6%)
Hillston-2*	9,533	1,370	1,095 (11.5%)
Hillston-3*	9,570	1,377	1,101 (11.5%)
Forbes-1*	9,465	1,362	1,090 (11.5%)
Forbes-2*	9,488	1,373	1,102 (11.6%)
Wagga-1*	9,547	1,383	1,108 (11.6%)
Wagga-2*	9,482	1,367	1,096 (11.6%)
Wagga-3*	9,521	1,204	1,095 (11.5%)
E. australis FL pathotype			
BRIP 52616a*	9,215	1,318	1,057 (11.5%)
E. australis SO pathotype			
Arg-1	9,253	1,368	1,091 (11.8%)
E. fawcettii			
BRIP 53147a*	10,080	1,578	1,280 (12.7%)
BRIP 54245a*	10,164	1,571	1,290 (12.7%)
BRIP 54425a*	10,014	1,554	1,272 (12.7%)
BRIP 54434a*	10,016	1,545	1,264 (12.6%)
DAR 70024	10,223	1,593	1,291 (12.6%)
SM16-1	10,003	1,588	1,296 (13.0%)
E. ampelina			
YL-1	9,804	1,469	1,167 (11.9%)
Comparative species:			
Necrotrophic			
Botrytis cinerea	11,481	1,936	1,294 (11.3%)
Parastagonospora nodorum	15,878	2,593	2,206 (13.9%)
Pyrenophora tritici-repentis	10,771	1,585	1,298 (12.1%)
Sclerotinia sclerotiorum	13,770	2,033	1,707 (12.4%)

Zymoseptoria tritici	11,936	1,963	1,514 (12.7%)
Hemi-biotrophic			
Leptosphaeria maculans	12,337	2,189	1,883 (15.3%)
Pyricularia oryzae	12,236	2,779	2,263 (18.5%)
Rhynchosporium commune	12,100	1,931	1,510 (12.5%)
Verticillium dahliae	10,441	1,749	1,407 (13.5%)
Biotrophic			
Ustilago maydis	6,692	1,087	856 (12.8%)

*Isolates sequenced in the current study

**Not including genes overlapping predicted TE regions

3.2.3 EFFECTOME PREDICTION

3.2.3.1 EFFECTOME PREDICTION

Effectome prediction was performed on the predicted secretomes using EffectorP 1.0 and EffectorP 2.0, as suggested by Sperschneider et al. (2018). For the *Elsinoë australis* JBS pathotype, the number of candidate effectors ranged between 233 (Hillston-1) and 245 (Hillston-3). Similar numbers were predicted for the *E. australis* FL and SO pathotypes, being 219 and 235, respectively. Higher numbers of effectors were predicted in the secretomes of *E. fawcettii*, ranging from 270 (BRIP 54425a) to 281 (SM16-1). Similarly, 270 effectors were predicted in the secretome of *E. ampelina*. Thus, for the *Elsinoë* spp. between 20.7% and 23.1% of secreted proteins were classed as candidate effectors, which was similar in comparison to *Botrytis* cinerea (22.0%), but lower than all other species analysed (Table 8).

Additionally, known effectors of the comparative species were run through the candidate effector pipeline (Figure 19) and indicated 43 out of 45 (95.6%) were correctly identified as secreted and 42 (93.3%) were also correctly identified as effectors. Vdlsc1 of *Verticillium dahliae* and MoCDIP2 of

Pyricularia oryzae were the two known effectors which were incorrectly not predicted as secreted.

Table 8. Candidate effectors (CE) of *Elsinoë* and comparative species.

Species / isolate	Total secreted proteins (SP)	Candidate effectors predicted with EffectorP 1.0 and/or 2.0 (% of total SP)	Known effectors predicted as secreted and as candidate effectors	Known effectors not predicted as secreted or as candidate effectors
Elsinoë species:				
E. australis JBS pathotype:				
Hillston-1*	1,039	233 (22.4 %)	-	-
Hillston-2*	1,095	236 (21.5 %)	-	-
Hillston-3*	1,101	245 (22.3 %)	-	-
Forbes-1*	1,090	235 (21.6 %)	-	-
Forbes-2*	1,102	238 (21.6 %)	-	-
Wagga-1*	1,108	239 (21.6 %)	-	-
Wagga-2*	1,096	238 (21.7 %)	-	-
Wagga-3*	1,095	237 (21.6 %)	-	-
E. australis FL pathotype				
BRIP 52616a*	1,057	219 (20.7 %)	-	-
E. australis SO pathotype				
Arg-1	1,091	235 (21.5 %)	-	-
E. fawcettii				
BRIP 53147a*	1,280	276 (21.6 %)	-	-
BRIP 54245a*	1,290	275 (21.3 %)	-	-
BRIP 54425a*	1,272	270 (21.2 %)	-	-
BRIP 54434a*	1,264	275 (21.8 %)	-	-
DAR 70024	1,291	274 (21.2 %)	-	-
SM16-1	1,296	281 (23.1 %)	-	-
E. ampelina				
YL-1	1,167	270 (23.1 %)	-	-
Comparative species:				
Necrotrophic				
Botrytis cinerea	1,294	285 (22.0 %)	NEP1	-

Parastagonospora nodorum	2,206	932 (42.2 %)	Tox1, ToxA	-
Pyrenophora tritici-repentis	1,298	388 (29.9 %)	ToxB	-
Sclerotinia sclerotiorum	1,707	692 (40.5 %)	SsSSVP1	-
Zymoseptoria tritici	1,514	597 (39.4 %)	Zt6, AvrStb6	-
Hemi-biotrophic				
Leptosphaeria maculans	1,883	787 (41.8 %)	AvrLM6, AvrLM11, AvrLM4-7	-
Pyricularia oryzae	2,263	1,055 (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	1,510	509 (33.7 %)	NIP1, NIP2, NIP3	-
Verticillium dahliae	1,407	413 (29.4 %)	PevD1, VdSCP7	Vdlsc1
Biotrophic				
Ustilago maydis	856	256 (30.0 %)	Pit2, Pep1, See1, Cmu1, Tin2	Eff1-1

*Isolates sequenced in the current study

3.2.3.2 KNOWN EFFECTOR ANALYSIS

The 42 known effectors from the 10 comparative fungal species, which were correctly predicted as effectors (section 3.2.3.1) were analysed for six features including gene density, GC content of CDS, inclusion in predicted SM gene clusters, species specificity, and number of base pairs to both the nearest TE region and nearest AT-rich region (Table 9). Results obtained for the known effectors were compared to results obtained for the entire proteomes of each species (Appendix D). Proportions of total genes in gene-dense locations (IFR on either side was < 1500 bp) within the proteome ranged between 17.66%, for *Pyrenophora tritici-repentis*, and 70.23%, for *Leptosphaeria maculans*, (Appendix D) with a mean of 49.4% for all comparative species. In comparison, there were 9 (21.4%) of the 42 known effectors that were classified as residing in gene-dense locations (Table 9). Hence, one point was allocated to a known effector unless it was found in a gene-dense location. The GC content of 32 (76.2%) of the 42 known effectors was either less than the Q_1 value or greater than the Q_3 value, of the specific species. As this was higher than a combined 50% in these two quartiles, one point was allocated to known effectors found in these two quartiles. None of the known effectors were involved in predicted SM gene clusters, thus one point was allocated for no predicted involvement in SM clusters. The proportion of species-specific genes within the proteomes of the comparative species ranged between 11.9%, for *Botrytis cinerea*, and 33.7%, for *Sclerotinia sclerotiorum*, with a mean of 25.4% for all species. While the proportion of genes which obtained a hit to the orthoMCL ID of a known effector was very low at < 0.3% for each comparative species. In comparison, of the 42 known effectors, 41 (97.6%) were classified as either species-specific (31) or obtained an orthoMCL hit to a current known effector (10). Therefore, one point was given to known effectors which were unique or had obtained an orthoMCL hit to the same group as a currently known effector. The distance between a known effector and the closest TE region was notable among the species that had > 2% of TE coverage across their genomes, including *Pyricularia oryzae*, *L. maculans*, *Rhynchosporium commune*, *B. cinerea*, *S. sclerotiorum* and *Zymoseptoria tritici*. Within the

genomes of these six species, a mean 47.1% of all genes were located within 20 Kb of a TE region. In comparison, of the known effectors of these same species, 29 (90.6%) were found within 20 Kb of a TE region. Thus, if a species' genomes had > 2% TE coverage, one point was given to known effectors found < 20 Kb from a TE region. Finally, known effectors of the comparative species, with genomes containing greater than 25% AT-rich regions, were all found to be fewer base pairs away from the nearest AT-rich region than the Q_1 value for the species. The genomes of *R. commune* and *L. maculans* had greater than 25% AT-rich region coverage. Thus, for species with genomes with greater than 25% AT-rich region coverage, one point was given to known effectors less than the Q_1 value in base pairs away from the nearest AT-rich region. Therefore, the known effectors of each species were scored out of between four and six points ("*n* points") depending on the degree of TE and AT-rich region content of a species. Of the 42 known effectors, 38 (90.5%) acquired either *n* or *n*-1 points. The known effector analysis was published as part of the current study (Jeffress et al., 2020).

Table 9. 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 region	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
Hemi-biotrophic:								
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
Pyricularia 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:								
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

Note: This table was published as part of the current study (Jeffress et al., 2020).

3.2.3.3 CANDIDATE EFFECTOR PRIORITISATION

The candidate effectors, predicted in section 3.2.3.1, were subjected to the
prioritisation pipeline (Figure 20) generated through the analysis of known effectors
(section 3.2.3.2). The reduction of candidate effectors to prioritised candidate
effectors, for each isolate of *Elsinoë* and comparative species is detailed in Table 10,
while the reduction from secreted proteins is illustrated in Figure 21. The genomic
features utilised during the prioritisation pipeline (gene-dense/gene-spare location,
GC contents of CDS, involvement in predicted SM clusters, potential uniqueness, TE
coverage and AT-rich region coverage) are detailed in Appendix D. The subset of
candidate effectors of each isolate of *Elsinoë* was more than halved during
application of the prioritisation pipeline. Specifically, the number of prioritised
candidate effectors for *Elsinoë australis* JBS pathotypes ranged between 107 and
116. The *E. australis* FL and SO pathotypes were similar, being 94 and 101,
respectively. The prioritisation pipeline led to a reduction of candidate effectors for
the *E. australis* pathotypes of between 52.36% and 57.08%. Looking at *E. fawcettii*,
the number of prioritised candidate effectors ranged between 118 and 120, again a
reduction by more than half (55.93% - 58.00%). The candidate effectors of
E. ampelina were reduced by 55.93%, leaving 119 prioritised candidate effectors.
While the percentage reduction of candidate effectors to those which were prioritised
was similar among all isolates of *Elsinoë*, there was a wide variation seen among the
comparative species (Table 10). For example, the highest reduction was seen in
Leptosphaeria maculans, being 77.13%, while the lowest reduction was observed in
Pyrenophora tritici-repentis, at 31.96%.

Figure 20. Points system for the prioritisation pipeline for candidate effectors.
Features in blue were used to prioritise candidate effectors of all species. Additional
features were used for prioritisation if candidate effectors were from genomes with
> 2% TE coverage (red) and > 25% AT-rich region coverage (green). Note: This
figure was published as part of the current study (Jeffress et al., 2020).

**Table 10. Reduction to prioritised candidate effectors (CE) for each *Elsino*ë
and comparative species.**

Species / isolate	Candidate effectors	Prioritised CE		Reduction of CE to prioritised CE (%)
		Number	% of total CE	
Elsinoë species:				
E. australis JBS pathotype:				
Hillston-1*	233	111	47.64%	52.36%
Hillston-2*	236	107	45.34%	54.66%
Hillston-3*	245	116	47.35%	52.65%
Forbes-1*	235	108	45.96%	54.04%
Forbes-2*	238	107	44.96%	55.04%
Wagga-1*	239	108	45.19%	54.81%
Wagga-2*	238	110	46.22%	53.78%
Wagga-3*	237	107	45.15%	54.85%

***E. australis* FL pathotype**				
BRIP 52616a*	219	94	42.92%	57.08%
***E. australis* SO pathotype**				
Arg-1	235	101	42.98%	57.02%
E. fawcettii				
BRIP 53147a*	276	120	43.48%	56.52%
BRIP 54245a*	275	119	43.27%	56.73%
BRIP 54425a*	270	119	44.07%	55.93%
BRIP 54434a*	275	118	42.91%	57.09%
DAR 70024	274	120	43.80%	56.20%
SM16-1	281	118	41.99%	58.00%
E. ampelina				
YL-1	270	119	44.07%	55.93%
Comparative species:				
Necrotrophic				
Botrytis cinerea	285	89	31.23%	68.77%
Parastagonospora nodorum	932	462	49.57%	50.43%
Pyrenophora tritici-repentis	388	264	68.04%	31.96%
Sclerotinia sclerotiorum	692	379	54.77%	45.23%
Zymoseptoria tritici	597	265	44.39%	55.61%
Hemi-biotrophic				
Leptosphaeria maculans	787	180	22.87%	77.13%
Pyricularia oryzae	1,055	619	58.67%	41.33%
Rhynchosporium commune	509	146	28.68%	71.32%
Verticillium dahliae	413	202	48.91%	51.09%
Biotrophic				
Ustilago maydis	256	153	59.77%	40.23%

*Isolates sequenced in the current study
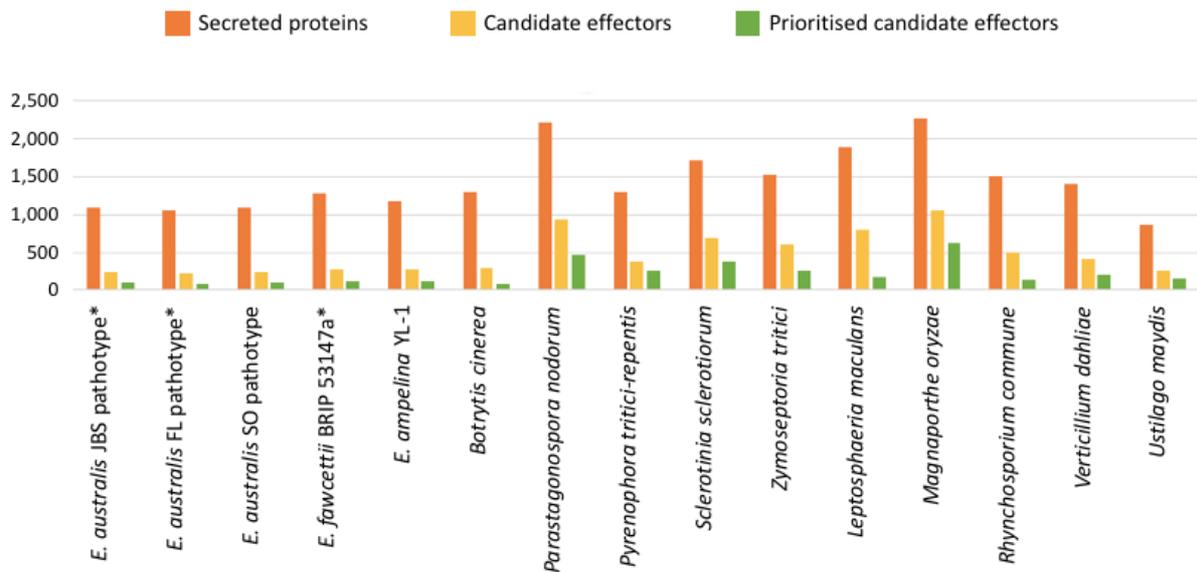
Figure 21. Secreted proteins, candidate effectors and prioritised candidate effectors among five isolates of *Elsinoë* and ten comparative fungal pathogens. Prediction of secreted proteins and candidate effectors was performed using the pathway in Figure 19. Prioritised candidate effectors were shortlisted the analysis of features listed in Figure 20.

3.2.3.4 ANALYSIS OF THE PRIORITISED *ELSINOË* CANDIDATE EFFECTORS

Analysis of the prioritised candidate effectors showed high degree of congruence among the *Elsinoë* spp. (Table 11). The total number of prioritised candidate effectors ranged from 94 (*Elsinoë australis* FL pathotype) to 120 (*E. fawcettii*). The median protein length ranged from 150 AA (*E. ampelina*) to 181 AA (*E. australis* JBS pathotype Hillston-1). The number of prioritised candidate effectors with high cysteine content (> 3%) was between 41 (38.32%) (*E. australis* JBS pathotype Wagga-3) and 50 (42.02%) (*E. fawcettii* BRIP 54425a), which was higher in comparison to the proportion of total proteins with high cysteine content, being between 4.8% (*E. ampelina*) and 3.8% (*E. australis* FL pathotype) of the predicted proteomes. A small number of prioritised candidate effectors, ranging between one (*E. fawcettii*) and nine (*E. australis* JBS pathotype Wagga-1) were predicted to target the host chloroplast. Lastly, between 6.67% (*E. fawcettii* BRIP 53147a) and 15.32% (*E. australis* JBS pathotype Hillston-1) of prioritised candidate effectors obtained hits to proteins in PHI-base.

Table 11. Analysis of the prioritised candidate effectors of the *Elsinoë* spp.

Species	Total prioritised candidate effectors (PCE)	Median protein length (AA)	High cysteine content (>3%)		Predicted to target host chloroplast		Hits to PHI-base	
			#	% of total PCE	#	% of total PCE	#	% of total PCE
Elsinoë isolates:								
E. australis JBS pathotype:								
Hillston-1*	111	181	42	37.84%	8	7.21%	17	15.32%
Hillston-2*	107	178	43	40.19%	8	7.48%	16	14.95%
Hillston-3*	116	178	46	39.66%	7	6.03%	16	13.79%
Forbes-1*	108	178	43	39.81%	8	7.41%	15	13.89%
Forbes-2*	107	178	43	40.19%	8	7.48%	15	14.02%
Wagga-1*	108	178	43	39.81%	9	8.33%	15	13.89%
Wagga-2*	110	178	44	40.00%	8	7.27%	17	15.45%
Wagga-3*	107	178	41	38.32%	8	7.48%	15	14.02%
E. australis FL pathotype:								
BRIP 52616a*	94	173	44	46.81%	6	6.38%	9	9.57%
E. australis SO pathotype:								
Arg-1	101	170	42	41.58%	5	4.95%	11	10.89%
E. fawcettii:								
BRIP 53147a*	120	166	45	37.50%	1	0.83%	8	6.67%
BRIP 54245a*	119	173	49	41.18%	1	0.84%	8	6.72%
BRIP 54425a*	119	171	50	42.02%	1	0.84%	8	6.72%
BRIP 54434a*	118	166	47	39.83%	1	0.85%	9	7.63%
DAR 70024	120	166	47	39.17%	2	1.67%	8	6.67%
SM16-1	118	163	42	35.59%	3	2.54%	9	7.63%
E. ampelina:								
YL-1	119	158	47	39.50%	3	2.52%	18	15.13%

*Isolates sequenced in the current study

3.2.4 PREDICTION OF CELL WALL DEGRADING ENZYMES

CWDE are another set of potential virulence-related genes that may be of importance during the pathogenesis of *Elsinoë* spp. CWDE prediction revealed the proteomes of the *Elsinoë* spp. consisted of between 3.64% (*E. ampelina*) and 3.96% (*E. australis* FL pathotype) potential CWDE (Table 12). Subsequent analysis indicated between 53.22% (*E. ampelina*) and 56.16% (*E. australis* FL pathotype) of predicted CWDE also contained a predictable signal peptide, highlighting approximately 200 potential virulence-related genes from the proteome of each isolate of *Elsinoë* examined that may be secreted and used to target host carbohydrates (Table 12).

Table 12. Proportions of predicted CWDE of *Elsinoë* species.

Species	Total # of predicted genes**	Predicted CWDE		Predicted secreted CWDE	
		#	% of total proteins	#	% of total predicted CWDE
E. australis **JBS pathotype:**					
Hillston-1*	8,967	341	3.80%	186	54.55%
Hillston-2*	9,533	355	3.72%	196	55.21%
Hillston-3*	9,570	356	3.72%	198	55.62%
Forbes-1*	9,465	354	3.74%	194	54.80%
Forbes-2*	9,488	356	3.75%	195	54.78%
Wagga-1*	9,547	355	3.72%	197	55.49%
Wagga-2*	9,482	355	3.74%	196	55.21%
Wagga-3*	9,521	355	3.73%	197	55.49%
E. australis **FL pathotype:**					
BRIP 52616a*	9,215	365	3.96%	205	56.16%
E. australis **SO pathotype:**					
Arg-1	9,253	356	3.85%	190	53.37%
E. fawcettii:					

BRIP 53147a*	10,080	394	3.91%	212	53.81%
BRIP 54245a*	10,164	390	3.84%	211	54.10%
BRIP 54425a*	10,014	392	3.91%	212	54.08%
BRIP 54434a*	10,016	393	3.92%	211	53.69%
DAR 70024	10,223	392	3.83%	211	53.83%
SM16-1	10,003	392	3.92%	219	55.87%
E. ampelina:					
YL-1	9,805	357	3.64%	190	53.22%

*Isolates sequenced in the current study

**Not including genes overlapping predicted TE regions

To prioritise the CWDE of the *Elsinoë* species, the predicted secreted CWDE of five genomes (*E. australis* JBS pathotype Hillston-2, *E. australis* FL pathotype, *E. australis* SO pathotype, *E. fawcettii* BRIP 53147a and *E. ampelina*) were searched against PHI-base and additionally orthologous relationships (section 3.1.3.4) were taken into account. Those which had at least one ortholog among the five *Elsinoë* species and at least one hit to a virulence-related gene of a plant pathogen (which showed evidence of reduced virulence or loss of pathogenicity) in PHI-base were prioritised. This amounted to between 21 (*E. australis* FL pathotype) and 25 (*E. ampelina*) prioritised candidate CWDE's. Looking at the orthologous relationships which existed, it was the presence of 13 orthoMCL group ID's which connected these prioritised CWDE's. These prioritised orthologous CWDE of the five *Elsinoë* species were cross referenced against the Pfam database, revealing multiple hits to pectate lyase, pectinesterase, pectin-degrading polygalacturonase (glycosyl hydrolases family 28), hemicellulose-degrading beta- xylanases (glycosyl hydrolases family 10 and family 11) and cellulose-degrading enzymes (Glycosyl hydrolase family 3 and family 5) (Table 13). Among the predicted pectin-degrading enzymes were genes with similarity to; the polygalacturonase gene *Pgx6* (66.4%) of *Fusarium oxysporum* (Bravo Ruiz et al., 2016); the *PecA* polygalacturonase gene (61.6% and 41.8%) of *Aspergillus flavus* (Shieh et al., 1997); the pectin methylesterase *Bcpme1* gene (48.0%) of *Botrytis cinerea* (Valette-Collet et al., 2003); the *PelA* and *PelD* pectate lyase genes (45.7% - 63.5%) of *Nectria haematococca* (Rogers et al., 2000); and the *Pnl1* pectin lyase gene (40.3% -

53.5%) of citrus pathogen *Penicillium digitatum* (López-Pérez et al., 2015). For the predicted hemicellulose-degrading enzymes, similarity was observed between five prioritised CWDE and the endo-1,4-beta-xylanases (glycosyl hydrolase families 10 and 11) of *Pyricularia oryzae* (syn. *Magnaporthe oryzae*) (46.7% - 61.6%) (Nguyen et al., 2011). For the predicted cellulose-degrading enzymes, there were seven prioritised CWDE; five of which showed 41.2% - 54.9% similarity to the avenacinase gene of *Gaeumannomyces graminis* (Osbourn et al., 1991); and two with similarity to the *Glu1* glucanase gene (51.9% - 52.9%) of *Pyrenophora tritici-repentis* (Fu et al., 2013).

Table 13. Prioritised candidate cell wall degrading enzymes of the *Elsinoë* species with predicted functions.

OrthoMCL Group ID	*Elsinoë fawcettii* (BRIP 53147a) GenBank Accession	PHI-base hit	Similarity (%)	Top Pfam hit	*E. australis* JBS pathotype (Hillston-2) orthologs	*E. australis* FL pathotype orthologs	*E. australis* SO pathotype orthologs	*E. ampelina* orthologs
Predicted pectin-degrading enzymes:								
OG5_149950	KAF4548260	*PGX6 Fusarium oxysporum* (PHI:4880)	66.39	Glycosyl hydrolases family 28 (GH28)	Ee.00g050350	Ee.00g042530	Ea.00g047870	Ea.00g033310
OG5_130068	KAF4550523	PECA *Aspergillus flavus* (PHI:88)	61.64	GH28	Ee.00g068310	Ee.00g079190	Ea.00g039680	Ea.00g006560
	KAF4547067	PECA *A. flavus* (PHI:88)	41.80	GH28				
OG5_133550	KAF4547800	BCPME1 *Botrytis cinerea* (PHI:278)	47.97	Pectinesterase	Ee.00g037750	Ee.00g032780	Ea.00g015350	Ea.00g088710
OG5_132299	KAF4549166	*PelD Nectria haematococca* (PHI:180)	47.27	Pectate lyase (PL)	Ee.00g005100	Ee.00g003740	Ea.00g021700	Ea.00g050210
	KAF4550092	*PelA N. haematococca* (PHI:179)	46.38	PL	Ee.00g013440	Ee.00g057430	Ea.00g049140	Ea.00g066210

	KAF4548090	*PelA N. haematococca* (PHI:179)	45.69	PL	Ee.00g062730		Ea.00g074360	Ea.00g099210
OG5_168898	KAF4552448	*PelD N. haematococca* (PHI:180)	63.45	PL	Ee.00g057150	Ee.00g030520	Ea.00g031130	Ea.00g012850
OG5_134795	KAF4549258	PNL1 *Penicillium digitatum* (PHI:3226)	53.46	Pectate lyase C	Ee.00g021020	Ee.00g041480	Ea.00g037110	Ea.00g033090
	KAF4555488	PNL1 *P. digitatum* (PHI:3226)	41.70	Pectate lyase C				
	KAF4556483	PNL1 *P. digitatum* (PHI:3226)	40.33	Pectate lyase C				
Predicted Hemicellulose-degrading enzymes:								
OG5_134270	KAF4552838	Endo-1,4-beta-xylanase *Pyricularia oryzae* (PHI:2204)	61.56	Glycosyl hydrolase family 10 (GH10)	Ee.00g062640 Ee.00g064370 Ee.00g083830	Ee.00g057530 Ee.00g074050	Ea.00g020520 Ea.00g049240 Ea.00g067710	Ea.00g033270 Ea.00g034010 Ea.00g068580 Ea.00g099130
	KAF4550100	Endo-1,4-beta-xylanase *Pyricularia oryzae* (PHI:2204)	57.69	GH10				
OG5_135408	KAF4555167	Endo-1,4-beta-xylanase	46.67	GH10	Ee.00g011180	Ee.00g016460	Ea.00g056150	Ea.00g086050

OG5	Accession	Name	Score	Family	Ee col 1	Ee col 2	Ea col 1	Ea col 2
		Pyricularia oryzae (PHI:2208)						
OG5_152744	KAF4547778	Endo-1,4-beta-xylanase I *Pyricularia oryzae* (PHI:2214)	58.87	Glycosyl hydrolases family 11 (GH11)	Ee.00g048860 Ee.00g092530	Ee.00g032360 Ee.00g069370	Ea.00g014920 Ea.00g065390	Ea.00g036880 Ea.00g081320 Ea.00g088810 Ea.00g096160
	KAF4556368	Endo-1,4-beta-xylanase I *Pyricularia oryzae* (PHI:2213)	56.72	GH11				
Predicted Cellulose-degrading enzymes:								
OG5_133540	KAF4547532	GLU1 *Pyrenophora tritici-repentis* (PHI:3859)	52.89	Cellulase - glycosyl hydrolase family 5 (GH5)	Ee.00g053060 Ee.00g086140	Ee.00g007470 Ee.00g052780	Ea.00g014070 Ea.00g060060 Ea.00g079020	Ea.00g035030 Ea.00g053860 Ea.00g090300
	KAF4552889	GLU1 *Pyrenophora tritici-repentis* (PHI:3859)	51.93	GH5				
OG5_130280	KAF4552192	Avenacinase *Gaeumannomyces graminis* (PHI:24)	46.40	Glycosyl hydrolase family 3 (GH3)	Ee.00g085000 Ee.00g095350	Ee.00g026520 Ee.00g046080	Ea.00g066560 Ea.00g085230	Ea.00g063700 Ea.00g072210

	KAF4556659	Avenacinase *Gaeumannomyces graminis* (PHI:24)	42.37	GH3				
OG5_142731	KAF4552311	Avenacinase *Gaeumannomyces graminis* (PHI:24)	54.92	GH3	Ee.00g015910 Ee.00g054480 Ee.00g076790	Ee.00g027130 Ee.00g027420 Ee.00g063970 Ee.00g085740	Ea.00g008420 Ea.00g028670 Ea.00g048930 Ea.00g090920	Ea.00g053610 Ea.00g058360
	KAF4553311	Avenacinase *Gaeumannomyces graminis* (PHI:24)	53.25	GH3				
	KAF4549450	Avenacinase *Gaeumannomyces graminis* (PHI:24)	41.23	GH3				
OG5_132365	KAF4548795	Fooch1 Fusarium oxysporum (PHI:4096)	53.23	Glycosyltransferase sugar-binding region containing DXD motif	Ee.00g031790	Ee.00g078480	Ea.00g040410	Ea.00g005700

3.3 SEARCH FOR EVIDENCE OF A TWO-SPEED GENOME IN THE *ELSINOË* SPECIES

3.3.1 ANALYSIS OF FEATURES ASSOCIATED WITH GENOME PLASTICITY

3.3.1.1 GENE-SPARSE REGIONS

Genes in gene-sparse regions were identified in all *Elsinoë* examined as genes which lay > 1500 bp from the edge of an adjacent gene or the edge of the contig, both upstream and downstream. The *Elsinoë australis* JBS pathotypes contained between 404 (Wagga-2) and 412 (Forbes-2), which was an average of 4.31% of total genes. The *E. australis* FL pathotype and SO pathotype had fewer genes in gene-sparse regions, with 283 (3.07%) and 333 (3.60%), respectively. More variation was observed among *E. fawcettii*, ranging from 358 (3.57%) in BRIP 54434a to 480 (4.7%) in DAR 70024. Finally, there were 278 (2.84%) genes found within gene-sparse locations in the *E. ampelina* assembly. Full data for the numbers and percentages of genes in gene-sparse regions is available in Appendix D Table 1.

3.3.1.2 TRANSPOSABLE ELEMENTS

Genome assembly coverage of TE regions was determined, with results detailed in section 3.1.3.2. Further to this, for the species and isolates of *Elsinoë*, the location of gene models in reference to TE regions was determined and detailed in Table 14. All *E. australis* JBS pathotypes, except for Hillston-1, showed similarity in these analyses. The Hillston-1 assembly showed only 0.36% coverage of TE regions, while the other isolates showed between 1.13% (Wagga-1) and 1.36% (Wagga-3). The number of genes which lay on a scaffold that also contained a TE region was 2,123 (23.68% of total predicted genes) for Hillston-1, while for the other isolates it ranged between 7,638 (80.55%) for the Wagga-2 isolate and 8,924 (93.73%) for the Wagga-3 isolate. The proportion of genes within 5 Kb, 10 Kb and 20 Kb of a TE region was also lower for the Hillston-1 isolate in comparison to all other JBS pathotypes, being 3.2%, 6.31% and 10.96%, respectively, for the Hillston-1 isolate. The other isolates representing *E. australis* JBS showed higher rates, with 5.31% to 6.09% of gene

models within 5 Kb of a TE region, 11.32% to 12.79% within 10 Kb, and 20.83% to 23.22% within 20 Kb (Table 14). In comparison to the majority of *E. australis* JBS pathotypes, the *E. australis* FL pathotype showed a lower degree of TE coverage, at only 0.43%, while the *E. australis* SO pathotype had a slightly higher proportion of 2.03% TE coverage. Both the *E. australis* FL and SO pathotypes had a high proportion of gene models on scaffolds with TE regions, being 93.98% and 99.76%, respectively. However, the proportion of gene models in the closer vicinity of TE region was lower when compared to the *E. australis* JBS pathotype. The *E. australis* FL pathotype showed 2.60%, 5.77% and 11.61% of gene models within 5 Kb, 10 Kb and 20 Kb, while the *E. australis* SO pathotype showed 4.01%, 7.96% and 14.78%, respectively (Table 14).

Table 14. Transposable element coverage and locality to nearby genes of the *Elsinoë* spp.

Species/isolate	Total genes (not including genes overlapping predicted TE regions)	Transposable element (TE) coverage (%)	Genes on a contig with a TE region		Genes within 5 Kb of a TE region		Genes within 10 Kb of a TE region		Genes within 20 Kb of a TE region	
			#	%	#	%	#	%	#	%
E. australis JBS pathotype:										
Hillston-1*	8,967	0.36	2123	23.68%	287	3.20%	566	6.31%	983	10.96%
Hillston-2*	9,533	1.33	8662	90.86%	562	5.90%	1176	12.34%	2174	22.80%
Hillston-3*	9,570	1.36	8895	92.95%	549	5.74%	1175	12.28%	2144	22.40%
Forbes-1*	9,465	1.29	8203	86.67%	560	5.92%	1167	12.33%	2141	22.62%
Forbes-2*	9,488	1.32	8253	86.98%	540	5.69%	1150	12.12%	2104	22.18%
Wagga-1*	9,547	1.13	8343	87.39%	507	5.31%	1081	11.32%	1989	20.83%
Wagga-2*	9,482	1.35	7638	80.55%	526	5.55%	1122	11.83%	2032	21.43%
Wagga-3*	9,521	1.19	8924	93.73%	580	6.09%	1218	12.79%	2211	23.22%
E. australis FL pathotype:										
BRIP 52616a*	9,215	0.43	8660	93.98%	240	2.60%	532	5.77%	1070	11.61%
E. australis SO pathotype:										
Arg-1	9253	2.03	9231	99.76%	371	4.01%	737	7.96%	1368	14.78%
E. fawcettii:										
BRIP 53147a*	10,080	0.37	8106	80.42%	286	2.84%	532	5.28%	962	9.54%
BRIP 54245a*	10,164	0.36	7902	77.74%	273	2.69%	515	5.07%	938	9.23%
BRIP 54425a*	10,014	0.32	7005	69.95%	245	2.45%	454	4.53%	827	8.26%

BRIP 54434a*	10,016	0.32	6920	69.09%	256	2.56%	476	4.75%	860	8.59%
DAR 70024	10,223	0.35	9738	95.26%	403	3.94%	759	7.42%	1367	13.37%
SM16-1	10,003	0.23	8317	83.15%	302	3.02%	596	5.96%	1083	10.83%
E. ampelina:										
YL-1	9804	1.59	9804	100%	862	8.79%	1987	20.27%	3670	37.43%

*Isolates sequenced in the current study

Analysis of the *E. fawcettii* genomes showed between 69.09% (BRIP 54434a) and 95.26% (DAR 70024) of gene models lay on a scaffold which also contained a TE region. Furthermore, between 245 (BRIP 54425a) and 403 (DAR 70024) genes were within 5 Kb of a TE region, between 454 (BRIP 54425a) and 759 (DAR 70024) genes were within 10 Kb and between 827 (BRIP 54425a) and 1,367 (DAR 70024) genes were within 20 Kb. Finally, analysis of the *E. ampelina* genome indicted all gene models were located on scaffolds that contained TE regions. A high proportion of *E. ampelina* genes were found in the close vicinity of TE regions, with 862 found within 5 Kb of a TE region, 1,987 genes within 10 Kb and 3,670 genes within 20 Kb.

3.3.1.3 SIMPLE SEQUENCE REPEATS

The search for SSR within CDS regions of the *Elsinoë* genomes indicated a low proportion of genes, between 1.62% and 3.42%, contained SSR. *E. fawcettii* had the lowest rate, with between 1.62% (BRIP 54245a) and 1.78% (SM16-1). All *E. australis* genomes were closely aligned in proportion of SSR-containing genes, ranging from 2.58% (*E. australis* FL pathotype) to 2.86% (*E. australis* JBS pathotype Wagga-1). Lastly, *E. ampelina* had the highest proportion of all *Elsinoë* examined, with 3.42% of genes containing an SSR (Table 15). Additionally, a search for polyAA repeats (five or more identical and adjacent residues) was performed throughout the proteomes of all genomes. The lowest rate was recorded for *E. fawcettii*, having between 10.56% (SM16-1) and 10.67% (BRIP 54425a) of polyAA-containing genes. The *E. australis* pathotypes contained between 11.35% (*E. australis* JBS pathotype Hillston-1) to 11.89% (*E. australis* JBS pathotype Wagga-1) genes with polyAA repeats. Finally, *E. ampelina* was found to have the highest rate, with 12.14% polyAA-containing genes.

Table 15. Genes of the *Elsinoë* spp. containing poly-amino acid (polyAA) repeats and Simple Sequence Repeats (SSR).

Species/isolate	Total genes (not including genes overlapping predicted TE regions)	Genes containing polyAA repeats		Genes containing SSR	
		#	%	#	%
E. australis **JBS pathotype:**					
Hillston-1*	8,967	1,018	11.35%	239	2.67%
Hillston-2*	9,533	1,128	11.83%	258	2.71%
Hillston-3*	9,570	1,121	11.71%	267	2.79%
Forbes-1*	9,465	1,122	11.85%	255	2.69%
Forbes-2*	9,488	1,127	11.88%	262	2.76%
Wagga-1*	9,547	1,135	11.89%	273	2.86%
Wagga-2*	9,482	1,120	11.81%	259	2.73%
Wagga-3*	9,521	1,127	11.84%	263	2.76%
E. australis **FL pathotype:**					
BRIP 52616a*	9,215	1,094	11.87%	238	2.58%
E. australis **SO pathotype:**					
Arg-1	9253	1,093	11.86%	241	2.60%
E. fawcettii:					
BRIP 53147a*	10080	1,073	10.64%	164	1.63%
BRIP 54245a*	10,164	1,074	10.57%	165	1.62%
BRIP 54425a*	10,014	1,068	10.67%	165	1.65%
BRIP 54434a*	10,016	1,068	10.66%	165	1.65%
DAR 70024	10223	1,083	10.59%	167	1.63%
SM16-1	10,003	1,056	10.56%	178	1.78%
E. ampelina:					
YL-1	9804	1,190	12.14%	335	3.42%

*Isolates sequenced in the current study

3.3.1.4 AT-RICH REGIONS

While a search for AT-rich regions within the *Elsinoë* genomes was conducted in section 3.1.3.3, further analysis was performed to establish where these AT-rich

regions were located within the genome assembly, if they formed a part of the scaffolds which contained genes and how close genes were to the edges of the AT-rich regions. The coverage of AT-rich regions among the assemblies of *Elsinoë* examined varied greatly, from not detected (four isolates out of six for *E. fawcettii*: BRIP 54245a, BRIP 54425a, BRIP 54434a and SM16-1) to 17.9% (*E. ampelina*) coverage, details can be found in section 3.1.3.3.

AT-rich regions were observed within the assemblies of two *E. fawcettii* (BRIP 53147a and DAR 70024) with low coverage (1.03% and 0.88%, respectively). Eight genes of BRIP 53147a and nine genes of DAR 70024 were found to overlap an AT-rich region. A total of 57 (0.56%), 112 (1.11%), 191 (1.89%) and 348 (3.45%) BRIP 53147a genes were found within 2 Kb, 5 Kb, 10 Kb and 20 Kb, respectively, with comparable numbers found in the DAR 70024 assembly (Table 16). Similar proportions were found in the *E. australis* SO pathotype, being 0.42%, 0.96%, 1.79% and 3.52%, found within 2 Kb, 5 Kb, 10 Kb and 20 Kb of an AT-rich region, respectively. The *E. australis* FL pathotype had a slightly increased number of genes in all regions, with 102 (1.11%), 209 (2.27%), 359 (3.90% and 654 (7.10%) genes at increasing distances from the AT-rich regions (Table 16). The *E. australis* JBS pathotypes showed higher proportions again. Using the Hillston-2 isolate as an example, 25 genes were found to overlap AT-rich regions, while 398 (4.17%), 768 (8.06%), 1,324 (13.89%) and 2,216 (23.25%) were observed within 2 Kb, 5 Kb, 10 Kb and 20 Kb. *E. ampelina* had the highest number of genes overlapping an AT-rich region, with 73, and also the highest numbers of genes at increasing distances, with 878 (8.96%), 1,724 (17.58%), 2,917 (29.75%) and. 4,699 (47.93%) within 2 Kb, 5 Kb, 10 Kb and 20 Kb (Table 16).

Table 16. Distance of genes of the *Elsinoë* spp. to AT-rich genomic regions.

Species/isolate	Total genes (not including genes overlapping predicted TE regions)	Number of AT-rich regions	AT-rich region coverage (%)	Genes overlapping an AT-rich region	Genes within 2 Kb of an AT-rich region**		Genes within 5 Kb of an AT-rich region**		Genes within 10 Kb of an AT-rich region**		Genes within 20 Kb of an AT-rich region**	
					#	%	#	%	#	%	#	%
E. australis JBS pathotype:												
Hillston-1*	8,967	342	3.4	7	184	2.05%	358	3.99%	622	6.94%	1030	11.49%
Hillston-2*	9,533	313	9.13	25	398	4.17%	768	8.06%	1324	13.89%	2216	23.25%
Hillston-3*	9,570	348	9.67	30	401	4.19%	782	8.17%	1341	14.01%	2280	23.82%
Forbes-1*	9,465	368	9.11	21	390	4.12%	760	8.03%	1313	13.87%	2195	23.19%
Forbes-2*	9,488	339	9.33	27	417	4.40%	801	8.44%	1385	14.60%	2334	24.60%
Wagga-1*	9,547	391	8.28	21	410	4.29%	810	8.48%	1387	14.53%	2314	24.24%
Wagga-2*	9,482	341	9.23	37	417	4.40%	794	8.37%	1354	14.28%	2246	23.69%
Wagga-3*	9,521	338	8.06	18	407	4.27%	788	8.28%	1351	14.19%	2262	23.76%
E. australis FL pathotype:												
BRIP 52616a*	9,215	81	2.48	11	102	1.11%	209	2.27%	359	3.90%	654	7.10%
E. australis SO pathotype:												
Arg-1	9253	32	0.92	3	39	0.42%	89	0.96%	166	1.79%	326	3.52%
E. fawcettii:												
BRIP 53147a*	10080	59	1.03	8	57	0.56%	112	1.11%	191	1.89%	348	3.45%
BRIP 54245a*	10,164	0	0	-	-	-	-	-	-	-	-	-
BRIP 54425a*	10,014	0	0	-	-	-	-	-	-	-	-	-

BRIP 54434a*	10,016	0	0	-	-	-	-	-	-	-	-	-
DAR 70024	10223	36	0.88	9	52	0.51%	103	1.01%	172	1.68%	327	3.20%
SM16-1	10,003	0	0	-	-	-	-	-	-	-	-	-
E. ampelina:												
YL-1	9804	380	17.9	73	878	8.96%	1724	17.58%	2917	29.75%	4699	47.93%

*Isolates sequenced in the current study
**Including genes overlapping an AT-rich region

3.3.1.5 FRAGMENTED GENOMIC REGIONS

The degree of genome fragmentation among the *Elsinoë* assemblies varied widely. The *Elsinoë ampelina* assembly was made up of 13 scaffolds and represented the most complete assembly. The *E. australis* SO pathotype, *E. fawcettii* DAR 70024 and *E. australis* SO pathotype were slightly fragmented in comparison with 21, 53 and 61 scaffolds, respectively. The *E. fawcettii* BRIP 53147a assembly was made up of 286 scaffolds. Three *E. fawcettii* assemblies (BRIP 54245a, BRIP 54425a and BRIP 54434a) and five *E. australis* JBS pathotype assemblies (Hillston-2, Forbes-2, Wagga-1, Wagga-2 and Wagga-3) contained between 438 and 495 scaffolds. Slightly more fragmented than most *Elsinoë* assemblies were the remaining two *E. australis* JBS pathotype assemblies (Hillston-3 and Forbes-1). Finally, *E. fawcettii* SM16-1 and *E. australis* JBS pathotype Hillston-1 were the most fragmented of the assemblies with 1,266 and 1,378 scaffolds, respectively. While the majority of the scaffolds making up these genomes were particularly fragmented (< 20 Kb in length) (Figure 22 and Appendix E Table 1), very few genes were located in these regions. The highly fragmented assembly of *E. australis* JBS pathotype Hillston-1 contained the most genes, being 497 (5.54%) on scaffolds < 20 Kb in length. This was presumed to be an outlier, as < 2% of genes were located on these short scaffolds for all other *Elsinoë* assemblies (Table 17). Genes located at a scaffold edge were also identified and again made up only a small subset, with all *Elsinoë* examined being between 0.27% (*E. ampelina*) and 3.31% (*E. australis* JBS pathotype Wagga-2), except for *E. australis* JBS pathotype Hillston-1 , again as an outlier, at 10.75% (Table 17).
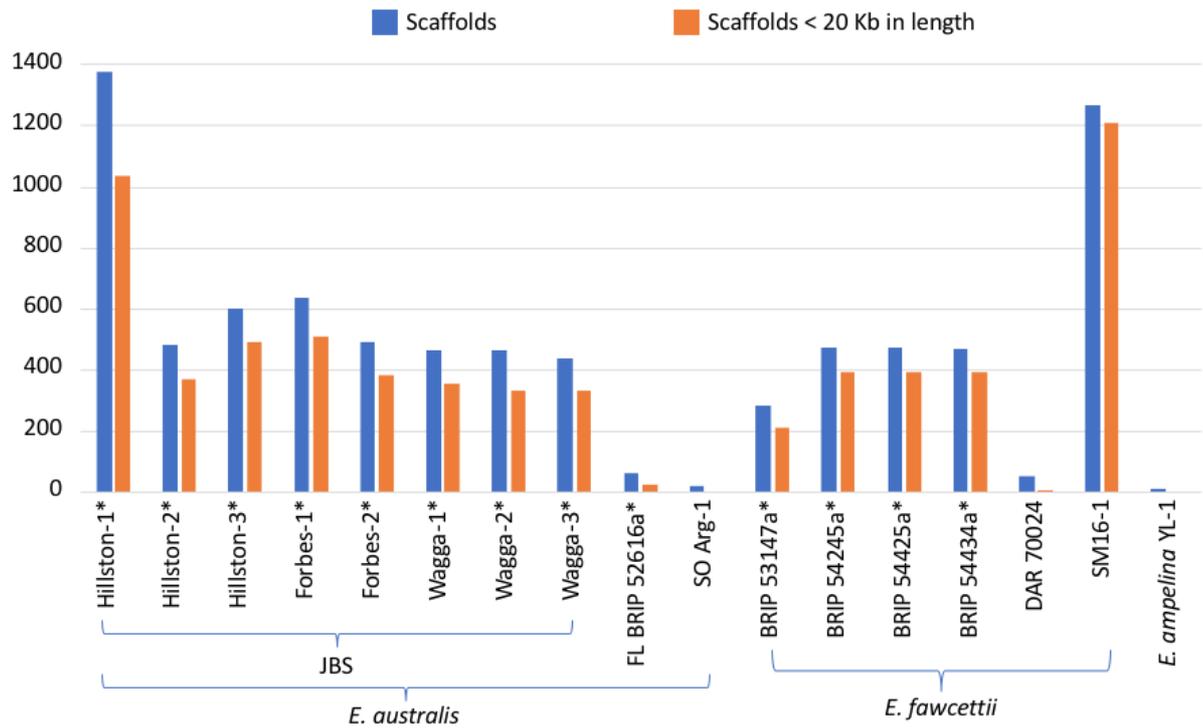
Figure 22. Level of fragmentation of genome assemblies for the *Elsino*ë spp. The total number of scaffolds in an assembly (blue) compared with the number of scaffolds < 20 Kb in length (orange). Full data available in Appendix E Table 1.

Table 17. Proportions of genes of the *Elsino*ë spp. in fragmented genomic regions.

Species/isolate	Total genes (not including genes overlapping predicted TE regions)	Genes on scaffolds < 20 Kb		Genes on scaffold edge	
		#	%	#	%
E. australis **JBS pathotype:**					
Hillston-1*	8,967	497	5.54%	964	10.75%
Hillston-2*	9,533	67	0.70%	253	2.65%
Hillston-3*	9,570	23	0.24%	230	2.40%
Forbes-1*	9,465	91	0.96%	293	3.10%

Forbes-2*	9,488	71	0.75%	259	2.73%
Wagga-1*	9,547	71	0.74%	251	2.63%
Wagga-2*	9,482	101	1.07%	314	3.31%
Wagga-3*	9,521	52	0.55%	230	2.42%
E. australis **FL pathotype:**					
BRIP 52616a*	9,215	8	0.09%	73	0.79%
E. australis **SO pathotype:**					
Arg-1	9253	7	0.08%	40	0.43%
E. fawcettii:					
BRIP 53147a*	10080	130	1.29%	252	2.50%
BRIP 54245a*	10,164	182	1.79%	291	2.86%
BRIP 54425a*	10,014	154	1.54%	279	2.79%
BRIP 54434a*	10,016	152	1.52%	283	2.83%
DAR 70024	10223	27	0.26%	101	0.99%
SM16-1	10,003	100	1.00%	168	1.68%
E. ampelina:					
YL-1	9804	0	0.00%	26	0.27%

*Isolates sequenced in the current study

3.3.1.6 SINGLE NUCLEOTIDE POLYMORPHISMS

SNP were determined for *Elsinoë fawcettii* (BRIP 53147a) in comparison to three isolates of *E. fawcettii* (BRIP 54245a, BRIP 54425a and BRIP 54434a). In total, 5,972 SNP were identified across the genome, of which 1,027 were found within a predicted gene model. A total of 340 genes contained at least one SNP, of which 76 contained more than one. The highest number of SNP counted within one gene model was 62, located within KAF4547161, a gene model coding for a hypothetical protein 531 amino acids in length and found <5kb from a TE region. The 340 genes containing at least one SNP were found scattered over 93 contigs. 126 were found in gene-sparse locations, while 67 were found in gene-dense locations.

Similarly, SNP were identified in the *E. australis* JBS pathotype Hillston-2, by comparison with reads from several JBS pathotypes (Forbes-1, Wagga-2

and Forbes-2). The total number of SNP found throughout the Hillston-2 assembly was 7,297, with 826 located within the CDS of a gene model. This amounted to 375 predicted genes with at least one SNP, which were again scattered throughout the genome over 80 scaffolds. Almost half of these genes (46.54%) contained only one SNP. The greatest number of SNP with a single gene was 11.

3.3.2 CANDIDATE VIRULENCE GENE LOCATION WITH RESPECT TO REGIONS WITH POTENTIAL GENOME PLASTICITY

The genomic location of candidate virulence-associated genes of five isolates of *Elsinoë* were analysed for locality in gene-sparse regions, proximity to TE regions and AT-rich regions, repetitive sequences and locality to fragmented genomic regions. Candidate virulence-associated genes, including genes involved in SM clusters, secreted proteins, candidate effectors, CWDE and *Elsinoë*-specific genes, were analysed. The isolates included in the analysis were *E. australis* JBS pathotype Hillston-2, *E. australis* FL pathotype BRIP 52616a, *E. australis* SO pathotype Arg-1, *E. fawcettii* BRIP 53147a and *E. ampelina* YL-1. The potential virulence-associated genes of both *E. fawcettii* BRIP 53147a and *E. australis* JBS pathotype Hillston-2 were also analysed for the inclusion of SNP.

3.3.2.1 CANDIDATE VIRULENCE GENES IN GENE-SPARSE GENOMIC LOCATIONS

Potential virulence-associated genes were cross-referenced with those found in gene-sparse locations. Genes in gene-sparse locations accounted for 4.28% (*Elsinoë australis* JBS pathotype Hillston-2), 3.07% (*E. australis* FL pathotype), 3.60% (*E. australis* SO pathotype), 4.05% (*E. fawcettii* BRIP 53147a) and 2.84% (*E. ampelina*) of total predicted genes (Figure 23). Genes in gene-sparse locations were found to be overrepresented among candidate effectors of three of the five isolates of *Elsinoë* examined,

accounting for 6.78% (*p*=0.04630) of candidate effectors in *E. australis* JBS pathotype, 8.33% (*p*=0.00082) in *E. fawcettii* BRIP 53147a and 5.56% (*p*=0.00997) in *E. ampelina*. *Elsinoë*-specific genes were also overrepresented among genes found in gene-sparse locations of all isolates tested, aside from *E. ampelina*. A total of 9.41% (*p*=6.01449e-24), 5.43% (*p*=1.57438e-08), 6.46% (*p*=1.64322e-08) and 9.35% (*p*=1.33508e-31) of genes in gene-sparse locations were also *Elsinoë*-specific genes of *E. australis* JBS pathotype, *E. australis* FL pathotype, *E. australis* SO pathotype and *E. fawcettii*, respectively.
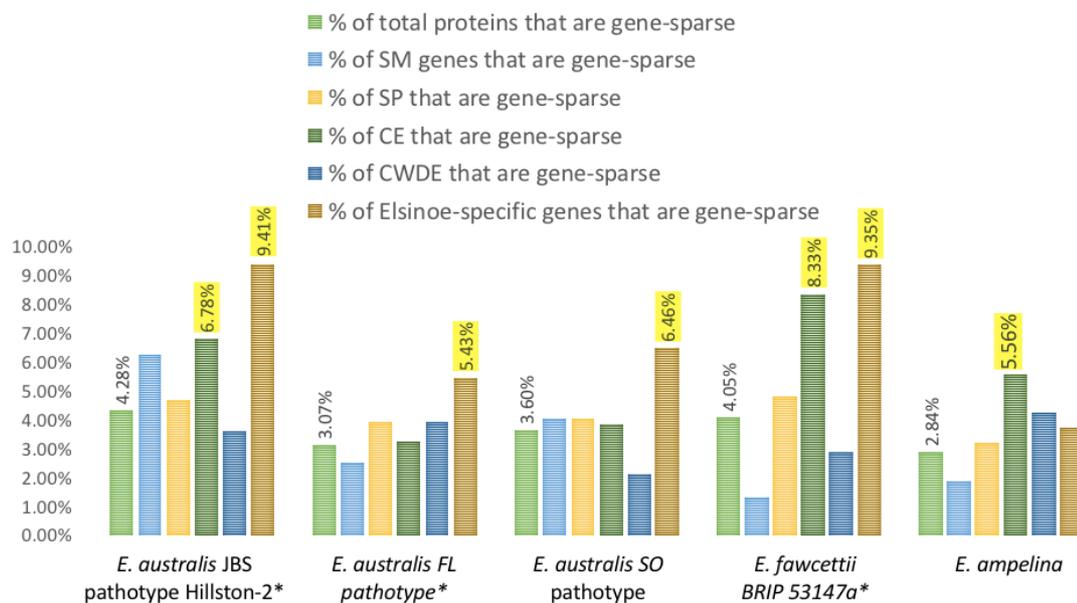
Figure 23. Proportions of candidate virulence genes in gene-sparse locations of the *Elsinoë* spp. The proportion of genes found in gene-sparse locations out of all predicted genes is shown on the left for each species (light green). Genes involved in predicted secondary metabolite (SM) gene clusters (light blue), predicted secreted proteins (SP) (yellow), candidate effectors (CE) (dark green), cell wall degrading enzymes (CWDE) (dark blue) and *Elsinoë*-specific genes (brown) are illustrated. Proportions of overrepresented candidate virulence genes in gene-sparse locations, that are statistically significant, are highlighted yellow. Full, including p values for statistically significant proportions, is available in Appendix E Table 2.

3.3.2.2 CANDIDATE VIRULENCE GENES NEARBY TO TRANSPOSABLE ELEMENT REGIONS

The proportions of candidate virulence genes located within 20 Kb of a TE region were identified. Those which were found to be at statistically significant levels are highlighted yellow in Figure 24. The baseline proportions of total genes that were < 20 Kb from a TE region were 22.80% (*Elsinoë australis* JBS pathotype), 11.61% (*E. australis* FL pathotype), 14.78% (*E. australis* SO pathotype), 9.54% (*E. fawcettii*) and 37.43% (*E. ampelina*). Genes < 20 Kb from a TE region and which were also involved in SM gene clusters were elevated among the *E. australis* FL and SO pathotypes, being 15.92% (p=0.03887) and 19.50% (p=0.03960), respectively. Secreted proteins were only overrepresented in the *E. australis* FL pathotype, at 14.19% (p=0.00377). Proportions of candidate effectors among genes close to TE regions were higher in both the *E. australis* SO pathotype (19.15%, p=0.03796) and *E. ampelina* (44.81%, p=0.00702). CWDE were overrepresented in the *E. australis* JBS pathotype, at 30.61% (p=0.00663). Lastly, proportions of *Elsinoë*-specific genes were increased in three of the five isolates, being *E. australis* JBS pathotype (27.49%, p=7.42819e-07), *E. fawcettii* (12.72, p=2.56171e-07) and *E. ampelina* (40.53, p=0.00143).
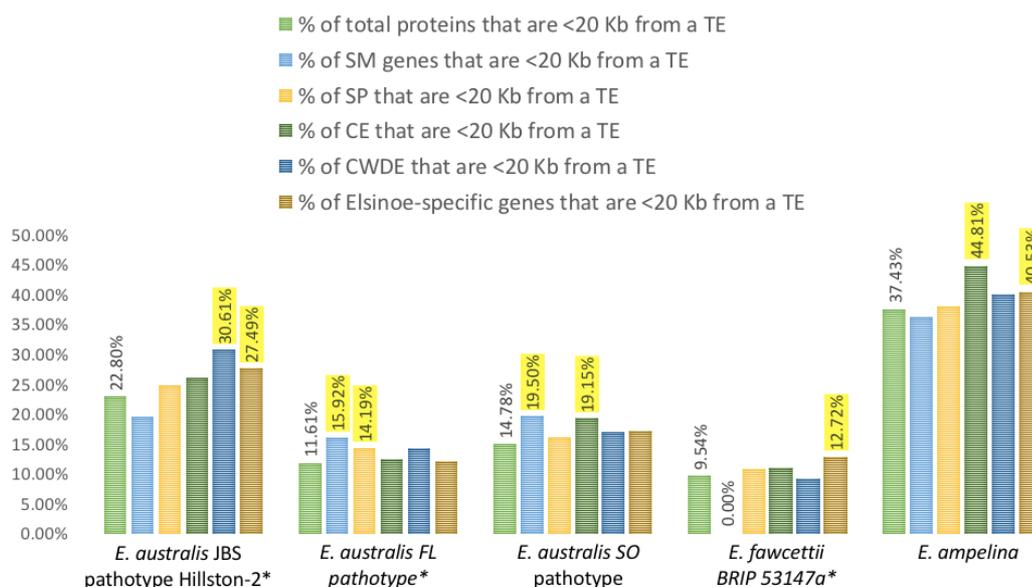
Figure 24. Proportions of candidate virulence genes of the *Elsinoë* spp. located within 20 Kb of a TE region. The proportion of total predicted genes < 20 Kb from a TE region is shown on the left for each species (light green). Gene involved in predicted secondary metabolite (SM) gene clusters (light blue), predicted secreted proteins (SP) (yellow), candidate effectors (CE) (dark green), cell wall-degrading enzymes (CWDE) (dark blue) and *Elsinoë*-specific genes (brown) are illustrated. Genes within 20 Kb of a TE region that are overrepresented within their candidate virulence gene group, at a statistically significant level, are highlighted yellow. Full data, including *p* values for statistically significant proportions, is available in Appendix E Table 3.

3.3.2.3 CANDIDATE VIRULENCE GENES CONTAINING SHORT REPETITIVE SEQUENCES

Potential virulence-associated genes were cross-referenced with those containing short repetitive regions, including SSR and polyAA repeats. Genes containing either or both of these short repetitive regions accounted for 12.22% (*Elsinoë australis* JBS pathotype Hillston-2), 12.22% (*E. australis* FL pathotype BRIP 52616a), 12.29% (*E. australis* SO pathotype Arg-1), 10.96% (*E. fawcettii* BRIP 53147a) and 12.95% (*E. ampelina*) of total

predicted genes (Figure 25). Genes containing short repetitive regions were found to be overrepresented only among the *Elsinoë*-specific genes of four of the five *Elsinoë* tested, being the *E. australis* JBS pathotype (15.48%, *p*=1.13710e-05), *E. australis* FL pathotype (18.74% (*p*=4.92263e-17), *E. australis* SO pathotype (15.85% *p*=1.93040e-05), and *E. fawcettii* (14.84%, *p*=4.38589e-09).
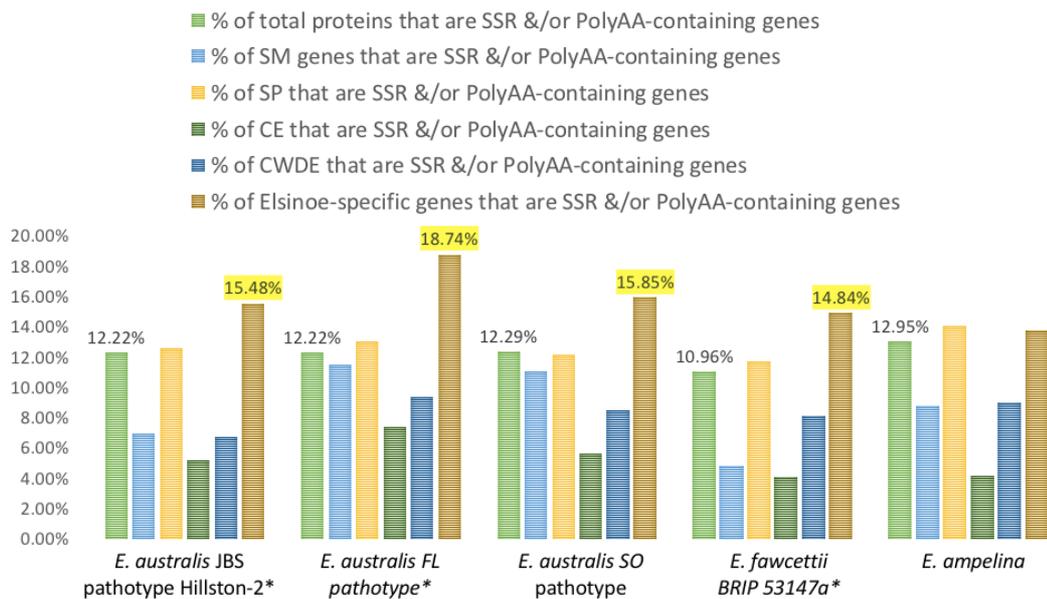
Figure 25. Proportions of candidate virulence genes of the *Elsinoë* spp. which contain an SSR and/or a polyAA. The proportion of total predicted genes that contain an SSR and/or a polyAA is shown on the left for each species (light green). Genes involved in predicted secondary metabolite (SM) gene clusters (light blue), predicted secreted proteins (SP) (yellow), candidate effectors (CE) (dark green), cell wall-degrading enzymes (CWDE) (dark blue) and *Elsinoë*-specific genes (brown) are illustrated. SSR- and/or polyAA-containing genes which are overrepresented within their candidate virulence gene group, at a statistically significant level, are highlighted yellow. Full data, including *p* values for statistically significant proportions, is available in Appendix E Table 4.

3.3.2.4 CANDIDATE VIRULENCE GENES NEARBY TO AT-RICH REGIONS

The proportions of candidate virulence genes located within 20 Kb of an AT-rich region were identified, with statistically significant proportions highlighted in yellow within Figure 26. For comparison, total genes that were < 20 Kb from an AT-rich region were 23.25% (*Elsinoë australis* JBS pathotype), 7.10% (*E. australis* FL pathotype), 3.52% (*E. australis* SO pathotype), 3.45% (*E. fawcettii*) and 47.93% (*E. ampelina*) (Figure 26). Genes < 20 Kb from an AT-rich region which were also involved in SM clusters were elevated only in the *E. australis* FL, at 20.90% (p=1.28387e-10). Secreted proteins were overrepresented among all three *E. australis* pathotypes and *E. fawcettii*, at 28.22% (p=2.79901e-05) (JBS pathotype), 10.79% (p=1.77967e-06) (FL pathotype), 5.04% (p=0.00358) (SO pathotype) and 5.23% (p=0.000266) (*E. fawcettii*). Proportions of candidate effectors among genes close to AT-rich regions were higher in the *E. australis* FL pathotype (13.24%, p=0.00079), *E. australis* SO pathotype (6.81%, p=0.00887) and *E. ampelina* (55.56%, p=0.00654). CWDE were overrepresented in all five isolates tested, accounting for 36.22% (p=2.38834e-05) (*E. australis* JBS pathotype) 13.66% (p=0.00059) (*E. australis* FL pathotype), 6.84% (p=0.01678) (*E. australis* SO pathotype), 6.60% (p=0.01518) (*E. fawcettii*) and 60.00% (p=0.00049) (*E. ampelina*) of genes < 20 Kb of an AT-rich region. Lastly, proportions of *Elsinoë*-specific genes located close to AT-rich regions were increased in only the *E. australis* JBS pathotype (28.67%, p=1.64244e-08).
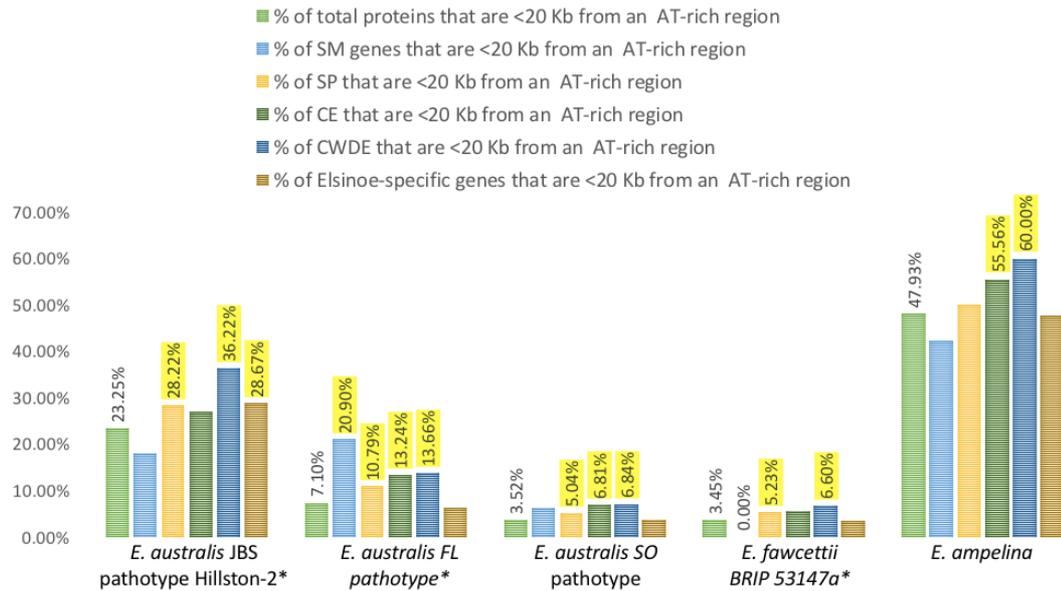
Figure 26. Proportions of candidate virulence genes of the *Elsinoë* spp. which are within 20 Kb of an AT-rich region. The proportion of total predicted genes that are within 20 Kb of an AT-rich region is shown on the left for each species (light green). Genes involved in predicted secondary metabolite (SM) gene clusters (light blue), predicted secreted proteins (SP) (yellow), candidate effectors (CE) (dark green), cell wall-degrading enzymes (CWDE) (dark blue) and *Elsinoë*-specific genes (brown) are illustrated. Genes within 20 Kb of an AT-rich region which are overrepresented within their candidate virulence gene group, at a statistically significant level, are highlighted yellow. Full data, including *p* values for statistically significant proportions, is available in Appendix E Table 5.

Table 18. Location of AT-rich regions with respect to TE regions among the *Elsinoë* spp.

Species/Pathotype	Number of AT-rich regions	Number of AT-rich regions >2 Kb from a TE or not on a contig with a TE region	Number of AT-rich regions which overlap or a <2Kb from a TE region
E. australis JBS pathotype Hillston-2*	313	118 (37.7%)	195 (62.3%)
E. australis FL pathotype BRIP 52616a*	81	40 (49.4%)	41 (50.6%)
E. australis SO pathotype Arg-1	32	16 (50.0%)	16(50.0%)
E. fawcettii BRIP 53147a*	59	38 (64.4%)	21 (35.6%)
E. ampelina YL-1	380	139 (36.6%)	241 (63.4%)

*Isolates sequenced in the current study

3.3.2.5 CANDIDATE VIRULENCE GENES FRAGMENTED GENOMIC LOCATIONS

Potential virulence-associated genes were cross-referenced with those in fragmented regions of the assemblies, including those located at the edge of a scaffold and those contained within scaffolds < 20Kb in length. Firstly, the total proportion of genes found on scaffolds < 20Kb in length were 0.70% (*Elsinoë australis* JBS pathotype), 0.09% (*E. australis* FL pathotype), 0.08% (*E. australis* SO pathotype) and 1.29% (*E. fawcettii*) of total predicted genes (Figure 27). These genes found on short scaffolds were found to be overrepresented only among CWDE of *E. australis* FL pathotype (0.98%, p=0.01262), and the *Elsinoë*-specific genes of *E. australis* JBS pathotype (1.61%, p=1.59040e-05) and *E. fawcettii* (3.33%, p=1.14938e-14).
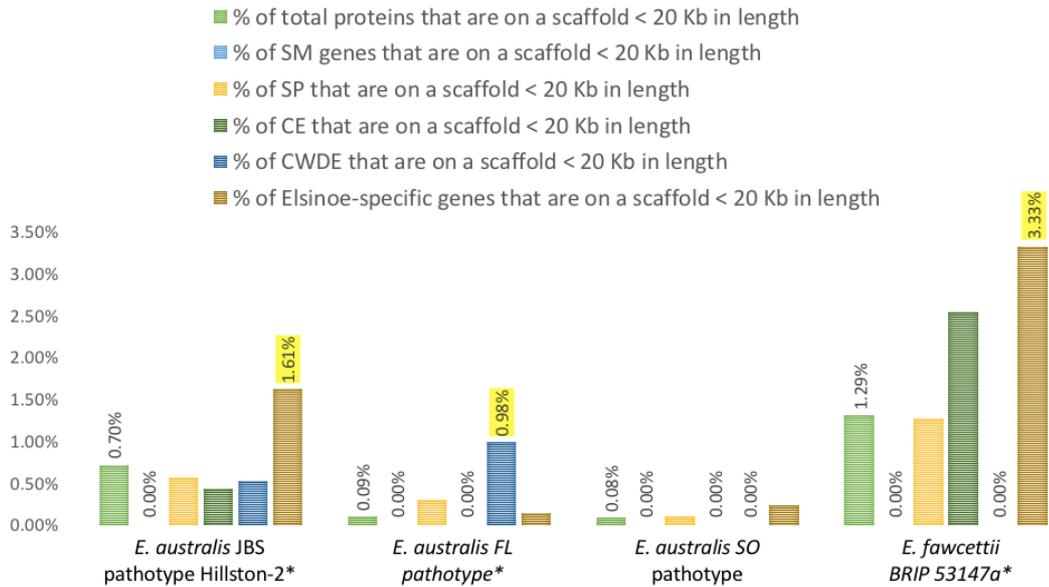
Figure 27. Proportions of candidate virulence genes of the *Elsinoë* spp. on scaffolds < 20 Kb in length. The proportion of total predicted genes found on scaffolds < 20 Kb in length. is shown on the left for each species (light green). Genes involved in predicted secondary metabolite (SM) gene clusters (light blue), predicted secreted proteins (SP) (yellow), candidate effectors (CE) (dark green), cell wall-degrading enzymes (CWDE) (dark blue) and *Elsinoë*-specific genes (brown) are illustrated. Genes on scaffolds < 20 Kb in length which are overrepresented within their candidate virulence gene group, at a statistically significant level, are highlighted yellow. Full data, including *p* values for statistically significant proportions, is available in Appendix E Table 6.

Secondly, the proportions of candidate virulence genes located on the edge of a scaffold, with statistically significant proportions, are highlighted yellow in Figure 28. Proportions of total genes found at the edge of a scaffold were 2.65% (*E. australis* JBS pathotype), 0.79% (*E. australis* FL pathotype), 0.43% (*E. australis* SO pathotype), 2.50% (*E. fawcettii*) and 0.27% (*E. ampelina*). Secreted proteins were elevated in the *E. australis* JBS pathotype (3.56%, *p*=0.03338), *E. australis* FL pathotype (1.61%, *p*=0.003112) and *E. ampelina* (0.60%, *p*=0.02852). Proportions of candidate

effectors among genes on the edge of a scaffold were higher in the *E. australis* JBS pathotype (5.93%, *p*=0.00410), *E. fawcettii* (5.07%, *p*=0.00946) and *E. ampelina* (1.48%, *p*=0.00521). Lastly, proportions of *Elsinoë*-specific genes located close to the edge of a scaffold were increased in *E. australis* JBS pathotype (4.77%, *p*=5.80495e-08), *E. australis* SO pathotype (0.98%, *p*=5.80495e-08) and *E. fawcettii* (5.76%, *p*=1.25758e-19).
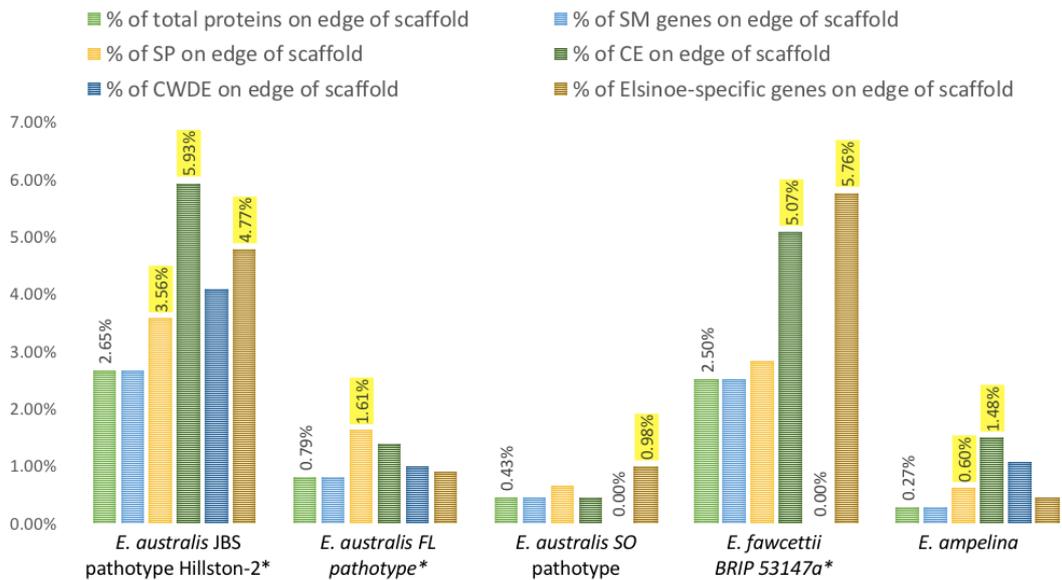
Figure 28. Proportions of candidate virulence genes of the *Elsinoë* spp. located on the edge of a scaffold. The proportion of total predicted genes found on the edge of a scaffold is shown on the left for each species (light green). Genes involved in predicted secondary metabolite (SM) gene clusters (light blue), predicted secreted proteins (SP) (yellow), candidate effectors (CE) (dark green), cell wall-degrading enzymes (CWDE) (dark blue) and *Elsinoë*-specific genes (brown) are illustrated. Genes located on the edge of a scaffold which are overrepresented within their candidate virulence gene group, at a statistically significant level, are highlighted yellow. Full data, including *p* values for statistically significant proportions, is available in Appendix E Table 7.

3.3.2.6 CANDIDATE VIRULENCE GENES CONTAINING SNP

Potential virulence-associated genes of *Elsinoë fawcettii* (BRIP 53147a) were cross-referenced with those containing SNP among the three other isolates of *E. fawcettii* examined (BRIP 54245a, BRIP 54425a and BRIP 54434a). In total, 3.37% of predicted *E. fawcettii* BRIP 53147a genes contained at least one SNP, however none of the potential virulence-related gene groups contained significantly higher proportions than this total proportion (Table 19). Similarly, predicted virulence-associated genes of *E. australis* JBS pathotype Hillston-2 which contained SNP among several JBS pathotype isolates (Forbes-1, Wagga-2 and Forbes-2) were identified. However, as for *E. fawcettii*, no groups of potential virulence-related genes contained significantly higher proportions of SNP (Table 19).

Table 19. Proportions of candidate virulence genes of *E. fawcettii* and *E. australis* JBS pathotype containing Single Nucleotide Polymorphisms.

Elsinoë species	*E. fawcettii* BRIP 53147a*	*E. australis* JBS pathotype Hillston-2*
Total genes	10080	9533
Total genes with Single Nucleotide Polymorphisms (SNP)	340	375
% of total genes with SNP	3.37%	3.93%
Genes involved in predicted Secondary Metabolite (SM) clusters	233	145
SM genes with SNP	5	4
% of SM genes with SNP	2.15%	2.76%
Secreted proteins (SP)	1280	1095
SP with SNP	36	33
% of SP with SNP	2.81%	3.01%
Candidate effectors (CE)	276	236
CE with SNP	6	5

% of CE with SNP	2.17%	2.12%
Cell wall degrading enzymes (CWDE)	212	196
CWDE with SNP	4	3
% of CWDE with SNP	1.89%	1.53%
Elsinoë*-specific genes	1894	1615
***Elsinoë*-specific genes with SNP**	76	59
% of *Elsinoë*-specific genes with SNP	4.01%	3.65%

*Sequenced in the current study

**Predicted secreted CAZymes

***Unique among the ten comparative species and no orthoMCL ID obtained

CHAPTER 4: DISCUSSION

4.1 GENOMIC ANALYSES

The *Elsinoë* genomes sequenced in the current study were similar in size to one another, and to the recently published *Elsinoë* genomes (Shanmugam et al., 2020), ranging between 23.8 Mb (*E. australis* FL pathotype) and 26.6 Mb (*E. australis* JBS pathotype) (Table 1, Figure 9). However, the assemblies were smaller than the average genome size of an Ascomycota species (36.91 Mb) (Mohanta & Bae, 2015) and contained low proportions of TE regions (Table 2). As low TE coverage can be a consequence of a fragmented genome due to short read sequencing (Thomma et al., 2016), TE coverage may potentially be improved with further genome assemblies generated using long read sequencing. While the assembly for *E. ampelina* produced using long read sequencing (Li et al., 2020) contained a higher TE coverage than all other *Elsinoë*, it was still relatively low. Degree of fragmentation varied among the *Elsinoë* genomes (Figure 12), yet TE coverage was low among all *Elsinoë*, suggesting TE were not proliferative within the *Elsinoë* genomes. In general, the *Elsinoë* genomes were more fragmented in comparison to most other fungal pathogen genomes analysed, with the exception of *Pyrenophora tritici-repentis* and *Leptosphaeria maculans* (Figure 12). This was likely due to differences in genome sequencing technology. For example, the *Elsinoë* genomes, with the exception of *E. ampelina*, were sequenced using short read Illumina technology. The overall GC content of the *Elsinoë* genomes ranged between 49.5% for *E. ampelina* and 52.6% for the *E. australis* SO pathotype. AT-rich regions likely contributed towards this variance across the assemblies, for example *E. ampelina* had the highest degree of AT-rich regions (17.9%), while the SO pathotype had one of the lowest (0.9%) (Table 3). The majority of AT-rich regions were located in intergenic regions, a pattern which is seen in the genomes of other species which contain AT-rich regions. For example, in *Leptosphaeria maculans* only 5% of predicted genes were found within AT-rich regions (Rouxel et al., 2011) and similarly only 3.6% of *Epichloë festucae* genes were found to overlap an AT-rich region (Winter et al., 2018). An effect of intergenic regions housing the majority of AT-rich regions was

observed in the mean GC content of CDS of each *Elsinoë*. These were more closely aligned with one another, ranging only between 54.1% and 54.9%. There was, however, a wide variety of GC content between individual genes within each species. For example, the genes of *E. fawcettii* BRIP 53147a ranged between 44.3% and 71.5%, while genes of *E. australis* Hillston-2 lay between 44.0% and 72.9%. A similarly wide variety was observed among all *Elsinoë* and comparative species. This attribute is frequently seen within the proteomes of fungal species, as noted in the comparative analysis of 443 fungal organisms by Wilken et al. (2020) This observed variation in GC content among individual genes was utilised in downstream analyses for candidate effector prioritisation, as both high and low, as opposed to average, proportions of GC content was identified in the known effectors analysed.

Multiple AT-rich regions in the *Elsinoë* genomes were found to overlap TE regions and further AT-rich sections were found within 2 Kb of a TE region. This was significant with between 35.6% (*E. fawcettii*) and 63.4% (*E. ampelina*) (Table 18) of AT-rich regions signifying RIP-like affected sections of the genomes (Rouxel et al., 2011; Testa et al., 2016). It may be suggested that the remaining AT-rich regions have become RIP-affected to the point where the TE region is not identifiable. A small number of genes were found within the identified boundaries of the AT-rich areas, specifically between 3 (*E. australis* SO pathotype) and 73 (*E. ampelina*) (Table 16). Of these, one belonging to *E. fawcettii*, and one to *E. australis* FL pathotype were also predicted as effectors, an observation of importance as known effector genes of *L. maculans* have been located within AT-rich regions (Fudal et al., 2009; Gout et al., 2006). In comparison to the *Elsinoë* genomes, there was a larger range of diversity of the levels of AT-rich region coverage among the comparative fungal pathogen genomes. While no AT-rich regions were identifiable in some genomes such as *P. tritici-repentis, U. maydis, Pyricularia oryzae* and *S. sclerotiorum*, others such as *R. commune* and *L. maculans* had relatively high proportions of AT-rich regions (29.5% and 37%, respectively). This high degree of AT-rich region coverage in *R. commune* and *L. maculans* has been reported by other

studies (Frantzeskakis et al., 2019; Rouxel et al., 2011). No correlation was seen between the degree of AT-rich region coverage and lifestyle classification (necrotroph, hemibiotroph or biotroph), nor to variety of host range (narrow or broad). While this may be due to the small number of species analysed in the current study, Testa et al. (2016) analysed the genomes of 392 fungal species, reporting those which contained AT-rich regions were most frequently found within the Pezizomycotina subphylum and were found among a variety of lifestyle classifications including saprobes, symbionts, necrotrophs, hemibiotrophs and non-obligate biotrophs.

Small genome size has potentially contributed to the smaller predicted proteomes of *Elsinoë* (Figure 10) in comparison to that of the average Ascomycota proteome size (11129.45) (Mohanta & Bae, 2015). The predicted proteomes of *Elsinoë*, which ranged in size from 9,215 (FL pathotype) to 10,164 (*E. fawcettii* BRIP 54245a) proteins, contained 95.9% - 98.6% complete and single copy Ascomycota orthologs (Table 1). The proteomes were comparable in size to two of the three recently published *Elsinoë* genomes (DAR 70024 and Arg-1) (Shanmugam et al., 2020). This suggests a high level of completeness and further supports the observed smaller than average proteome sizes. Utilising the methods in this current study, the larger proteome of the recently published *E. fawcettii* SM16-1 genome was further analysed. Shanmugam et al., (2020), reported that the proteome consisted of 10,340 protein-coding genes. The current study identified a proteome of 10,519 protein-coding genes and, with additional analysis, indicated likely contamination of this genome assembly with bacterial sequences (Figure 15), potentially contributing to the larger genome size reported. As this study used a stringent method to determine potential contamination, only 20 contigs containing 516 gene models were removed, leaving the possibility that further contaminating sequences may remain in the assembly.

The number of proteins which obtained at least one Pfam hit in the proteome of each *Elsinoë* and comparative species indicated that species with higher

numbers of predicted proteins had lower proportions of proteins with at least one Pfam hit. Similarly, species with lower numbers of predicted proteins had higher proportions of proteins with at least one Pfam hit. While this is likely a product of the limitation provided by the number of models relevant to fungal organisms in the Pfam database, it can still provide useful information regarding the potential function of proteins or highlight proteins that may have novel functions when obtaining no hit. As fungal effectors are frequently species-specific and often have unique functions, a protein obtaining no hits to the Pfam database can be a useful shortlisting tool.

Phylogenetic analyses provided insight into closely related species of the *Elsinoë australis* pathotypes and *E. fawcettii*. Analysis of the *E. australis* pathotypes, indicated that the JBS pathogen may be more closely related to *E. punicae*, which infects the fruit and leaves of pomegranate (*Punica granatum*) (Fan et al., 2017), than to the original *E. australis* pathogen of sweet orange (Figure 3). Similarly, the FL pathotype appeared to be more closely related to *E. genipae-americanae*, a pathogen of *Genipa americana* (Fan et al., 2017), a species of tree found in North and South America, than to the SO pathotype. Further comparisons are needed between these closely aligned species, as it has been shown that the four sequence regions used in the analysis (ITS and partial TEF1-α, LSU and RPB2) have a high degree of similarity among species of *Elsinoë* (Fan et al., 2017). For example, *E. mattiroloanum* and *E. piri* have only 12-14 base changes over the same four regions (Figure 3). As *E. punicae*, *E. genipae-americanae* and the three pathotypes of *E. australis* differ in host range, it is likely that further differentiation between these species will appear with additional sequencing. There is, at present, no evidence that it is closely related to any other species of *Elsinoë* in Australia. However, as more *Elsinoë* have been collected from plants in Australia (DAF Biological Collections, 2019) than are currently sequenced, future investigations may assist in unravelling evolutionary relationships.

E. fawcettii were identical in the ITS and partial TEF1-α regions, with the exception of the Jingeul pathotype (Figure 4), despite analysis of multiple

pathotypes including the Lemon, Tyron's, FBHR, FNHR, SRGC (Hyun et al., 2009; Timmer et al., 1996). Further whole genome sequencing of international pathotypes including FBHR, FNHR, SRGC and Jingeul, is a suggested future step. This could accompany the comprehensive classification of pathotypes based on host pathogenicity testing of 61 isolates (Hyun et al., 2009) and may enable identification of genomic regions associated with host specificity. As only the Tyron's and Lemon pathotypes have been found in Australia, it is assumed all *E. fawcettii* sequenced in the current project belong to either of these two pathotypes. However, for validation, host pathogenicity testing of these isolates is a suggested future strategy for further clarification and investigation. *E. fawcettii* was most closely aligned with *E. citricola* (Figure 4), also a pathogen of *Citrus limon* identified in Brazil; however labelled as a different species based on changes within the partial TEF1-α and RPB2 regions. (Fan et al., 2017). Whether *E. citricola* is a true species of its own, or whether it could be described as an additional pathotype, perhaps will be determined by future investigation.

The high level of similarity among the species and pathotypes of *Elsinoë* exemplified the need to uncover alternate genetic regions for use in phylogenetic analysis and diagnostics. Three individual gene regions were identified for the differentiation of the *E. australis* pathotypes, including β-tubulin, Histone 4 and plasma membrane ATPase (*Pma1*) (Figures 6-8). Importantly, these regions were identical among all eight isolates of the JBS pathotype. This will need to be investigated among multiple isolates of both the FL and SO pathotypes in the future, to ensure identity is conserved within each pathotype. Interestingly, within the *Pma1* region, two base substitutions were observed between *E. fawcetii* SM16-1 and all other isolates (Figure 8), raising the possibility of a gene region that may be used to differentiate the FBHR pathotype from other *E. fawcetii* pathotypes.

Gene orthology analyses indicated the majority of *Elsinoë* genes were classified as core, being shared among all species of *Elsinoë* analysed and with 10 comparative fungal plant pathogens (Figures 13-16). These analyses illustrated *U. maydis* was an outlier among the group. As *U. maydis* was the

sole Basidiomycete and biotroph with the comparative group, its low rate of accessory genes and high rate of unique genes was understandable (Figure 16, Appendix C Table 5). Phylogenetic analysis also illustrated *U. maydis* was separated from all Ascomycota (Figure 5).

As expected, gene models of the JBS pathotypes were highly conserved within this pathotype, with very few unique genes and low proportions of accessory genes among most of the JBS pathotype genomes.
E. australis JBS pathotype Hillston-1 was an outlier, however, with on average 495 less core genes when analysed against the other seven JBS pathotypes (Figures 13, Appendix C Table 2). No biological significance was placed on this observation, as it was thought this may be a consequence of its smaller genome and proteome sizes, being potential artefacts of a highly fragmented assembly (Figures 9, 10 and 12). The low numbers of accessory and unique genes, among the JBS pathotype genomes, may be associated with their pathogenicity towards only one known host plant species. In comparison, similarly low rates of unique genes, but higher rates of accessory genes were seen within the *E. fawcettii* group (Figure 15, Appendix C Table 4). This higher number of accessory genes may be explained by the ability of the *E. fawcettii* pathotypes to each infect a different collection of citrus varieties to one another, with some overlap of pathogenicity range among the pathotypes. Furthermore, among the three pathotypes of *E. australis* (JBS, SO and FL) higher rates of both accessory and unique genes were observed (Figure 14, Appendix C Table 3). This may be associated with their pathogenicity towards three different host plants; jojoba, sweet orange and finger lime. However, whether the differences reported in host choice of these species and pathotypes can be attributed to whole unique genes, mutations in core genes or a combination of both is currently unknown, and therefore, it is suggested that both core and species-specific genes of *Elsinoë* be included in future experimental validation studies. The high number of unique genes seen within *E. fawcettii* SM16-1 (Figure 15) were determined to be produced by contaminating contigs within the assembly and were removed prior to subsequent analyses.

4.2 VIRULENCE GENE PREDICTION

Elsinochrome, the product of a SM gene cluster which contributes to necrotic activities, has been the focus of most research endeavouring to uncover the pathogenesis pathway of *Elsinoë* (Liao & Chung, 2008a, 2008b, Chung & Liao, 2008, Wang, et al., 2009). Analysis of *Elsinoë* genomes in this study has enabled the prediction of additional genes with potential involvement in the elsinochrome gene cluster, and additionally the prediction of extra SM gene clusters throughout the assemblies. Liao and Chung (2008a) showed the polyketide synthase *EfPKS1* of the elsinochrome gene cluster plays an important role in the formation of necrotic lesions on host plant leaves inoculated on day seven after emergence. However, they demonstrated that more unknown factors needed to be involved in full pathogenesis as necrotic lesions still formed on host plant leaves inoculated on day three after emergence despite the disruption of the *EfPKS1* (Liao & Chung, 2008a). These additional factors may involve further SM gene clusters, effector proteins or CWDE. Each of these potential pathogenicity-related factors will be discussed.

Six candidate SM gene clusters, with >80% similarity among the predicted core biosynthetic genes of the five different species and pathotypes of *Elsinoë* analysed, were identified (Table 6). Gene cluster #1 (Table 6, Figure 17) was determined to be the most likely candidate for the elsinochrome gene cluster, with high similarity to six elsinochrome genes previously identified (Chung & Liao, 2008a). A further nine genes, not included in previous studies, were also predicted (Figure 17), providing potential further candidates involved in elsinochrome production. These additional genes obtained Pfam hits for the THUMP domain, peptidase M3, Apolipoprotein O, Gar1/Naf1 RNA binding region and Endonuclease/Exonuclease/phosphatase family, which may indicate these nearby genes could function in RNA or lipid binding, peptide cleavage and intracellular signalling. It is suggested that these additional genes could be utilised in future experimental validation studies for further investigation into

the elsinochrome biosynthesis pathway. Furthermore, a second *T1PKS* cluster was predicted, being gene cluster #3 (Table 6), which showed similarity to the elsinochrome gene cluster of *Parastagonospora nodorum* and which appeared in all five species and pathotypes of *Elsinoë* analysed. This provides initial evidence that *Elsinoë* may hold a second elsinochrome gene cluster that may have been utilised by the pathogen during EfPKS1 disruption studies (Liao & Chung, 2008a), potentially explaining how the pathogen was able to cause necrotic lesions to young leaves despite the gene knockout. Recent work of Ebert et al. (2018) demonstrates the elsinochrome gene cluster of *E. fawcettii* is involved in the production of melanin, potentially highlighting the second predicted *T1PKS* cluster (gene cluster #3) as the primary genes involved in elsinochrome production. Lastly, there were four additional SM gene clusters, classed as terpenes and conserved among *Elsinoë* (Table 6) which are also prime candidates for future experimental validation studies. In particular, gene cluster #2 showed similarity to the clavaric acid gene cluster of *Hypholoma sublateritium*, a known antitumor SM which inhibits Ras farnesyl transferase (Godio et al., 2004; Jayasuriya et al., 1998). These additional predicted SM gene clusters are suggested candidates for future experimental validation studies.

Prediction of secretomes and effectomes among the *Elsinoë* genomes was a step forward in identifying a second subgroup of potential pathogenicity-related genes for future experimental investigation. The methods used to predict the secretomes, effectomes and prioritised candidate effectors (Figures 19 and 20) appear effective, with the majority of known fungal effectors (95.6%, 93.3% and 84.4%, respectively) being correctly predicted in each group (Table 8 and Table 9). As the number of candidate effectors for a genome can range in the hundreds, shortlisting is required to make experimental validation a viable future option. The predicted secretomes of the *Elsinoë* genomes were similar in size (11.5% - 13% of total predicted proteins) and within the range of the secretome size of the comparative species (11.3% - 18.5% of total predicted proteins) (Table 7). However, the proportion of candidate effectors from the predicted secretomes of *Elsinoë* were smaller in comparison to the other analysed species. The predicted

effectomes of *Elsinoë* ranged between 20.7% and 23.1% of total predicted secreted proteins, with only *Botrytis cinerea* being similar with 22.0%. Predicted effectors of all other comparative species ranged between 29.4% and 46.4% of predicted secreted proteins. This may be a possible effect of the smaller genome size of *Elsinoë*. The three known effectors which were not correctly predicted as candidate effectors included Vdlsc1, MoCDIP2 and Eff1-1. Vdlsc1, of *Verticillium dahlia*, does not contain an N-terminal signal peptide and is instead unconventionally secreted (Liu et al., 2014), however it was not predicted as non-classically secreted protein. MoCDIP2, of *Pyricularia oryzae*, was removed for containing a predicted GPI-anchor. Lastly, the effector of *Ustilago maydis*, Eff1-1, while predicted as a secreted protein, was not identified as a candidate effector by EffectorP. Sperschneider et al. (2018) have acknowledged Eff1-1 and MoCDIP2 as known false negatives of EffectorP.

Analysis of the 42 known effectors from the 10 comparative fungal species identified six features which were more prominent among the known effectors in comparison to all predicted proteins within the assemblies. Acknowledgment of these enabled the creation of a prioritisation pipeline to shortlist candidate effectors for experimental validation (Figure 20). These features, which were more commonly seen among known effectors, included >1500 bp to the nearest CDS of a gene both upstream and downstream, no involvement in predicted SM gene clusters, GC content of CDS being either $<Q_1$ or $>Q_3$ values and being labelled as unique or obtaining the same orthoMCL hit as a known effector. Known effectors, of species with assemblies with >2% TE coverage, were more frequently found within 20 Kb of a TE region. Lastly, known effectors, of species with assemblies with >25% AT-rich region coverage, were more frequently found nearby to AT-rich regions. Shortlisting candidate effectors of *Elsinoë*, based on these features, led to the prioritisation of between 94 and 120 for each species or pathotype of *Elsinoë*, a reduction of at least 52% of total candidate effectors (Table 10, Figure 21).

Analysis of the prioritised candidate effectors of *Elsinoë* indicated they were shorter in length (median protein length ranging between 150 AA and 181 AA) (Table 11), when compared to the predicted proteomes (median protein length varying between 404 AA and 419 AA) (Figure 11, Appendix C Table 1). Furthermore, a higher proportion (37.5% - 46.8%) of prioritised candidate effectors had high cysteine content of >3% (Table 11), compared to that of the predicted proteomes (3.8% - 4.8%). Both short protein length and high cysteine content have been observed among effectors in other studies (Duplessis et al., 2011; Stergiopoulos et al., 2013). Additionally, a small proportion of prioritised candidate effectors (0.8% - 8.3%) were predicted to be targeted to the host plant chloroplast (Table 11), suggesting *Elsinoë* may have the capacity to disrupt host function within the chloroplast. Function disruption within the host chloroplast has been reported previously, for example the SsITL effector of *Sclerotinia sclerotiorum* (Tang et al., 2020) and the Pst_12806 effector of *Puccinia striiformis* f. sp. *tritici* (Xu et al., 2019). Likewise, a low proportion (6.7% - 15.5%) obtained hits to PHI-base, indicating some similarity exists between known pathogenicity-related genes and several prioritised candidate effectors of *Elsinoë*. For example, prioritised candidate effectors with hits to the *Mogis2* gene of *Pyricularia oryzae* were identified in *E. ampelina* and all three *E. australis* pathotypes. *Mogis2* of *Pyricularia oryzae* is thought to be involved in appressorium-mediated penetration during plant infection (Zhou et al., 2011). Additionally, hits to the *Pep2* gene, a virulence factor of *Nectria haematococca* (Temporini & VanEtten, 2004), were observed in *E. ampelina* and *E. australis* JBS pathotype. Subsequent analysis indicated the prioritised candidate effectors were scattered throughout each assembly, being located on numerous contigs and not forming any apparent clusters. This is in contrast to the clustering of secreted effectors observed in the genome of *Ustilago maydis* (Brefort et al., 2014; Kämper et al., 2006). In comparison to the 10 comparative species, *Elsinoë* had a low number of prioritised candidate effectors, only *Botrytis cinerea* with 89, having a lower number (Table 10). This may be attributed to smaller genome and proteome size of *Elsinoë* or potentially to effector prediction software identifying greater numbers of

candidate effectors in species in which known effectors have previously been determined and subsequently utilised in the training of the software.

Limitations caused by the shortlisting of candidate effectors should be acknowledged. For example, this study prioritised candidate effectors which were either species-specific or obtained the same orthoMCL group ID as a known effector. If only prioritised candidate effectors are investigated in future experimental validation, this runs the risk of unintentionally eliminating novel effectors that may be shared by multiple pathogens. To avoid this, or similar situations, the cross-referencing of transcriptomic data against all candidate effectors, not only those which are prioritised, is a suggested future step.

Prediction of CWDE revealed the proteomes of *Elsinoë* may contain between 186 and 219 secreted CWDE (Table 12), including predicted pectin-degrading enzymes, hemicellulose-degrading enzymes and cellulose-degrading enzymes. Pectin-degrading enzymes and hemicellulose-degrading enzymes target pectin and xyloglucan in the plant cell wall (Kubicek, 2014), as these two components are abundant in the primary cell wall, this suggests *Elsinoë* may have a repertoire of virulence factors that could be effective against young plant tissues. Of the candidate CWDE of *Elsinoë*, 19 obtained hits to PHI-base and had orthologs in *E. fawcettii*, *E. ampelina* and all three *E. australis* pathotypes (Table 13), providing a more viable number of potential CWDE for future experimental validation. PHI-base hits included; the *PecA* polygalacturonase gene of *Aspergillus flavus*, which targets pectin for degradation and increases pathogen spread during infection (Shieh et al., 1997); the endo-1,4-beta-xylanases (glycosyl hydrolase families 10 and 11) of *Pyricularia oryzae*, a hemicellulose-degrading enzyme which has been observed to improve pathogenicity (Nguyen et al., 2011); the *Glu1* glucanase gene of *Pyrenophora tritici-repentis*, which contributes towards this pathogen's virulence to wheat (Fu et al., 2013); and, the avenacinase gene of *Gaeumannomyces graminis*, an enzyme which deglucosylates avenacin, a plant host inhibitor (Osbourn et al., 1991). The CWDE of *Elsinoë* have the

potential to cause degradation to components of a growing cell wall, perhaps providing a second virulence factor that may explain the necrosis observed by Liao and Chung (2008a) from inoculation of *E. fawcettii* on young leaves (3 day since emergence) despite the lack of the *EfPKS* elsinochrome-producing gene. The candidate CWDE identified and prioritised in this study provides an additional subgroup of potential virulence factors of *Elsinoë* for future investigation. The next suggested step is further shortlisting by cross-referencing with transcriptomic data when available, with additional gene knockout studies focusing on CWDE.

4.3 SEARCH FOR EVIDENCE OF GENOME COMPARTMENTALISATION

It is thought that virulence genes of the pathogen and their target proteins of the host plant co-evolve (Raffaele & Kamoun, 2012). Regions of the pathogen genome that have greater degrees of plasticity may provide locations suitable for the rapid evolution of virulence genes which code for proteins that interact with host proteins. Therefore, these regions of greater variability potentially highlight important virulence factors, aiding prediction and prioritisation. This study analysed the location of potential virulence factors including *Elsinoë*-specific genes, SM gene clusters, and genes coding for secreted proteins, candidate effectors and CWDE. The location of these predicted virulence factors was investigated in regard to the location of gene-sparse regions, TE regions, AT-rich regions, short repetitive sequences and fragmented regions across five species and pathotypes of *Elsinoë*. Within fungal pathogen genomes, gene-sparse regions and areas nearby to TE regions have been known to harbour virulence-related genes. For example, candidate effector genes of the fungal pathogen *Colletotrichum higginsianum* were frequently found in gene-sparse regions which contained a higher proportion of TE regions (Tsushima et al., 2019). Additionally, AT-rich regions, which are located throughout only some fungal genomes, differ to the majority of the genome by having a higher AT content (Testa et al., 2016). For example, the genome of *Leptosphaeria maculans* contains 36% AT-rich regions with an average of 33.9% GC content (Rouxel et al.,

2011). AT-rich regions have been found nearby to TE regions and are assumed to develop through Repeat-Induced Point mutation (RIP). Cytosine to thymine polymorphisms, occur during RIP to degrade TE and prevent potential destruction to necessary genes by TE movement. These polymorphisms can continue within genes in the close vicinity of TE regions (Cambareri et al., 1989; Gladyshev, 2017; Selker, 1990; Selker et al., 1987), increasing variation in these areas.

The overrepresentation of *Elsinoë*-specific genes in gene-sparse locations (Figure 23), nearby to TE regions (Figure 24), containing short repetitive sequences (Figure 25) and in fragmented regions (Figures 27 and 28) may suggest an ease of evolution of species-specific genes in regions devoid of densely packed genes and with local repetitive sequences. The location of these genes in regions of the genome with potential plasticity, while at the same time be identified as *Elsinoë*-specific, may indicate their potential utilisation as virulence factors. Investigation into the function of *Elsinoë*-specific genes is a suggested future step towards a better understanding of the pathogenesis pathway of *Elsinoë*.

Genes coding for predicted secreted proteins were found overrepresented in regions nearby to AT-rich regions (Figure 26) and at the edge of scaffolds (Figure 28). Genes coding for candidate effectors were also overrepresented in both of these areas, as well as in gene-sparse regions (Figure 23) and in locations nearby to TE regions (Figure 24). A distinction should be made between these two groups, being genes coding simply for predicted secreted proteins, pathogenicity-related or not, and those coding for candidate effectors, which are more likely to be associated with pathogenicity. The results suggest that potentially the more likely a gene is to be involved in pathogenesis, the more frequently it may be found in a region of the assembly that is associated with plasticity. While some fungal effector genes have been shown to contain repetitive sequences (Yu et al., 2017), this was not the case within the candidate effectors of *Elsinoë*, which showed markedly reduced proportions of short repeat-containing genes among those coding for candidate effectors (Figure 25).

Aside from genes which code for predicted secreted proteins and candidate effectors, genes coding for predicted CWDE were also found to be over-represented in locations in the close vicinity of AT-rich regions (Figure 26). This suggests these candidate virulence factors may reside nearby to AT-rich regions where they have the potential to continually evolve (Rouxel et al., 2011; Testa et al., 2016). This was in contrast to SM gene clusters, which overall did not appear to be located in any areas associated with potential plasticity identified in this study. This finding aligns with the knowledge that the elsinochrome SM gene cluster is well conserved among *Elsinoë* species and therefore may require a higher degree of genomic stability to maintain its function.

The *E. ampelina* genome had the greatest proportion of AT-rich regions, with 17.9% of the assembly having a mean GC content of 30.5% (Table 16). As discussed previously, the *E. ampelina* assembly was also made up of the lowest number of scaffolds of the *Elsinoë* spp. (Figure 12, Appendix C Table 1), due to the use of long read sequencing. It is therefore questionable whether the current assemblies for *E. fawcettii* and the three *E. australis* pathotypes provide a true indication of the proportion of AT-rich region content. It is possible that some AT-rich regions were lost during the assembly process due to a lack of diversity within the short-read sequences. While this would unlikely affect AT-rich regions which contain predictable gene models, it may simply lead to a lower overall reported proportion of AT-rich region content.

Finally, SNP identified during the analysis were cross-referenced against the location of predicted virulence factors for *E. fawcettii* and *E. australis* JBS pathotype, yet these regions failed to show any propensity for the candidate virulence factors predicted in this study (Table 18). While some studies have utilised diversifying selection as further evidence and support in identifying virulence factors (Guyon et al., 2014; Poppe et al., 2015; Sperschneider et al., 2014), no evidence of this process was identified in the *Elsinoë* assemblies.

Much research surrounding *Elsinoë* has focused solely on the elsinochrome SM gene cluster, it is therefore easy to neglect the idea that these pathogens potentially utilise other virulence factors by these pathogens. However, observing the patterns which form when cross-referencing the candidate virulence genes with markers for genome plasticity, suggests these genes may have the means to continually evolve, highlighting that they potentially provide a significant purpose for the organism.

CHAPTER 5: CONCLUSION

Elsinoë fawcettii and *E. australis* cause loss to agricultural crops around the world. Genomic analyses can predict virulence related factors, which, with experimental validation, can identify individual genes which contribute to pathogenesis. Genome mining, prediction and prioritisation of candidate virulence factors is an important step in this process, as a pathogen's genome can have a multitude of predicted gene models which far exceeds the resources and time available for experimental validation. This study has generated nine genome assemblies for multiple isolates of *Elsinoë*, two of which were annotated. Prediction and analysis of SM gene clusters indicated that in addition to the elsinochrome gene cluster, a further five clusters are present which are conserved among the species of *Elsinoë* analysed. Prediction of CWDE and further comparison against the PHI-base led to the prioritisation of between 21 and 25 candidate CWDE for the species and pathotypes of *Elsinoë*. Prediction of the effectomes of the *Elsinoë* spp., with subsequent analysis of known effectors and their respective genome assemblies led to the generation of points-based pathway for candidate effector prioritisation. Using this method resulted in the shortlisting of between 94 and 120 candidate effectors for the *Elsinoë* spp.. Additionally, this prioritisation pathway can be utilised by other studies wishing to shortlist candidate effectors prior to experimental validation. Thus, this study has generated comprehensive resources for candidate virulence-related genes for two species of *Elsinoë* which can be used to guide future experimental research. This study also searched for genomic regions with potential plasticity within the *Elsinoë* assemblies. Some candidate virulence genes, including CWDE and effectors, were more likely to be found nearby to TE, AT-rich and fragmented regions, providing a potential mechanism for variability of some pathogenicity-related genes. Overall, this study added to knowledge surrounding potential virulence factors of *Elsinoë*, beyond the currently known elsinochrome gene cluster. Importantly, the candidate virulence factors presented may provide an explanation for the observed host-specificity seen among the different pathotypes. The resources provided

here will aid the viability of future experimental validation of virulence genes for *Elsinoë*.

REFERENCES

Afgan, E., Baker, D., Batut, B., Van Den Beek, M., Bouvier, D., Čech, M., . . . Grüning, B. A. (2018). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Research, 46*(W1), W537-W544.

Afgan, E., Sloggett, C., Goonasekera, N., Makunin, I., Benson, D., Crowe, M., . . . Lonie, A. (2015). Genomics Virtual Laboratory: A Practical Bioinformatics Workbench for the Cloud.(Research Article). *PLoS ONE, 10*(10), e0140829. doi:10.1371/journal.pone.0140829

Ahmed, Y., Hussein, A., Hubert, J., Fourrier-Jeandel, C., Aguayo, J., & Ioos, R. (2020). New multiplex conventional PCR and quadruplex real-time PCR assays for one-tube detection of *Phyllosticta citricarpa*, *Elsinoë fawcettii*, *Elsinoë australis*, and *Pseudocercospora angolensis* in Citrus: development and validation. *Applied Microbiology and Biotechnology, 104*(21), 9363-9385.

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Research, 25*(17), 3389-3402. doi:10.1093/nar/25.17.3389

Amselem, J., Cuomo, C. A., van Kan, J. A. L., Viaud, M., Benito, E. P., Couloux, A., . . . Dickman, M. (2011). Genomic analysis of the necrotrophic fungal pathogens *Sclerotinia sclerotiorum* and *Botrytis cinerea*. *PLoS Genetics, 7*(8), e1002230. doi:10.1371/journal.pgen.1002230

Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data.

Armenteros, J. J. A., Salvatore, M., Emanuelsson, O., Winther, O., Von Heijne, G., Elofsson, A., & Nielsen, H. (2019). Detecting sequence signals in targeting peptides using deep learning. *Life science alliance, 2*(5).

Ash, G., Stodart, B., & Hyun, J.W. (2012). Black scab of jojoba (Simmondsia chinensis) in Australia caused by a putative new pathotype of *Elsinoë australis*. *Plant disease, 96*(5), 629-634.

Ash, G. J., Lang, J. M., Triplett, L. R., Stodart, B. J., Verdier, V., Cruz, C. V., . . . Leach, J. E. (2014). Development of a genomics-based LAMP (loop-mediated isothermal amplification) assay for detection of *Pseudomonas fuscovaginae* from rice. *Plant disease, 98*(7), 909-915.

Aurrecoechea, C., Barreto, A., Basenko, E. Y., Brestelli, J., Brunk, B. P., Cade, S., . . . Fischer, S. (2017). EuPathDB: the eukaryotic pathogen genomics database resource. *Nucleic Acids Research, 45*(D1), D581-D591.

Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., . . . Noble, W. S. (2009). MEME Suite: Tools for motif discovery and searching. *Nucleic Acids Research, 37*(2), W202-W208. doi:10.1093/nar/gkp335

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., . . . Pevzner, P. A. (2012). SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *Journal of Computational Biology, 19*(5), 455-477. doi:10.1089/cmb.2012.0021

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

Batuman, O., & Ritenour, M. (2020). Diseases caused by fungi and oomycetes. In M. Talon, M. Caruso, F. G. Gmitter (Eds.), The Genus Citrus (pp. 349-369). https://doi.org/10.1016/B978-0-12-812163-4.00017-6.

Beier, S., Thiel, T., Münch, T., Scholz, U., & Mascher, M. (2017). MISA-web: a web server for microsatellite prediction. *Bioinformatics (Oxford, England), 33*(16), 2583-2585. doi:10.1093/bioinformatics/btx198

Bitancourt, A. A., & Jenkins, A. E. (1936). *Elsinoe fawcettii*, the perfect stage of the Citrus scab fungus. *Phytopathology, 26*(4), 393-395.

Bitancourt, A. A., & Jenkins, A. E. (1937). Sweet orange fruit scab caused by australis. *Journal of Agricultural Research, 54*, 0001-0018.

Blin, K., Shaw, S., Steinke, K., Villebro, R., Ziemert, N., Lee, S. Y., . . . Weber, T. (2019). antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. *Nucleic Acids Research, 47*(W1), W81-W87.

Blin, K., Wolf, T., Chevrette, M. G., Lu, X., Schwalen, C. J., Kautsar, S. A., . . . Medema, M. H. (2017). AntiSMASH 4.0 - improvements in chemistry prediction and gene cluster boundary identification. *Nucleic Acids Research, 45*(1), W36-W41. doi:10.1093/nar/gkx319

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics, 30*(15), 2114-2120. doi:10.1093/bioinformatics/btu170

Bolton, M. D., Thomma, B. P. H. J., & Nelson, B. D. (2006). *Sclerotinia sclerotiorum* (Lib.) de Bary: biology and molecular traits of a cosmopolitan pathogen (Vol. 7, pp. 1-16). Oxford, UK: Blackwell Science Ltd.

Bravo Ruiz, G., Di Pietro, A., & Roncero, M. I. G. (2016). Combined action of the major secreted exo- and endopolygalacturonases is required for full virulence of *Fusarium oxysporum*. *Molecular plant pathology, 17*(3), 339-353. doi:10.1111/mpp.12283

Brefort, T., Tanaka, S., Neidig, N., Doehlemann, G., Vincon, V., & Kahmann, R. (2014). Characterization of the largest effector gene cluster of *Ustilago maydis*. PLoS Pathog, 10(7), e1003866.

Brito, N., Espino, J. J., & González, C. (2006). The endo-β-1,4-xylanase Xyn11A is required for virulence in *Botrytis cinerea*. *Molecular plant-microbe interactions, 19*(1), 25-32. doi:10.1094/MPMI-19-0025

Buchfink, B., Xie, C., & Huson, D., H. (2014). Fast and sensitive protein alignment using DIAMOND. *Nature Methods, 12*(1). doi:10.1038/nmeth.3176

Buiate, E. A., Xavier, K. V., Moore, N., Torres, M. F., Farman, M. L., Schardl, C. L., & Vaillancourt, L. J. (2017). A comparative genomic analysis of putative pathogenicity genes in the host-specific sibling species *Colletotrichum graminicola* and *Colletotrichum sublineola*. *BMC genomics, 18*(1), 67.

Busk, P. K., Pilgaard, B., Lezyk, M. J., Meyer, A. S., & Lange, L. (2017). Homology to peptide pattern for annotation of carbohydrate-active enzymes and prediction of function.(Report). *BMC Bioinformatics, 18*(1). doi:10.1186/s12859-017-1625-9

CABI, 2017. Elsinoë australis (citrus scab). In: Invasive Species Compendium. Wallingford, UK: CAB International. www.cabi.org/isc.

Cambareri, E. B., Jensen, B. C., Schabtach, E., & Selker, E. U. (1989). Repeat-induced G-C to A-T mutations in *Neurospora*. (glycine-

Bravo Ruiz, G., Di Pietro, A., & Roncero, M. I. G. (2016). Combined action of the major secreted exo- and endopolygalacturonases is required for full virulence of *Fusarium oxysporum*. *Molecular plant pathology, 17*(3), 339-353. doi:10.1111/mpp.12283

Brefort, T., Tanaka, S., Neidig, N., Doehlemann, G., Vincon, V., & Kahmann, R. (2014). Characterization of the largest effector gene cluster of *Ustilago maydis*. PLoS Pathog, 10(7), e1003866.

Brito, N., Espino, J. J., & González, C. (2006). The endo-β-1,4-xylanase Xyn11A is required for virulence in *Botrytis cinerea*. *Molecular plant-microbe interactions, 19*(1), 25-32. doi:10.1094/MPMI-19-0025

Buchfink, B., Xie, C., & Huson, D., H. (2014). Fast and sensitive protein alignment using DIAMOND. *Nature Methods, 12*(1). doi:10.1038/nmeth.3176

Buiate, E. A., Xavier, K. V., Moore, N., Torres, M. F., Farman, M. L., Schardl, C. L., & Vaillancourt, L. J. (2017). A comparative genomic analysis of putative pathogenicity genes in the host-specific sibling species *Colletotrichum graminicola* and *Colletotrichum sublineola*. *BMC genomics, 18*(1), 67.

Busk, P. K., Pilgaard, B., Lezyk, M. J., Meyer, A. S., & Lange, L. (2017). Homology to peptide pattern for annotation of carbohydrate-active enzymes and prediction of function.(Report). *BMC Bioinformatics, 18*(1). doi:10.1186/s12859-017-1625-9

CABI, 2017. Elsinoë australis (citrus scab). In: Invasive Species Compendium. Wallingford, UK: CAB International. www.cabi.org/isc.

Cambareri, E. B., Jensen, B. C., Schabtach, E., & Selker, E. U. (1989). Repeat-induced G-C to A-T mutations in *Neurospora*. (glycine-

cysteine to alanine-threonine). *Science, 244*(4912), 1571. doi:10.1126/science.2544994

Cantarel, B. I., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V., & Henrissat, B. (2009). The Carbohydrate-Active EnZymes database (CAZy): An expert resource for glycogenomics. *Nucleic Acids Research, 37*(1), D233-D238. doi:10.1093/nar/gkn663

Carisse, O., & Lefebvre, A. (2011). Evaluation of northern grape hybrid cultivars for their susceptibility to anthracnose caused by *Elsinoe ampelina*. *Plant Health Progress, 12*(1), 9.

Chen, S., Songkumarn, P., Venu, R. C., Gowda, M., Bellizzi, M., Hu, J., . . . Wang, G. L. (2013). Identification and characterization of in planta-expressed secreted effector proteins from magnaporthe oryzae that induce cell death in rice. *Molecular plant-microbe interactions, 26*(2), 191-202. doi:10.1094/MPMI-05-12-0117-R

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

Chung, K. R., & Liao, H. L. (2008). Determination of a transcriptional regulator-like gene involved in biosynthesis of elsinochrome phytotoxin by the citrus scab fungus, *Elsinoë fawcettii*. *Microbiology, 154*(11), 3556-3566.

Chung, K. R. (2011). *Elsinoë fawcettii* and *Elsinoë australis*: the fungal pathogens causing citrus scab. *Molecular plant pathology, 12*(2), 123-135.

Citrus Australis (2019). Australian citrus exports surpass $500m mark. Retrieved on 31st December 2020 from https://citrusaustralia.com.au/news/latest-news/australian-citrus-exports-pass-500m-mark

Citrus Australis (2020). Fresh Australian citrus at risk without Government intervention. Retrieved on 31st December 2020 from https://citrusaustralia.com.au/media-release/fresh-australian-citrus-at-risk-without-government-intervention

Coleman, J. J., Rounsley, S. D., Rodriguez-Carres, M., Kuo, A., Wasmann, C. C., Grimwood, J., . . . Zhou, S. (2009). The genome of *Nectria haematococca*: contribution of supernumerary chromosomes to gene expansion. *PLoS Genet, 5*(8), e1000618.

Conesa, A., Götz, S., García-Gómez, J. M., Terol, J., Talón, M., & Robles, M. (2005). Blast2GO: A universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics, 21*(18), 3674-3676. doi:10.1093/bioinformatics/bti610

Cosgrove, D. J. (2005). Growth of the plant cell wall (Vol. 6, pp. 850): Nature Publishing Group.

Croll, D., & McDonald, B. A. (2012). The accessory genome as a cradle for adaptive evolution in pathogens. *PLoS Pathog, 8*(4), e1002608.

Dean, R. A., Talbot, N. J., Ebbole, D. J., Farman, M. L., Mitchell, T. K., Orbach, M. J., . . . Pan, H. (2005). The genome sequence of the rice blast fungus *Magnaporthe grisea*. *Nature, 434*(7036), 980.

Delort, E., & Yuan, Y. M. (2018). Finger lime/The Australian Caviar—Citrus australasica. *Exotic fruits* (pp. 203-210). Academic Press.

Department of Agriculture and Water Resources (DAWR). (2013). Changes to import requirements for Californian citrus (Notice to Industry 103-2013) Retrieved 10 August 2017 from http://www.agriculture.gov.au/import/industry-advice/13/103-2013

Djamei, A., Schipper, K., Rabe, F., Ghosh, A., Vincon, V., Kahnt, J., . . . Kahmann, R. (2011). Metabolic priming by a secreted fungal effector. *Nature, 478*(7369), 395. doi:10.1038/nature10454

dos Santos, R. F., Spósito, M. B., Ayres, M. R., & Sosnowski, M. R. (2018). Phylogeny, morphology and pathogenicity of *Elsinoë ampelina*, the causal agent of grapevine anthracnose in Brazil and Australia. *Journal of Phytopathology, 166*(3), 187-198.

Duplessis, S., Cuomo, C. A., Lin, Y.C., Aerts, A., Tisserant, E., Veneault-Fourrey, C., . . . Cantarel, B. L. (2011). Obligate biotrophy features unraveled by the genomic analysis of rust fungi. *Proceedings of the National Academy of Sciences, 108*(22), 9166-9171.

Ebert, M. K., Spanner, R. E., de Jonge, R., Smith, D. J., Holthusen, J., Secor, G. A., ... & Bolton, M. D. (2019). Gene cluster conservation identifies melanin and perylenequinone biosynthesis pathways in multiple plant pathogenic fungi. Environmental microbiology, 21(3), 913-927.

Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research, 32*(5), 1792-1797. doi:10.1093/nar/gkh340

Ekblom, R., & Wolf, J. B. (2014). A field guide to whole-genome sequencing, assembly and annotation. *Evolutionary applications, 7*(9), 1026-1042.

Fan, X., Barreto, R., Groenewald, J., Bezerra, J., Pereira, O., Cheewangkoon, R., . . . Crous, P. W. (2017). Phylogeny and

taxonomy of the scab and spot anthracnose fungus *Elsinoë*
(Myriangiales, Dothideomycetes). *Studies in mycology, 87*, 1-41.

Finn, R. D., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Mistry, J., Mitchell, A.
L., . . . Bateman, A. (2016). The Pfam protein families database:
Towards a more sustainable future. *Nucleic Acids Research, 44*(1),
D279-D285. doi:10.1093/nar/gkv1344

Fischer, S., Brunk, B. P., Chen, F., Gao, X., Harb, O. S., Iodice, J. B., . . .
Stoeckert Jr, C. J. (2011). Using OrthoMCL to assign proteins to
OrthoMCL-DB groups or to cluster proteomes into new ortholog
groups. *Current protocols in bioinformatics, 35*(1), 6.12. 11-16.12. 19.

Frantzeskakis, L., Kracher, B., Kusch, S., Yoshikawa-Maekawa, M., Bauer,
S., Pedersen, C., … & Panstruga, R. (2018). Signatures of host
specialization and a recent transposable element burst in the dynamic
one-speed genome of the fungal barley powdery mildew pathogen.
BMC genomics, 19(1), 1-23.

Frantzeskakis, L., Németh, M. Z., Barsoum, M., Kusch, S., Kiss, L.,
Takamatsu, S., & Panstruga, R. (2019). The *Parauncinula polyspora*
draft genome provides insights into patterns of gene erosion and
genome expansion in powdery mildew fungi. mBio, 10(5).

Friesen, T. L., Faris, J. D., Solomon, P. S., & Oliver, R. P. (2008). Host-
specific toxins: effectors of necrotrophic pathogenicity (Vol. 10, pp.
1421-1428). Oxford, UK: Blackwell Publishing Ltd.

Fu, H., Feng, J., Aboukhaddour, R., Cao, T., Hwang, S. F., & Strelkov, S. E.
(2013). An exo-1,3-[beta]-glucanase GLU1 contributes to the virulence
of the wheat tan spot pathogen *Pyrenophora tritici-repentis*. *Fungal
Biology, 117*(10), 673.

Fudal, I., Ross, S., Brun, H., Besnard, A. L., Ermel, M., Kuhn, M. L., . . . Rouxel, T. (2009). Repeat-Induced Point Mutation (RIP) as an alternative mechanism of evolution toward virulence in *Leptosphaeria maculans*. *Molecular plant-microbe interactions, 22*(8), 932-941. doi:10.1094/MPMI-22-8-0932

Garrison, E., & Marth, G. (2012). Haplotype-based variant detection from short-read sequencing. *arXiv preprint arXiv:1207.3907*.

Garrison, E. (2015). vcflib. GitHub repository. GitHub. Retrieved from https://github.com/ekg/vcflib

Gladman, S. (2012). VelvetOptimiser. *Victorian Bioinformatics Consortium, Clayton, Australia: http://bioinformatics.net.ausoftware.velvetoptimiser.shtml*.

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

Godio, R., Fouces, R., Gudina, E., & Martin, J. (2004). *Agrobacterium tumefaciens*-mediated transformation of the antitumor clavaric acid-producing basidiomycete *Hypholoma sublateritium*. *Current Genetics, 46*(5), 287-294.

Gout, L., Fudal, I., Kuhn, M. L., Blaise, F., Eckert, M., Cattolico, L., . . . Rouxel, T. (2006). Lost in the middle of nowhere: the AvrLm1 avirulence gene of the Dothideomycete *Leptosphaeria maculans*. *Molecular Microbiology, 60*(1), 67-80. doi:10.1111/j.1365-2958.2006.05076.x

Grandaubert, J., Lowe, R. G., Soyer, J. L., Schoch, C. L., Van de Wouw, A. P., Fudal, I., . . . Ollivier, B. (2014). Transposable element-assisted evolution and adaptation to host plant within the *Leptosphaeria*

maculans-Leptosphaeria biglobosa species complex of fungal pathogens. *BMC genomics, 15*(1), 1-27.

Grant, C. E., Bailey, T. L., & Noble, W. S. (2011). FIMO: Scanning for occurrences of a given motif. *Bioinformatics, 27*(7), 1017-1018. doi:10.1093/bioinformatics/btr064

Grigoriev, I. V., Nikitin, R., Haridas, S., Kuo, A., Ohm, R., Otillar, R., . . . Korzeniewski, F. (2014). MycoCosm portal: gearing up for 1000 fungal genomes. *Nucleic Acids Research, 42*(D1), D699-D704.

Guyon, K., Balagué, C., Roby, D., & Raffaele, S. (2014). Secretome analysis reveals effector candidates associated with broad host range necrotrophy in the fungal plant pathogen *Sclerotinia sclerotiorum*. *BMC genomics, 15*(1), 336.

Hane, J. K., Lowe, R. G., Solomon, P. S., Tan, K. C., Schoch, C. L., Spatafora, J. W., . . . Galagan, J. E. (2007). Dothideomycete–plant interactions illuminated by genome sequencing and EST analysis of the wheat pathogen *Stagonospora nodorum*. *The Plant Cell, 19*(11), 3347-3368.

Hartmann, F. E., Sánchez-Vallet, A., McDonald, B. A., & Croll, D. (2017). A fungal wheat pathogen evolved host specialization by extensive chromosomal rearrangements. The ISME journal, 11(5), 1189-1204.

Hogenhout, S. A., Van Der Hoorn, R. A. L., Terauchi, R., & Kamoun, S. (2009). Emerging concepts in effector biology of plant-associated organisms. *Molecular plant-microbe interactions, 22*(2), 115-122. doi:10.1094/MPMI-22-2-0115

Hyun, J., Peres, N., Yi, S. Y., Timmer, L., Kim, K., Kwon, H.-M., & Lim, H. C. (2007). Development of PCR assays for the identification of species

and pathotypes of *Elsinoë* causing scab on citrus. *Plant disease, 91*(7), 865-870.

Hyun, J., Timmer, L., Lee, S., Yun, S., Ko, S., & Kim, K. (2001). Pathological characterization and molecular analysis of *Elsinoe* isolates causing scab diseases of citrus in Jeju Island in Korea. *Plant disease, 85*(9), 1013-1017.

Hyun, J., Yi, S., MacKenzie, S., Timmer, L., Kim, K., Kang, S., . . . Lim, H. (2009). Pathotypes and genetic relationship of worldwide collections of *Elsinoë* spp. causing scab diseases of citrus. *Phytopathology, 99*(6), 721-728.

Jayasuriya, H., Silverman, K. C., Zink, D. L., Jenkins, R. G., Sanchez, M., Pelaez, F., . . . Singh, S. B. (1998). Clavaric Acid: A Triterpenoid Inhibitor of Farnesyl-Protein Transferase from *Clavariadelphus truncatus. Journal of natural products, 61*(12), 1568-1570.

Jeffress, S., Arun-Chinnappa, K., Stodart, B., Vaghefi, N., Tan, Y. P., & Ash, G. (2020). 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, 15*(5), e0227396.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., & Charpentier, E. (2012). A programmable dual-RNA–guided DNA endonuclease in adaptive bacterial immunity. *Science, 337*(6096), 816-821.

Johnson, L. S., Eddy, S. R., & Portugaly, E. (2010). Hidden Markov model speed heuristic and iterative HMM search procedure. *BMC Bioinformatics, 11*(1), <xocs:firstpage xmlns:xocs=""/>. doi:10.1186/1471-2105-11-431

Käll, L., Krogh, A., & Sonnhammer, E. L. L. (2004). A Combined Transmembrane Topology and Signal Peptide Prediction Method. *Journal of Molecular Biology, 338*(5), 1027-1036. doi:10.1016/j.jmb.2004.03.016

Kamoun, S. (2006). A catalogue of the effector secretome of plant pathogenic oomycetes. *Annu. Rev. Phytopathol., 44*, 41-60.

Kämper, J., Kahmann, R., Bölker, M., Ma, L. J., Brefort, T., Saville, B. J., . . . Müller, O. (2006). Insights from the genome of the biotrophic fungal plant pathogen *Ustilago maydis*. *Nature, 444*(7115), 97.

Kautsar, S. A., Blin, K., Shaw, S., Navarro-Muñoz, J. C., Terlouw, B. R., van der Hooft, J. J., . . . Pascal Andreu, V. (2020). MIBiG 2.0: a repository for biosynthetic gene clusters of known function. *Nucleic Acids Research, 48*(D1), D454-D458.

Kim, M., Lee, J., Heo, L., & Han, S. W. (2020). Putative bifunctional chorismate mutase/prephenate dehydratase contributes to the virulence of Acidovorax citrulli. Frontiers in plant science, 11, 1482.

King, B. C., Waxman, K. D., Nenni, N. V., Walker, L. P., Bergstrom, G. C., & Gibson, D. M. (2011). Arsenal of plant cell wall degrading enzymes reflects host preference among plant pathogenic fungi.(Research)(Report). *Biotechnology for Biofuels, 4*, 4.

Kirsten, S., Navarro-Quezada, A., Penselin, D., Wenzel, C., Matern, A., Leitner, A., . . . Knogge, W. (2012). Necrosis-inducing proteins of rhynchosporium commune, effectors in quantitative disease resistance. *Molecular plant-microbe interactions, 25*(10), 1314-1325. doi:10.1094/MPMI-03-12-0065-R

Klosterman, S., Subbarao, K., Kang, S., Veronese, P., Gold, S., Thomma, B., . . . Park, J. (2010). Comparative genomics of the plant vascular wilt

pathogens, *Verticillium dahliae* and *Verticillium albo-atrum*. *Phytopathology, 100*(6), S64-S64.

Koch, A., & Kogel, K. H. (2014). New wind in the sails: improving the agronomic value of crop plants through RNA i-mediated gene silencing. *Plant biotechnology journal, 12*(7), 821-831.

Kokoa, P. (2001). *Review of sweet potato diseases in PNG.* Paper presented at the Food Security for Papua New Guinea. Proceedings of the Papua New Guinea Food and Nutrition 2000 Conference. ACIAR Proceedings.

Kriventseva, E. V., Kuznetsov, D., Tegenfeldt, F., Manni, M., Dias, R., Simão, F. A., & Zdobnov, E. M. (2019). OrthoDB v10: sampling the diversity of animal, plant, fungal, protist, bacterial and viral genomes for evolutionary and functional annotations of orthologs. *Nucleic Acids Research, 47*(D1), D807-D811.

Krogh, A., Larsson, B., Von Heijne, G., & Sonnhammer, E. L. L. (2001). Predicting transmembrane protein topology with a hidden markov model: application to complete genomes. *Journal of Molecular Biology, 305*(3), 567-580. doi:10.1006/jmbi.2000.4315

Kubicek, C. P., Starr, T. L., & Glass, N. L. (2014). Plant Cell WallDegrading Enzymes and Their Secretion in Plant-Pathogenic Fungi. *Annual Review of Phytopathology, 52*(1), 427-451. doi:10.1146/annurev-phyto-102313-045831

Kumar, S., Stecher, G., & Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Molecular biology and evolution, 33*(7), 1870-1874. doi:10.1093/molbev/msw054

Kunta, M., Rascoe, J., Sa, P. B. d., Timmer, L. W., Palm, M. E., Graça, J. V. d., . . . Satpute, A. (2013). Sweet orange scab with a new scab disease" syndrome" of citrus in the USA associated with *Elsinoë australis*. *Tropical Plant Pathology, 38*(3), 203-212.

Lagesen, K., Hallin, P., Rødland, E. A., Stærfeldt, H. H., Rognes, T., & Ussery, D. W. (2007). RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Research, 35*(9), 3100-3108.

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods, 9*(4), 357.

Lee, T., Peace, C., Jung, S., Zheng, P., Main, D., & Cho, I. (2011). GenSAS - An online integrated genome sequence annotation pipeline (Vol. 4, pp. 1967-1973).

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., . . . Durbin, R. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics, 25*(16), 2078-2079.

Li, Z., Fan, Y., Chang, P., Gao, L., & Wang, X. (2020). Genome sequence resource for *Elsinoë ampelina*, the causal organism of grapevine anthracnose. *Molecular plant-microbe interactions, 33*(4), 576-579.

Liao, H. L., & Chung, K. R. (2008a). Genetic dissection defines the roles of elsinochrome phytotoxin for fungal pathogenesis and conidiation of the citrus pathogen *Elsinoë fawcettii. Molecular plant-microbe interactions, 21*(4), 469-479.

Liao, H. L., & Chung, K. R. (2008b). Cellular toxicity of elsinochrome phytotoxins produced by the pathogenic fungus, *Elsinoë fawcettii* causing citrus scab. *New Phytologist, 177*(1), 239-250.

Liu, T., Song, T., Zhang, X., Yuan, H., Su, L., Li, W., . . . Dou, D. (2014). Unconventionally secreted effectors of two filamentous pathogens target plant salicylate biosynthesis. *Nature Communications, 5*(1). doi:10.1038/ncomms5686

Lo Presti, L., Lanver, D., Schweizer, G., Tanaka, S., Liang, L., Tollot, M., . . . Kahmann, R. (2015). Fungal Effectors and Plant Susceptibility. *Annual Review of Plant Biology, 66*(1), 513-545. doi:10.1146/annurev-arplant-043014-114623

Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M., & Henrissat, B. (2014). The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Research, 42*(1), D490-D495. doi:10.1093/nar/gkt1178

Lomsadze, A., Ter-Hovhannisyan, V., Chernoff, Y. O., & Borodovsky, M. (2005). Gene identification in novel eukaryotic genomes by self-training algorithm. *Nucleic Acids Research, 33*(20), 6494-6506. doi:10.1093/nar/gki937

López-Pérez, M., Ballester, A. R., & González-Candelas, L. (2015). Identification and functional analysis of *Penicillium digitatum* genes putatively involved in virulence towards citrus fruit. *Molecular plant pathology, 16*(3), 262-275. doi:10.1111/mpp.12179

Lowe, T. M., & Eddy, S. R. (1997). tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Research, 25*(5), 955-964.

Lyu, X., Shen, C., Fu, Y., Xie, J., Jiang, D., Li, G., & Cheng, J. (2016). A Small Secreted Virulence-Related Protein Is Essential for the Necrotrophic Interactions of *Sclerotinia sclerotiorum* with Its Host Plants.(Research Article). *PLoS Pathogens, 12*(2), e1005435. doi:10.1371/journal.ppat.1005435

Ma, L. J., Van Der Does, H. C., Borkovich, K. A., Coleman, J. J., Daboussi, M.-J., Di Pietro, A., . . . Henrissat, B. (2010). Comparative genomics reveals mobile pathogenicity chromosomes in Fusarium. *Nature, 464*(7287), 367-373.

Malnoy, M., Viola, R., Jung, M.-H., Koo, O. J., Kim, S., Kim, J. S., . . . Nagamangala Kanchiswamy, C. (2016). DNA-free genetically edited grapevine and apple protoplast using CRISPR/Cas9 ribonucleoproteins. *Frontiers in Plant Science, 7*, 1904.

Manning, V. A., Pandelova, I., Dhillon, B., Wilhelm, L. J., Goodwin, S. B., Berlin, A. M., . . . Ciuffetti, L. M. (2013). 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.), 3*(1), 41-63. doi:10.1534/g3.112.004044

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

McTaggart, A., (2007). Sweet Orange Fruit Scab (Elsinoë australis). Updated on 12/21/2007. Retrieved from http://www.padil.gov.au on 20th July 2017.

Menardo, F., Praz, C. R., Wyder, S., Ben-David, R., Bourras, S., Matsumae, H., ... & Keller, B. (2016). Hybridization of powdery mildew strains gives rise to pathogens on novel agricultural crop species. Nature Genetics, 48(2), 201-205.

Mentlak, T. A., Kombrink, A., Shinya, T., Ryder, L. S., Otomo, I., Saitoh, H., . . . Thomma, B. P. (2012). Effector-mediated suppression of chitin-

triggered immunity by *Magnaporthe oryzae* is necessary for rice blast disease. *The Plant Cell, 24*(1), 322-335.

Miles, A. K., Tan, Y. P., Shivas, R. G., & Drenth, A. (2015). Novel pathotypes of *Elsinoë australis* associated with *Citrus australasica* and *Simmondsia chinensis* in Australia. *Tropical Plant Pathology, 40*(1), 26-34.

Mohanta, T. K., & Bae, H. (2015). The diversity of fungal genome (Vol. 17): BioMed Central Ltd.

Moolhuijzen, P., See, P. T., Hane, J. K., Shi, G., Liu, Z., Oliver, R. P., & Moffat, C. S. (2018). Comparative genomics of the wheat fungal pathogen *Pyrenophora tritici-repentis* reveals chromosomal variations and genome plasticity. *BMC genomics, 19*(1), 279.

Nei, M. (2000). *Molecular evolution and phylogenetics*. Oxford ;: Oxford University Press.

Nekrasov, V., Wang, C., Win, J., Lanz, C., Weigel, D., & Kamoun, S. (2017). Rapid generation of a transgene-free powdery mildew resistant tomato by genome deletion. *Scientific reports, 7*(1), 1-6.

Nguyen, Q. B., Itoh, K., Van Vu, B., Tosa, Y., & Nakayashiki, H. (2011). Simultaneous silencing of endo-β-1,4 xylanase genes reveals their roles in the virulence of *Magnaporthe oryzae*. *Molecular Microbiology, 81*(4), 1008-1019. doi:10.1111/j.1365-2958.2011.07746.x

Noda, J., Brito, N., & Gonzalez, C. (2010). The *Botrytis cinerea* xylanase Xyn11A contributes to virulence with its necrotizing activity, not with its catalytic activity.(Research article)(Report). *BMC Plant Biology, 10*, 38.

Nowara, D., Gay, A., Lacomme, C., Shaw, J., Ridout, C., Douchkov, D., . . .
Schweizer, P. (2010). HIGS: host-induced gene silencing in the
obligate biotrophic fungal pathogen Blumeria graminis. *The Plant Cell,
22*(9), 3130-3141.

Osbourn, A., Clarke, B., Dow, J., & Daniels, M. (1991). Partial
characterization of avenacinase from *Gaeumannomyces graminis* var.
avenae. *Physiological and Molecular Plant Pathology, 38*(4), 301-312.

O'Sullivan, J. Leaf and stem scab (*E. batatas*). Retrieved from
http://keys.lucidcentral.org/keys/sweetpotato/key/Sweetpotato%20Dia
gnotes/Media/Html/TheProblems/DiseasesFungal/Scab/Scab.htm on
22nd July 2017.

Panwar, V., McCallum, B., & Bakkeren, G. (2013). Host-induced gene
silencing of wheat leaf rust fungus *Puccinia triticina* pathogenicity
genes mediated by the Barley stripe mosaic virus. *Plant molecular
biology, 81*(6), 595-608.

Paudyal, D. P., & Hyun, J. W. (2015). Physical changes in satsuma mandarin
leaf after infection of *Elsinoë fawcettii* causing citrus scab disease.
The plant pathology journal, 31(4), 421.

Paudyal, D. P., Hyun, J. W., & Hwang, R. Y. (2017). Infection and symptom
development by citrus scab pathogen Elsinoë fawcettii on leaves of
satsuma mandarin. European Journal of Plant Pathology, 148(4), 807-
816.

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

Petersen, T. N., Brunak, S., Heijne, G. V., & Nielsen, H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature Methods, 8*(10), 785. doi:10.1038/nmeth.1701

Pierleoni, A., Martelli, P., & Casadio, R. (2008). PredGPI: A GPI-anchor predictor. *BMC Bioinformatics, 9*(1), <xocs:firstpage xmlns:xocs=""/>. doi:10.1186/1471-2105-9-392

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

Poppe, S., Dorsheimer, L., Happel, P., & Stukenbrock, E. H. (2015). Rapidly evolving genes are key players in host specialization and virulence of the fungal wheat pathogen *Zymoseptoria tritici* (*Mycosphaerella graminicola*). PLoS Pathog, 11(7), e1005055.

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

Quinlan, A. R. (2014). BEDTools: the Swiss-army tool for genome feature analysis. *Current protocols in bioinformatics, 47*(1), 11.12. 11-11.12. 34.

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

Raffaele, S., & Kamoun, S. (2012). Genome evolution in filamentous plant pathogens: why bigger can be better (Vol. 10, pp. 417): Nature Publishing Group.

Roberts, F., Roberts, C. W., Johnson, J. J., Kyle, D. E., Krell, T., Coggins, J. R., ... & McLeod, R. (1998). Evidence for the shikimate pathway in apicomplexan parasites. Nature, 393(6687), 801-805.

Rodriguez-Moreno, L., Ebert, M. K., Bolton, M. D., & Thomma, B. P. H. J. (2018). Tools of the crook- infection strategies of fungal plant pathogens. *Plant Journal, 93*(4), 664-674. doi:10.1111/tpj.13810

Rogers, L. M., Kim, Y. K., Guo, W., Gonzalez-Candelas, L., Li, D., & Kolattukudy, P. E. (2000). Requirement for either a host- or pectin-induced pectate lyase for infection of *Pisum sativum* by *Nectria hematococca*. *Proceedings of the National Academy of Sciences of the United States, 97*(17), 9813. doi:10.1073/pnas.160271497

Rooney, H. C., van't Klooster, J. W., van der Hoorn, R. A., Joosten, M. H., Jones, J. D., & de Wit, P. J. (2005). *Cladosporium* Avr2 inhibits tomato Rcr3 protease required for Cf-2-dependent disease resistance. *Science, 308*(5729), 1783-1786.

Rouxel, T., Grandaubert, J., Hane, J. K., Hoede, C., Van De Wouw, A. P., Couloux, A., . . . Howlett, B. J. (2011). Effector diversification within compartments of the *Leptosphaeria maculans* genome affected by Repeat-Induced Point mutations. *Nature Communications, 2*(202), 202. doi:10.1038/ncomms1189

Sánchez, M., Avhad, M. R., Marchetti, J. M., Martínez, M., & Aracil, J. (2016). Jojoba oil: a state of the art review and future prospects. *Energy Conversion and Management, 129*, 293-304.

Saunders, D. G., Win, J., Cano, L. M., Szabo, L. J., Kamoun, S., & Raffaele, S. (2012). Using hierarchical clustering of secreted protein families to classify and rank candidate effectors of rust fungi. *PLoS ONE, 7*(1), e29847.

Scheper, R., Wood, P., & Fisher, B. (2013). Isolation, spore production and Koch's postulates of *Elsinoe pyri. NZ Plant Prot, 66*, 308-316.

Selker, E. U. (1990). Premeiotic Instability of Repeated Sequences in *Neurospora crassa. Annual Review of Genetics, 24*(1), 579-613. doi:10.1146/annurev.ge.24.120190.003051

Selker, E. U., Cambareri, E. B., Jensen, B. C., & Haack, K. R. (1987). Rearrangement of duplicated DNA in specialized cells of *Neurospora*. *Cell, 51*(5), 741-752. doi:10.1016/0092-8674(87)90097-3

Shanmugam, G., Jeon, J., & Hyun, J. W. (2020). Draft genome sequences of *Elsinoë fawcettii* and *Elsinoë australis* causing scab diseases on citrus. *Molecular plant-microbe interactions, 33*(2), 135-137.

Shieh, M. T., Brown, R. L., Whitehead, M. P., Cary, J. W., Cotty, P. J., Cleveland, T. E., & Dean, R. A. (1997). Molecular genetic evidence for the involvement of a specific polygalacturonase, P2c, in the invasion and spread of *Aspergillus flavus* in cotton bolls. *Applied and Environmental Microbiology, 63*(9), 3548.

Silveira, S.F., Shivas, R.G. & McTaggart, A.R. (2007). Sour Orange Scab (*Elsinoë fawcettii*). Retrieved from http://www.padil.gov.au on 20th July 2017.

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics, 31*(19), 3210-3212. doi:10.1093/bioinformatics/btv351

Smit, A. F. A., Hubley, R., & Green, P. (2015). RepeatMasker Open-4.0. Obtained from http://www.repeatmasker.org.

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

Sperschneider, J., Dodds, P. N., Gardiner, D. M., Manners, J. M., Singh, K.
    B., & Taylor, J. M. (2015). Advances and challenges in computational
    prediction of effectors from plant pathogenic fungi. *PLoS Pathogens,
    11*(5), e1004806. doi:10.1371/journal.ppat.1004806

Sperschneider, J., Dodds, P. N., Gardiner, D. M., Singh, K. B., & Taylor, J.
    M. (2018). Improved prediction of fungal effector proteins from
    secretomes with EffectorP 2.0. *Molecular plant pathology, 19*(9),
    2094-2110. doi:10.1111/mpp.12682

Sperschneider, J., Gardiner, D. M., Dodds, P. N., Tini, F., Covarelli, L.,
    Singh, K. B., . . . Taylor, J. M. (2016). EffectorP: predicting fungal
    effector proteins from secretomes using machine learning. *New
    Phytologist, 210*(2), 743-761.

Sperschneider, J., Ying, H., Dodds, P. N., Gardiner, D. M., Upadhyaya, N.
    M., Singh, K. B., . . . Taylor, J. M. (2014). Diversifying selection in the
    wheat stem rust fungus acts predominantly on pathogen-associated
    gene families and reveals candidate effectors. *Frontiers in Plant
    Science, 5*, 372.

Stam, R., Münsterkötter, M., Pophaly, S. D., Fokkens, L., Sghyer, H.,
    Güldener, U., … & Hess, M. (2018). A new reference genome shows
    the one-speed genome structure of the barley pathogen Ramularia
    collo-cygni. Genome biology and evolution, 10(12), 3243-3249.

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

Stergiopoulos, I., Collemare, J., Mehrabi, R., & De Wit, P. J. (2013). Phytotoxic secondary metabolites and peptides produced by plant pathogenic Dothideomycete fungi. *FEMS microbiology reviews, 37*(1), 67-93.

Stukenbrock, E. H., Jørgensen, F. G., Zala, M., Hansen, T. T., McDonald, B. A., & Schierup, M. H. (2010). Whole-genome and chromosome evolution associated with host adaptation and speciation of the wheat pathogen *Mycosphaerella graminicola*. *PLoS Genet, 6*(12), e1001189.

Syme, R. A., Hane, J. K., Friesen, T. L., & Oliver, R. P. (2013). Resequencing and comparative genomics of *Stagonospora nodorum*: Sectional gene absence and effector discovery. *G3: Genes, Genomes, Genetics, 3*(6), 959-969. doi:10.1534/g3.112.004994

Tamura, K., & Nei, M. (1993). Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular biology and evolution, 10*(3), 512-526.

Tan, M., Timmer, L., Broadbent, P., Priest, M., & Cain, P. (1996). Differentiation by molecular analysis of *Elsinoë* spp. causing scab diseases of citrus and its epidemiological implications. *Phytopathology, 86*(10), 1039-1044.

Tanaka, S., Brefort, T., Neidig, N., Djamei, A., Kahnt, J., Vermerris, W., . . . Kahmann, R. (2014). A secreted *Ustilago maydis* effector promotes virulence by targeting anthocyanin biosynthesis in maize. *eLife, 2014*(3), <xocs:firstpage xmlns:xocs=""/>. doi:10.7554/eLife.01355.001

Tang, L., Yang, G., Ma, M., Liu, X., Li, B., Xie, J., . . . Chen, W. (2020). An effector of a necrotrophic fungal pathogen targets the calcium-sensing receptor in chloroplasts to inhibit host resistance. *Molecular plant pathology, 21*(5), 686-701.

Temporini, E. D., & VanEtten, H. D. (2004). An analysis of the phylogenetic
distribution of the pea pathogenicity genes of *Nectria haematococca*
MPVI supports the hypothesis of their origin by horizontal transfer and
uncovers a potentially new pathogen of garden pea: *Neocosmospora
boniensis*. Current Genetics, 46(1), 29-36.

Testa, A. C., Oliver, R. P., & Hane, J. K. (2016). OcculterCut: a
comprehensive survey of AT-rich regions in fungal genomes. *Genome
biology and evolution, 8*(6), 2044-2064.

The Uniprot Consortium. (2018). UniProt: the universal protein
knowledgebase. Nucleic Acids Research, 46(5), 2699.

Thomma, B. P. H. J., Seidl, M. F., Shi-Kunne, X., Cook, D. E., Bolton, M. D.,
van Kan, J. A. L., & Faino, L. (2016). Mind the gap; seven reasons to
close fragmented genome assemblies. *Fungal Genetics and Biology,
90*, 24-30. doi:10.1016/j.fgb.2015.08.010

Thorvaldsdóttir, H., Robinson, J. T., & Mesirov, J. P. (2013). Integrative
Genomics Viewer (IGV): high-performance genomics data
visualization and exploration. *Briefings in bioinformatics, 14*(2), 178-
192.

Timmer, L., Priest, M., Broadbent, P., & Tan, M. (1996). Morphological and
pathological characterization of species of *Elsinoë* causing scab
diseases of citrus. *Phytopathology, 86*(10), 1032-1038.

Timmer, L. W., Garnsey, S. M., & Graham, J. H. (2000). *Compendium of
Citrus Diseases* (2nd ed.).

Tsushima, A., Gan, P., Kumakura, N., Narusaka, M., Takano, Y., Narusaka,
Y., & Shirasu, K. (2019). Genomic plasticity mediated by transposable

elements in the plant pathogenic fungus *Colletotrichum higginsianum*. *Genome biology and evolution, 11*(5), 1487-1500.

Urban, M., Cuzick, A., Rutherford, K., Irvine, A., Pedro, H., Pant, R., . . . Hammond-Kosack, K. E. (2017). PHI-base: A new interface and further additions for the multi-species pathogen-host interactions database. *Nucleic Acids Research, 45*(1), D604-D610. doi:10.1093/nar/gkw1089

Valette-Collet, O., Cimerman, A., Reignault, P., Levis, C., & Boccara, M. (2003). Disruption of *Botrytis cinerea* pectin methylesterase gene Bcpme1 reduces virulence on several host plants. *Molecular plant-microbe interactions, 16*(4), 360-367. doi:10.1094/MPMI.2003.16.4.360

van Esse, H. P., van't Klooster, J. W., Bolton, M. D., Yadeta, K. A., Van Baarlen, P., Boeren, S., . . . Thomma, B. P. (2008). The *Cladosporium fulvum* virulence protein Avr2 inhibits host proteases required for basal defense. *The Plant Cell, 20*(7), 1948-1963.

Van Kan, J. A., Stassen, J. H., Mosbach, A., Van Der Lee, T. A., Faino, L., Farmer, A. D., . . . Cottam, E. (2017). A gapless genome sequence of the fungus *Botrytis cinerea*. *Molecular plant pathology, 18*(1), 75-89.

Ve, T., Williams, S. J., Catanzariti, A. M., Rafiqi, M., Rahman, M., Ellis, J. G., . . . Kobe, B. (2013). Structures of the flax-rust effector AvrM reveal insights into the molecular basis of plant-cell entry and effector-triggered immunity.(PLANT BIOLOGY)(Report)(Author abstract). *Proceedings of the National Academy of Sciences of the United States, 110*(43), 17594. doi:10.1073/pnas.1307614110

Verma, S., Gazara, R. K., Nizam, S., Parween, S., Chattopadhyay, D., & Verma, P. K. (2016). Draft genome sequencing and secretome

analysis of fungal phytopathogen *Ascochyta rabiei* provides insight into the necrotrophic effector repertoire. *Scientific reports, 6*(1), 1-14.

Walkowiak, S., Rowland, O., Rodrigue, N., & Subramaniam, R. (2016). Whole genome sequencing and comparative genomics of closely related Fusarium Head Blight fungi: *Fusarium graminearum, F. meridionale* and *F. asiaticum. BMC genomics, 17*(1), 1014.

Wang, L. Y., Bau, H. J., & Chung, K. R. (2009). Accumulation of elsinochrome phytotoxin does not correlate with fungal virulence among *Elsinoë fawcettii* isolates in Florida. *Journal of Phytopathology, 157*(10), 602-608.

Wang, L. Y., Liao, H. L., Bau, H. J., & Chung, K. R. (2009). 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, 31*(1), 28-37.

Wang, Q., Jiang, C., Wang, C., Chen, C., Xu, J. R., & Liu, H. (2017). Characterization of the two-speed subgenomes of *Fusarium graminearum* reveals the fast-speed subgenome specialized for adaption and infection. *Frontiers in Plant Science, 8*, 140.

Wang, X., Jiang, N., Liu, J., Liu, W., & Wang, G.-L. (2014). The role of effectors and host immunity in plant–necrotrophic fungal interactions (Vol. 5, pp. 722-732): Taylor & Francis.

Wang, X., Xue, B., Dai, J., Qin, X., Liu, L., Chi, Y., ... & Li, H. (2018). A novel Meloidogyne incognita chorismate mutase effector suppresses plant immunity by manipulating the salicylic acid pathway and functions mainly during the early stages of nematode parasitism. Plant pathology, 67(6), 1436-1448.

Whiteside, J. (1975). Biological characteristics of *Elsinoe fawcettii* pertaining to the epidemiology of sour orange scab. *Phytopathology, 65*(10), 1170-1175.

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

Wilken, S.E., Seppälä, S., Lankiewicz, T. S., Saxena, M., Henske, J. K., Salamov, A. A., Grigoriev, I. V., & O'Malley, M. A. (2020). Genomic and proteomic biases inform metabolic engineering strategies for anaerobic fungi. Metabolic engineering communications, 10, e00107.

Winter, D. J., Ganley, A. R., Young, C. A., Liachko, I., Schardl, C. L., Dupont, P.-Y., . . . Cox, M. P. (2018). Repeat elements organise 3D genome structure and mediate transcription in the filamentous fungus *Epichloë festucae*. PLoS Genetics, 14(10), e1007467.

Xin, H., Huang, F., ZHANG, T., XU, J., Hyde, K. D., & LI, H. (2014). Pathotypes and genetic diversity of Chinese collections of *Elsinoë fawcettii* causing citrus scab. *Journal of Integrative Agriculture, 13*(6), 1293-1302.

Xu, Q., Tang, C., Wang, X., Sun, S., Zhao, J., Kang, Z., & Wang, X. (2019). An effector protein of the wheat stripe rust fungus targets chloroplasts and suppresses chloroplast function. *Nature Communications, 10*(1), 1-13.

Yakoby, N., Beno-Moualem, D., Keen, N. T., Dinoor, A., Pines, O., & Prusky, D. (2001). *Colletotrichum gloeosporioides* pelB is an important virulence factor in avocado fruit-fungus interaction. *Molecular plant-microbe interactions, 14*(8), 988-995. doi:10.1094/MPMI.2001.14.8.988

Yin, Y., Mao, X., Yang, J., Chen, X., Mao, F., & Xu, Y. (2012). DbCAN: A web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Research, 40*(1), W445-W451. doi:10.1093/nar/gks479

Yoshida, K., Saunders, D. G., Mitsuoka, C., Natsume, S., Kosugi, S., Saitoh, H., ... & Terauchi, R. (2016). Host specialization of the blast fungus Magnaporthe oryzae is associated with dynamic gain and loss of genes linked to transposable elements. BMC genomics, 17(1), 1-18.

Yu, Y., Xiao, J., Zhu, W., Yang, Y., Mei, J., Bi, C., . . . Tan, W. (2017). Ss-Rhs1, a secretory Rhs repeat-containing protein, is required for the virulence of *Sclerotinia sclerotiorum*. *Molecular plant pathology, 18*(8), 1052-1061. doi:10.1111/mpp.12459

Zerbino, D. R., & Birney, E. (2008). Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome research, 18*(5), 821-829. doi:10.1101/gr.074492.107

Zhang, H., Yohe, T., Huang, L., Entwistle, S., Wu, P., Yang, Z., . . . Yin, Y. (2018). DbCAN2: A meta server for automated carbohydrate-active enzyme annotation. *Nucleic Acids Research, 46*(1), W95-W101. doi:10.1093/nar/gky418

Zhou, X., Liu, W., Wang, C., Xu, Q., Wang, Y., Ding, S., & Xu, J. R. (2011). A MADS-box transcription factor MoMcm1 is required for male fertility, microconidium production and virulence in *Magnaporthe oryzae*. Molecular Microbiology, 80(1), 33-53.

Zuppini, A., Navazio, L., Sella, L., Castiglioni, C., Favaron, F., & Mariani, P. (2005). An endopolygalacturonase from *Sclerotinia sclerotiorum* induces calcium-mediated signaling and programmed cell death in soybean cells. *Molecular plant-microbe interactions, 18*(8), 849-855. doi:10.1094/MPMI-18-0849

APPENDIX A: METHODS FOR FUNGAL CULTURE AND WGS
SEQUENCING OF *ELSINOË* ISOLATES

For the *Elsinoë australis* JBS pathotype isolates, single-spore cultures were
bulked in 250 ml of clarified Campbell's V8 juice (20%), within 500 ml sterile
flasks. Broth cultures were incubated at 27 ºC on an orbital shaker for 14
days @ 150 rpm. Following incubation, cultures were collected on
Whatmann filter papers, under vacuum. The fungal biomass was air dried
within a laminar flow cabinet for 1 hour and then collected and weighed
before being snap frozen in liquid nitrogen. Genomic DNA was extracted
from single-spored cultures using the NucleoSpin Plant II (Machery Nagel)
DNA extraction kit, following manufacturer's instructions. Following QA via
gel electrophoresis and spectrophotometric analysis, purified DNA samples
underwent paired-end sequencing at the AGRF. Illumina ligation libraries for
sequencing were prepared following manufacturer's directions, with TruSeq
adapters. Whole genome sequencing was conducted on an Illumina HiSeq
according to Illumina protocols for 150b paired-end reads. Raw sequence
reads were trimmed via Trimmomatic, for the removal of Illumina adapters,
maintaining base quality of >Q30, and a minimum read length of 50 bases.

The *E. fawcettii* (BRIP 53147a) isolate was cultured on potato dextrose agar
(Difco) and incubated at 23-25 °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 were prepared according to
Illumina NexteraTM DNA Flex Library Prep Reference Guide using a
NexteraTM DNA Flex Library Prep Kit and NexteraTM DNA CD Indexes.
WGS sequencing was performed on Illumina MiSeq platform (600-cycles) at
the molecular laboratories of the Centre for Crop Health, USQ.

The *E. fawcettii* (BRIP 54245a, 54425a and 54434a) and *E. australis* (FL)
(BRIP 52616a) isolates were cultured on malt extract agar and incubated at
25 °C for three weeks. Whole genomic DNA was extracted using the Gentra
Puregene Tissue Kit (QIAGEN) according to the manufacturer's protocol with

one exception; twice the amount of Proteinase K was used during cell lysis. Paired end libraries were prepared and WGS sequencing performed on Illumina HiSeq platform at the Australian Genome Research Facility (AGRF).

APPENDIX B: SUPPLEMENTARY INFORMATION FOR PHYLOGENETIC
ANALYSES

**Table 1. GenBank accessions of TEF1-α and ITS regions included in the
phylogenetic analysis with the species of *Elsinoë* from Australian and
New Zealand utilised in tree I (Figure 2).**

Species	Strain	GenBank Accession	
		ITS	TEF1-α
Elsinoë fawcettii	BRIP 53147a	MN784182	MN787508
E. fawcettii	BRIP 54245a	WLYY00000000 Contig 157, 1806:2328	WLYY00000000 Contig 5, 207668:208034
E. fawcettii	BRIP 54425a	WLYZ00000000 Contig 4, 740609:741131	WLYZ00000000 Contig 7, 2893:3259
E. fawcettii	BRIP 54434a	WLZA00000000 Contig 39, 193156:193678	WLZA00000000 Contig 22, 2891:3257
E. australis-FL	BRIP 52616a	WLZB00000000 Contig 43, 2698:3213	WLZB00000000 Contig 3, 863844:864213
E. australis-JBS	Hillston-1	QGIJ00000000 Contig 7, 4225:3710	QGIJ00000000 Contig 823, 47312:46943
E. australis -JBS	Hillston-2	PTQR00000000 Contig 140, 4857:4342	PTQR00000000 Contig 128, 323353:322984
E. australis-JBS	Hillston-3	QGIG00000000 Contig 19, 4428:4943	QGIG00000000 Contig 227, 64462:64831
E. australis-JBS	Wagga-1	QGII00000000 Contig 4, 4909:4394	QGII00000000 Contig 98, 845956:846325
E. australis-JBS	Wagga-2	PTQP00000000 Contig 7, 1:512	PTQP00000000 Contig 18, 63171:63540
E. australis-JBS	Wagga-3	QGIH00000000	QGIH00000000

		Contig 8, 2737:3252	Contig 193, 369290:368921
E. australis-JBS	Forbes-1	PTQO00000000 Contig 16, 4857:4342	PTQO00000000 Contig 129, 370600:370231
E. australis-JBS	Forbes-2	PTQQ00000000 Contig 3, 4701:4186	PTQQ00000000 Contig 42, 741780:742149
E. ampelina	GAAUS9	KY684885.1	KY684920.1
E. piri	CBS 179.82	KX887268.1	KX886913.1
E. banksiicola	CBS 113734	KX887199.1	KX886845.1
E. leucospermi	CBS 112367	KX887246.1	KX886891.1
E. eucalypticola	CBS 124765	KX887215.1	KX886861.1
E. tectificae	CBS 124777	KX887292.1	KX886937.1
E. tiliae	CBS 350.73	KX887296.1	KX886940.1
E. sp. *'eucalyptorum'*	CBS 120084	KX887216.1	KX886862.1
Myriangium hispanicum	CBS 247.33	KX887304.1	KX886948.1

Table 2. GenBank accessions of TEF1-α, LSU, RPB2 and ITS regions included in the phylogenetic analysis with closely related species to the *E. australis* JBS pathotype and *E. australis* Finger Lime pathotype utilised in tree II (Figure 3).

Species	Strain	Genbank accession			
		TEF1-α	LSU	RPB2	ITS
E. australis-FL	BRIP 52616a	WLZB00000000 Contig 3, 863844:864213	WLZB00000000 Contig 43, 3371:4108	WLZB00000000 Contig10, 576105:576848	WLZB00000000 Contig 43, 2698:3213
E. australis-JBS	Hillston-1	QGIJ00000000 Contig 823, 47312:46943	QGIJ00000000 Contig 7 3552:2815	QGIJ00000000 Contig 717 24342:23599	QGIJ00000000 Contig 7, 4225:3710
E. australis-JBS	Hillston-2	PTQR00000000 Contig 128, 323353:322984	PTQR00000000 Contig 140 4184:3447	PTQR00000000 Contig 111 34142:33399	PTQR00000000 Contig 140, 4857:4342
E. australis-JBS	Hillston-3	QGIG00000000 Contig 227, 64462:64831	QGIG00000000 Contig 19 3533:4270	QGIG00000000 Contig 362 107635:108378	QGIG00000000 Contig 19, 4428:4943
E. australis-JBS	Wagga-1	QGII00000000 Contig 98, 845956:846325	QGII00000000 Contig 4 4236:3499	QGII00000000 Contig 461 29962:29219	QGII00000000 Contig 4, 4909:4394
E. australis-JBS	Wagga-2	PTQP00000000 Contig 18, 63171:63540	PTQP00000000 Contig 7 670:1407	PTQP00000000 Contig 392 176689:177432	PTQP00000000 Contig 7, 1:512
E. australis-JBS	Wagga-3	QGIH00000000 Contig 193, 369290:368921	QGIH00000000 Contig 8 3410:4147	QGIH00000000 Contig 259 34183:33440	QGIH00000000 Contig 8, 2737:3252
E. australis-JBS	Forbes-1	PTQO00000000 Contig 129, 370600:370231	PTQO00000000 Contig 16 4184:3447	PTQO00000000 Contig 363 112167:112910	PTQO00000000 Contig 16, 4857:4342
E. australis-JBS	Forbes-2	PTQQ00000000 Contig 42, 741780:742149	PTQQ00000000 Contig 3 4028:3291	PTQQ00000000 Contig 313 111906:112649	PTQQ00000000 Contig 3, 4701:4186
Elisnoë abutilonis	CBS 510.50[T]	KX886831.1	KX886949.1	KX887068.1	KX887185.1
E. anacardii	CBS 470.62[T]	KX886835.1	KX886953.1	KX887072.1	KX887189.1
E. australis	CBS 229.64	KX886842.1	KX886960.1	KX887079.1	KX887196.1
E. australis	CBS 314.32[T]	KX886844.1	KX886962.1	KX887081.1	KX887198.1
E. centrolobii	CBS 222.50[T]	KX886852.1	KX886969.1	KX887089.1	KX887206.1
E. fici-caricae	CBS 473.62[T]	KX886870.1	KX886987.1	KX887106.1	KX887224.1
E. freyliniae	CBS 128204[T]	KX886872.1	KX886989.1	KX887108.1	KX887226.1
E. genipae-americanae	CBS 516.50[T]	KX886874.1	KX886991.1	KX887110.1	KX887228.1

E. ichnocarpi	CBS 475.62[T]	KX886878.1	KX886995.1	KX887114.1	KX887232.1
E. mattiroloanum	CBS 287.64	KX886895.1	KX887013.1	KX887131.1	KX887250.1
E. oleae	CBS 227.59[T]	KX886901.1	KX887019.1	KX887137.1	KX887256.1
E. piri	CBS 163.29	KX886912.1	KX887030.1	KX887148.1	KX887267.1
E. piri	CBS 179.82	KX886913.1	KX887031.1	KX887149.1	KX887268.1
E. punicae	CPC 19968	KX886921.1	KX887039.1	KX887157.1	KX887276.1
E. rosarum	CBS 212.33[T]	KX886928.1	KX887046.1	KX887163.1	KX887283.1
E. rosarum	CBS 213.33	KX886929.1	KX887047.1	KX887164.1	KX887284.1
E. rosarum	CBS 235.64	KX886930.1	KX887048.1	KX887165.1	KX887285.1
E. salicina	CPC 17824[T]	KX886931.1	KX887049.1	KX887166.1	KX887286.1
E. semecarpi	CBS 477.62[T]	KX886932.1	KX887050.1	KX887167.1	KX887287.1
E. violae	CBS 336.35[T]	KX886946.1	KX887065.1	KX887182.1	KX887302.1
Myriangium hispanicum	CBS 247.33	KX886948.1	KX887067.1	KX887184.1	KX887304.1

[T]Type strains

Table 3. GenBank accessions included in the phylogenetic analysis with closely related species to *E. fawcettii*, utilised in tree III (Figure 4).

| Species | Strain | GenBank Accession | |
		ITS	TEF1-α
Elsinoë fawcettii	BRIP 53147a	MN784182	MN787508
E. fawcettii	BRIP 54245a	WLYY00000000 Contig 157, 1806:2328	WLYY00000000 Contig 5, 207668:208034
E. fawcettii	BRIP 54425a	WLYZ00000000 Contig 4, 740609:741131	WLYZ00000000 Contig 7, 2893:3259
E. fawcettii	BRIP 54434a	WLZA00000000 Contig 39, 193156:193678	WLZA00000000 Contig 22, 2891:3257
E. fawcettii	SM3-1	FJ010360.2	FJ010270.2
E. fawcettii	Jin-1	FJ010320.1	FJ010244.1
E. fawcettii	Jin-6	FJ010323.2	FJ010247.2
E. fawcettii	S38162	FJ010343.2	FJ010267.1
E. fawcettii	DAR 70187	FJ010290.2	FJ010214.1
E. fawcettii	CC-132	FJ010297.2	FJ010222.1
E. fawcettii	CC-3	FJ010295.1	FJ010220.2
E. fawcettii	DAR 70024	FJ010307.2	FJ010231.1
E. fawcettii	CBS 139.25	KX887219.1	KX886865.1
E. fawcettii	CBS 231.64	KX887220.1	KX886866.1
E. fawcettii	CBS 232.64	KX887221.1	KX886867.1
E. fawcettii	CBS 233.64	KX887222.1	KX886868.1
E. citricola	CPC 18535	KX887207.1	KX886853.1
E. citricola	CPC 18570	KX887208.1	KX886854.1
E. pitangae	CBS 227.50	KX887269.1	KX886914.1
E. fagarae	CBS 514.50	KX887218.1	KX886864.1
E. caleae	CBS 221.50	KX887205.1	KX886851.1
E. diospyri	CBS 223.50	KX887210.1	KX886856.1
E. fici-caricae	CBS 473.62	KX887224.1	KX886870.1
E. flacourtiae	CBS 474.62	KX887225.1	KX886871.1
E. zizyphi	CBS 378.62	KX887303.1	KX886947.1
E. eucalypticola	CBS 124765	KX887215.1	KX886861.1
E. tectificae	CBS 124777	KX887292.1	KX886937.1
Myriangium hispanicum	CBS 247.33	KX887304.1	KX886948.1

Table 4. GenBank accessions and genomic locations for RPB2 and TEF1-α sequences included in the phylogenetic analysis of *Elsinoë* and comparative species, utilised in tree IV (Figure 5).

Species (isolate), GenBank assembly accession	Genomic location	
	RPB2	TEF1-α
Elsinoë fawcettii (BRIP 53147a), SDJM00000000	SDJM01000016.1 566337:567078	Accession: MN787508
E. australis FL (BRIP 52616a), WLZB00000000	WLZB01000010.1 576105: 576846	WLZB01000003.1 863844: 864211
E. australis JBS (Hillston-2), PTQR00000000	PTQR01000111.1 33401: 34142	PTQR01000128.1 322984: 323348
E. fawcettii (DAR 70024), GCA_007556565.1	SWCR01000004.1 1491439:1492180	SWCR01000007.1 905414:905781
E. fawcettii (SM16-1), GCA_007556535.1	VAAB01000009.1 527249:527992	VAAB01000010.1 810556:810923
E. ampelina (YL-1), GCA_005959805.1	SMYM01000002.1 775093:775834	SMYM01000001.1 359907:360274
E. australis (Arg-1), GCA_007556505.1	SWCS01000005.1 550352:551093	SWCS01000007.1 804446:804813
Zymoseptoria tritici (ST99CH 1E4), GCA_900184115.1	LT854256.1 920363:921522	LT854256.1 2059985:2060352
Parastagonospora nodorum (SN15), GCF_000146915.1	NW_001884567.1 399978:400673	NW_001884568.1 56860:57227
Leptosphaeria maculans (JN3), GCF_000230375.1	NW_003533871.1 571201:571966	NW_003533847.1 1591860:1592227
Pyrenophora tritici-repentis (DW5), GCA_003231415.1	MUXC01000023.1 96534:97465	MUXC01000280.1 25823:26190
Rhynchosporium commune (UK7), GCA_900074885.1	FJUW01000001.1 941326:945224	FJUW01000015.1 587359:587709
Botrytis cinerea (B05.10), GCF_000143535.2	NC_037323.1 703730:704407	NC_037318.1 2020091:2020458
Sclerotinia sclerotiorum (UF-70), GCF_000146945.2	NW_001820835.1 2290220:2291017	NW_001820830.1 1823579:1823946
Pyricularia oryzae (syn. Magnaporthe oryzae) (70-15), GCF_000002495.2	NC_017844.1 6879084:6883798	NC_017851.1 1161008:1161375

Species	Strain	Genbank accession		
		β-tubulin	Histone 4	Pma1
Verticillium dahlia (Vdls.17), GCF_000150675.1		NW_009276922.1 1110816:1115035	NW_009276925.1 665591:665958	
Ustilago maydis (521), GCF_000328475.2		NC_026489.1 572250:576082	NC_026479.1 314654:315013	
Spizellomyces punctatus		DQ302773.1	XM_016754690.1	

Table 5. GenBank accessions and genomic locations for the partial β-tubulin (Tree V), Histone 4 (Tree VI) and *Pma1* (Tree VII) gene sequences included in the phylogenetic analyses of *Elsinoë* (Figures 6, 7 and 8).

Species	Strain	Genbank accession		
		β-tubulin	Histone 4	*Pma1*
E. australis Finger Lime pathotype	BRIP 52616a	WLZB01000001.1 1106915:1107329	WLZB01000022.1 264264:264622	WLZB01000003.1 314075:312588
E. australis Sweet Orange pathotype	Arg-1	SWCS01000001.1 1320277:1319863	SWCS01000006.1 1152435:1152793	SWCS01000007.1 1342204:1344024
E. australis Jojoba Black Scab (JBS) pathotype	Hillston-1	QGIJ01000379.1 133050:132644	QGIJ01000556.1 9884:10251	QGIJ01000485.1 6851:5364
E. australis JBS pathotype	Hillston-2	PTQR01000114.1 491411:491005	PTQR01000036.1 23802:24169	PTQR01000012.1 194170:192683
E. australis JBS pathotype	Hillston-3	QGIG01000537.1 291927:291521	QGIG01000401.1 351557:351924	QGIG01000226.1 677415:678902
E. australis JBS pathotype	Wagga-1	QGII01000353.1 229624:230030	QGII01000296.1 33222:33589	QGII01000422.1 265255:266742
E. australis JBS pathotype	Wagga-2	PTQP01000323.1 491356:490950	PTQP01000179.1 23923:24290	PTQP01000453.1 203035:204522
E. australis JBS pathotype	Wagga-3	QGIH01000437.1 504401:503995	QGIH01000436.1 33697:34064	QGIH01000425.1 945365:946852
E. australis JBS pathotype	Forbes-1	PTQO01000518.1 481491:481085	PTQO01000395.1 26788:27155	PTQO01000197.1 203031:204518

E. australis JBS pathotype	Forbes-2	PTQQ01000163.1 12070:12476	PTQQ01000244.1 33387:33754	PTQQ01000333.1 194183:192696
E. ampelina	YL-1	SMYM01000009.1 1017379:1016965	SMYM01000005.1 1804477:1804835	SMYM01000001.1 1595078:1593565
E. fawcettii	BRIP 53147a	SDJM01000001.1 2121067:2121481	SDJM01000002.1 1582558:1582916	SDJM01000008.1 266647:268158
E. fawcettii	BRIP 54245a	WLYY01000006.1 224777:224363	WLYY01000001.1 398014:398372	WLYY01000005.1 744936:743425
E. fawcettii	BRIP 54425a	WLYZ01000005.1 683007:683421	WLYZ01000001.1 26569:26927	WLYZ01000007.1 539544: 538033
E. fawcettii	BRIP 54434a	WLZA01000005.1 224764:224350	WLZA01000001.1 397999:398357	WLZA01000025.1 76331:74820
E. fawcettii	SM16-1	VAAB01000002.1 1635508:1635922	VAAB01000001.1 463870:464228	VAAB01000010.1 271384:272895
E. fawcettii	DAR 70024	SWCR01000010.1 684842:685256	SWCR01000001.1 1197390:1197748	SWCR01000007.1 1442695:1441184
Alternaria alternata	MLL3	KJ921779.1	XM_018534962.1	XM_018534612.1

APPENDIX C: DATA FOR THE GENERATION OF COMPARATIVE
GENOMICS FIGURES 9-12 (SECTION 3.1.3)

**Table 1. Comparison of genome assembly size (Mb), number of
predicted genes, protein length and number of scaffolds among newly
sequenced isolates of *Elsino*ë, additional *Elsino*ë spp., and further
fungal pathogens.**

Species/ isolate	GenBank accession	Genome size (Mb)	Predicted genes (not including genes overlapping predicted TE regions)	Median protein length (AA)	Number of scaffolds
Necrotrophic:					
***E. australis* SO pathotype (Arg-1)**	GCA_007556505.1	23.8	9253	417	21
***E. australis* FL pathotype (BRIP 52616a)**	GCA_016625675.1	23.8	9215	419	61
E. australis* JBS pathotype (Wagga-1)	GCA_004761985.1	26.3	9547	411	465
E. australis* JBS pathotype (Wagga-2)	GCA_005382395.1	26.6	9482	411	466
E. australis* JBS pathotype (Wagga-3)	GCA_004761955.1	26.3	9521	411	438
E. australis* JBS pathotype (Hillston-1)	GCA_004764485.1	25	8967	404	1378
***E. australis* JBS**	GCA_005382405.1	26.6	9533	411	484

pathotype (Hillston 2)*					
E. australis **JBS pathotype (Hillston 3)***	GCA_004761995.1	27.1	9570	410	603
E. australis **JBS pathotype (Forbes 1)***	GCA_005382415.1	26.6	9465	410	637
E. australis **JBS pathotype (Forbes 2)***	GCA_005382375.1	26.6	9488	412.5	495
E. fawcettii **(BRIP 53147a)***	GCA_012977835.1	26	10,080	409	286
E. fawcettii **(BRIP 54245a)***	GCA_016625505.1	26.4	10,164	408	475
E. fawcettii **(BRIP 54425a)***	GCA_016625515.1	25.9	10,014	409.5	474
E. fawcettii **(BRIP 54434a)***	GCA_016625525.1	25.9	10,016	410	473
E. fawcettii **(SM16-1)**	GCA_007556535.1	26.7	10,519	404	1266
E. fawcettii **(DAR 70024)**	GCA_007556565.1	26.3	10,223	407	53
E. ampelina **(YL-1)**	GCA_005959805.1	28.3	9,804	409	13
Botrytis cinerea	GCF_000143535.2	42.6	11,481	405	18
Parastagono-spora nodorum	GCF_000146915.1	37.2	15,878	304	108
Pyrenophora tritici-repentis	GCA_003231415.1	33.4	10,772	386	3964

Sclerotinia sclerotiorum	GCF_000146945.2	38.5	13,770	273	37
Zymoseptoria tritici	GCA_900184115.1	38.6	11,936	380	20
Hemibiotrophic:					
Leptosphaeria maculans	GCF_000230375.1	45.1	12,337	331	1717
Pyricularia oryzae (syn. Magnaporthe oryzae)	GCF_000002495.2	41	12,236	359	53
Rhynchosporium commune	GCA_900074885.1	55.6	12,100	373.5	164
Verticillium dahliae	GCF_000150675.1	33.9	10,441	395	55
Biotrophic:					
Ustilago maydis	GCF_000328475.2	19.7	6692	484	27

*Isolates sequenced in the current study

Table 2. Gene ortholog classification among eight isolates of *Elsinoë australis* JBS pathotype (Figure 13).

Isolate	GenBank accession	Core genes	Accessory genes	Unique genes
E. australis **JBS** pathotype (Wagga-1)*	GCA_004761985.1	9310	221	16
E. australis **JBS** pathotype (Wagga-2)*	GCA_005382395.1	9264	214	4
E. australis **JBS** pathotype (Wagga-3)*	GCA_004761955.1	9303	211	7

E. australis JBS pathotype (Hillston-1)*	GCA_004764485.1	8792	132	43
E. australis JBS pathotype (Hillston-2)*	GCA_005382405.1	9307	180	46
E. australis JBS pathotype (Hillston-3)*	GCA_004761995.1	9228	195	147
E. australis JBS pathotype (Forbes-1)*	GCA_005382415.1	9301	157	7
E. australis JBS pathotype (Forbes-2)*	GCA_005382375.1	9300	185	3

*Isolates sequenced in the current study

Table 3. Gene ortholog classification among three pathotypes of *Elsinoë australis* (Figure 14).

Pathotype	GenBank accession	Core genes	Accessory genes	Unique genes
E. australis SO pathotype (Arg-1)	GCA_007556505.1	8460	485	308
E. australis FL pathotype (BRIP 52616a)	GCA_016625675.1	8446	532	237
E. australis JBS pathotype (Hillston-2)*	GCA_005382405.1	8430	519	584

*Isolates sequenced in the current study

Table 4. Gene ortholog classification among six isolates of *Elsinoë fawcettii* (Figure 15).

Isolate	GenBank accession	Core genes	Accessory genes	Unique genes
E. fawcettii (BRIP 53147a)*	GCA_012977835.1	9520	541	19
E. fawcettii (BRIP 54245a)*	GCA_016625505.1	9555	591	18
E. fawcettii (BRIP 54425a)*	GCA_016625515.1	9496	513	5
E. fawcettii (BRIP 54434a)*	GCA_016625525.1	9522	482	12
E. fawcettii (SM16-1)	GCA_007556535.1	9592	223	704
E. fawcettii (DAR 70024)	GCA_007556565.1	9566	619	38

*Isolates sequenced in the current study

Table 5. Gene ortholog classification among the proteomes of five *Elsinoë* spp/pathotypes with a further ten fungal pathogens (Figure 16).

Isolate	Core genes	Ascomycota core genes	Accessory genes	Necrotrophic-specific genes	Core *Elsinoe*-specific genes	Accessory *Elsinoe*-specific genes	Species-specific genes (% of total genes)	Species-specific genes	
								With orthoMCL group ID	No orthoMCL group ID
Necrotrophic:									
E. fawcettii (BRIP 53147a)*	2963	1106	3582 (35.5%)	392	22 (0 with orthoMCL group ID)	1285 (88 with orthoMCL group ID)	730 (7.2%)	52	678
E. australis SO pathotype (Arg-1)	2919	1075	3190 (34.5%)	348	22 (0 with orthoMCL group ID)	1553 (98 with orthoMCL group ID)	146 (1.6%)	9 pathotype-specific	137 pathotype-specific
E. australis FL pathotype (BRIP 52616a)*	2926	1062	3189 (34.6%)	333	22 (0 with orthoMCL group ID)	1579 (90 with orthoMCL group ID)	104 (1.1%)	14 pathotype-specific	90 pathotype-specific
E. australis JBS pathotype (Hillston-2)*	2905	1073	3218 (33.8%)	356	22	1591	368 (3.9%)	7 pathotype-specific	361 pathotype-specific

					(0 with orthoMCL group ID)	(92 with orthoMCL group ID)			
E. ampelina (YL-1)	2874	1083	3330 (34.0%)	379	22 (0 with orthoMCL group ID)	1291 (66 with orthoMCL group ID)	825 (8.4%)	39	786
Botrytis cinerea	3063	1123	4612 (40.2%)	1393	N/A	N/A	1290 (11.2%)	90	1200
Parastagonospora nodorum	3217	1148	5666 (35.7%)	816	N/A	N/A	5031 (31.7%)	113	4918
Pyrenophora tritici-repentis	3024	1081	4561 (42.3%)	721	N/A	N/A	1385 (12.9%)	79	1306
Sclerotinia sclerotiorum	2952	1076	3799 (27.6%)	1394	N/A	N/A	4549 (33.0%)	61	4488
Zymoseptoria tritici	3043	1103	3797 (31.8%)	556	N/A	N/A	3437 (28.8%)	127	3310
Hemi-biotrophic:									
Leptosphaeria maculans	2988	1045	4662 (37.8%)	N/A	N/A	N/A	3642 (29.5%)	49	3593
Pyricularia oryzae	2976	1080	4558 (37.3%)	N/A	N/A	N/A	3622 (29.4%)	209	3413
Rhynchosporium commune	3057	1115	5014 (41.4%)	N/A	N/A	N/A	2914 (24.1%)	118	2796

Verticillium dahliae	3006	1102	4490 (43.0%)	N/A	N/A	N/A	1843 (17.7%)	215	1628
Biotrophic:									
Ustilago maydis	2710	N/A	1550 (23.2%)	N/A	N/A	N/A	2432 (36.3%)	515	1917

*Isolates sequenced in the current study

APPENDIX D: ADDITIONAL DATA FOR THE KNOWN EFFECTOR ANALYSIS (SECTION 3.2.3.3) AND SUBSEQUENT CANDIDATE EFFECTOR PRIORITISATION

Table 1. Genomic features utilised in the known effector analysis and candidate effector prioritisation.

Species	Total genes	# of genes in gene-dense locations (IFR < 1500 bp on both sides)		# of genes in gene-sparse locations (IFR > 1500 bp on both sides)		GC content of CDS (%)			# of genes involved in predicted secondary metabolite clusters (relaxed setting)		Unique genes not assigned an orthoMCL group ID**		Genes allocated the same orthoMCL ID as a known effector		TE coverage (%)	# of genes within 20,000 Kb of a TE region		AT-rich region coverage (%)	Distance of gene to closest AT-rich region, Q1 value (bp)
		#	% of total	#	% of total	Q1	Median	Q3	#	% of total	#	% of total	#	% of total		#	% of total		
Elsinoë isolates:																			
E. australis JBS pathotype:																			
Hillston-1*	8967	5757	64.20	411	4.58	52.75	54.74	57.35	251	2.80	1482	16.53	8	0.09	0.36	983	10.96	3.40	10371
Hillston-2*	9533	6654	69.80	408	4.28	52.80	54.85	57.55	304	3.19	1615	16.94	10	0.10	1.33	2174	22.80	9.13	19917
Hillston-3*	9570	6638	69.36	430	4.49	52.82	54.85	57.56	311	3.25	1683	17.59	10	0.10	1.36	2144	22.40	9.67	19847
Forbes-1*	9465	6556	69.27	405	4.28	52.82	54.86	57.57	304	3.21	1571	16.60	10	0.11	1.29	2141	22.62	9.11	18695
Forbes-2*	9488	6621	69.78	412	4.34	52.81	54.84	57.56	308	3.25	1572	16.57	10	0.11	1.32	2104	22.18	9.33	18488
Wagga-1*	9547	6687	70.04	408	4.27	52.82	54.86	57.56	310	3.25	1609	16.85	10	0.10	1.13	1989	20.83	8.28	19428
Wagga-2*	9482	6553	69.11	404	4.26	52.80	54.84	57.53	306	3.23	1580	16.66	10	0.11	1.35	2032	21.43	9.23	16874
Wagga-3*	9521	6652	69.87	407	4.27	52.82	54.86	57.57	310	3.26	1599	16.79	10	0.11	1.19	2211	23.22	8.06	21123
E. australis FL pathotype:																			

BRIP 52616a*	9215	6638	72.03	283	3.07	52.82	54.89	57.58	312	3.39	1601	17.37	11	0.12	0.43	1070	11.61	2.48	87842

***E. australis* SO pathotype:**

Arg-1	9253	6616	71.50	333	3.60	52.78	54.82	57.5	321	3.47	1331	14.38	11	0.12	2.03	1368	14.78	0.92	139846

E. fawcettii:

BRIP 53147a*	10080	7314	72.56	408	4.05	52.35	54.16	56.63	404	4.01	1989	19.73	13	0.13	0.37	962	9.54	1.03	95535
BRIP 54245a*	10164	7372	72.53	430	4.23	52.34	54.15	56.62	371	3.65	1750	17.22	13	0.13	0.36	938	9.23	N/A	N/A
BRIP 54425a*	10014	7309	72.99	365	3.64	52.36	54.18	56.65	371	3.70	1684	16.82	13	0.13	0.32	827	8.26	N/A	N/A
BRIP 54434a*	10016	7302	72.90	358	3.57	52.36	54.17	56.65	374	3.73	1680	16.77	13	0.13	0.32	860	8.59	N/A	N/A
DAR 70024	10223	7373	72.12	480	4.70	52.3	54.11	56.59	378	3.70	1806	17.67	13	0.13	0.35	1367	13.37	0.88	162499
SM16-1	10003	7,330	73.27	373	3.73	52.39	54.22	56.69	377	3.77	1669	16.68	12	0.12	0.23	1083	10.83	N/A	N/A

E. ampelina:

YL-1	9804	7309	74.55	278	2.84	53.03	54.88	57.6	316	3.22	1816	18.52	13	0.13	1.59	3670	37.43	17.9	7860

Utilised in known effector analysis:

Necrotrophic:

Botrytis cinerea	11481	4350	37.89	2162	18.83	44.72	46.22	48.08	590	5.14	1361	11.85	12	0.10	2.76	4454	38.79	N/A	N/A

Parastagon-ospora nodorum	15878	11009	69.33	480	3.02	51.9	54.28	56.86	724	4.56	5130	32.31	16	0.10	0.64	3107	19.57	6.64	52953
Pyrenophora tritici-repentis	10772	1902	17.66	3786	35.15	51.71	53.55	55.5	322	2.99	1489	13.82	12	0.11	1.63	2181	20.25	N/A	N/A
Sclerotinia sclerotiorum	13770	6560	47.64	1625	11.80	43.93	45.7	47.6	557	4.05	4647	33.75	10	0.07	3.49	10610	77.05	N/A	N/A
Zymoseptoria tritici	11936	7541	63.18	787	6.59	54.35	55.7	57.27	415	3.48	3583	30.02	11	0.09	2.46	4387	36.75	17.3	711
Hemi-biotrophic:																			
Leptosphaeria maculans	12337	8664	70.23	459	3.72	51.72	54.15	56.71	442	3.58	3746	30.36	14	0.11	10.7	5065	41.06	37	7635
Pyricularia oryzae	12236	5221	42.67	1676	13.70	54.47	57.41	61.1	684	5.59	3617	29.56	24	0.20	9.07	6105	49.89	N/A	N/A
Rhynchospo-rium commune	12100	4475	36.98	2409	19.91	48.54	49.91	51.5	396	3.27	2986	24.68	12	0.10	6.26	4700	38.84	29.5	9009
Verticillium dahliae	10441	4878	46.72	1245	11.92	56.38	59.43	63.03	329	3.15	1850	17.72	21	0.20	1.49	1893	18.13	1.47	85600
Biotrophic:																			
Ustilago maydis	6692	4133	61.76	330	4.93	54.36	55.87	57.45	157	2.35	1974	29.50	3	0.04	1	1406	21.01	N/A	N/A

*Isolates sequenced in the current study

**Results shown for isolates of *Elsinoe* compared individually to all 10 comparative species.

APPENDIX E: ADDITIONAL DATA FOR THE ANALYSIS OF FEATURES
ASSOCIATED WITH GENOME PLASTICITY

Table 1. Level of fragmentation of genome assemblies of the *Elsinoë* spp. (Figure 22).

Species/isolate	Total number of scaffolds	Scaffolds < 20 Kb in length	
		Number	%
E. australis **JBS pathotype:**			
Hillston-1*	1378	1038	75.33%
Hillston-2*	484	372	76.86%
Hillston-3*	603	495	82.09%
Forbes-1*	637	511	80.22%
Forbes-2*	495	383	77.37%
Wagga-1*	465	359	77.20%
Wagga-2*	466	334	71.67%
Wagga-3*	438	337	76.94%
E. australis **FL pathotype:**			
BRIP 52616a*	61	26	42.62%
E. australis **SO pathotype:**			
Arg-1	21	2	9.52%
E. fawcettii:			
BRIP 53147a*	286	211	73.78%
BRIP 54245a*	475	396	83.37%
BRIP 54425a*	474	395	83.33%
BRIP 54434a*	473	394	83.30%
DAR 70024	53	9	16.98%
SM16-1	1266	1208	95.42%
E. ampelina:			
YL-1	13	0	0.00%

*Isolates sequenced in the current study

Table 2. Proportions of candidate virulence genes in gene-sparse locations of the *Elsinoë* spp. (Figure 23). P values are shown for statistically significant overrepresentation of candidate virulence genes in gene-sparse locations.

	E. australis JBS pathotype Hillston-2*	*E. australis* FL pathotype BRIP 52616a*	*E. australis* SO pathotype Arg-1	*E. fawcettii* BRIP 53147a*	*E. ampelina* YL-1
Total genes	9533	9215	9253	10080	9804
Total genes in gene-sparse locations	408	283	333	408	278
% of total proteins that are in gene-sparse locations	4.28%	3.07%	3.60%	4.05%	2.84%
Genes involved in predicted Secondary Metabolite (SM) clusters	145	201	200	233	219
SM gene clusters in gene-sparse locations	9	5	8	3	4
% of SM genes that are in gene-sparse locations	6.21%	2.49%	4.00%	1.29%	1.83%
Secreted proteins (SP)	1095	1057	1091	1280	1167
SP in gene-sparse locations	51	41	44	61	37
% of SP that are in gene-sparse locations	4.66%	3.88%	4.03%	4.77%	3.17%
Candidate effectors (CE)	236	219	235	276	270

CE in gene-sparse locations	16	7	9	23	15
% of CE that are in gene-sparse locations	6.78% (p=0.04630)	3.20%	3.83%	8.33% (p=0.00082)	5.56% (p=0.00997)
Cell wall degrading enzymes (CWDE)	196	205	190	212	190
CWDE in gene-sparse locations	7	8	4	6	8
% of CWDE that are in gene-sparse locations	3.57%	3.90%	2.11%	2.83%	4.21%
Elsinoë-specific genes***	1615	1601	1331	1894	1816
Elsinoë-specific genes in gene-sparse locations	152	87	86	177	67
% of *Elsinoë*-specific genes in gene-sparse locations	9.41% (p=6.01449e-24)	5.43% (p=1.57438e-08)	6.46% (p=1.64322e-08)	9.35% (p=1.33508e-31)	3.69%

*Sequenced in the current study

**Predicted secreted CAZymes

***Unique among ten comparative species and no orthoMCL ID obtained

Table 3. Proportions of candidate virulence genes of the *Elsinoë* spp. that are within 20 Kb of a TE region (Figure 24). P values are shown for statistically significant overrepresentation of genes within 20 Kb of a TE region among the candidate virulence genes.

	E. australis JBS pathotype Hillston-2*	*E. australis* FL pathotype BRIP 52616a*	*E. australis* SO pathotype Arg-1	*E. fawcettii* BRIP 53147a*	*E. ampelina* YL-1
Total genes	9533	9215	9253	10080	9804
Total genes within 20 Kb of a TE region	2174	1070	1368	962	3670
% of total proteins within 20 Kb of a TE region	22.80%	11.61%	14.78%	9.54%	37.43%
Genes involved in predicted Secondary Metabolite (SM) clusters	145	201	200	233	219
SM genes within 20 Kb of a TE region	28	32	39	0	79
% of SM genes within 20 Kb of a TE region	19.31%	15.92% (p=0.03887)	19.50% (p=0.03960)	0.00%	36.07%
Secreted proteins (SP)	1095	1057	1091	1280	1167
SP within 20 Kb of a TE region	270	150	173	137	444
% of SP within 20 Kb of a TE region	24.66%	14.19% (p=0.00377)	15.86%	10.70%	38.05%
Candidate effectors (CE)	236	219	235	276	270

CE within 20 Kb of a TE region	61	27	45	30	121
% of CE within 20 Kb of a TE region	25.85%	12.33%	19.15% (p=0.03796)	10.87%	44.81% (p=0.00702)
Cell wall degrading enzymes** (CWDE)	196	205	190	212	190
CWDE within 20 Kb of a TE region	60	29	32	19	76
% of CWDE within 20 Kb of a TE region	30.61% (p=0.00663)	14.15%	16.84%	8.96%	40.00%
Elsinoë-specific genes***	1615	1601	1331	1894	1816
Elsinoë-specific genes within 20 Kb of a TE region	444	192	227	241	736
% of *Elsinoë*-specific genes within 20 Kb of a TE region	27.49% (p=7.42819e-07)	11.99%	17.05%	12.72% (p=2.56171e-07)	40.53% (p=0.00143)

*Sequenced in the current study

**Predicted secreted CAZymes

***Unique among ten comparative species and no orthoMCL ID obtained

Table 4. Proportions of candidate virulence genes of the *Elsinoë* spp. which contain an SSR and/or a PolyAA (Figure 25). P values are shown for statistically significant overrepresentation of genes containing an SSR and/or a PolyAA repeat among the candidate virulence genes.

	E. australis JBS pathotype Hillston-2*	*E. australis* FL pathotype BRIP 52616a*	*E. australis* SO pathotype Arg-1	*E. fawcettii* BRIP 53147a*	*E. ampelina* YL-1
Total genes	9533	9215	9253	10080	9804
Total genes containing an SSR and/or a PolyAA	1165	1126	1137	1105	1270
% of total proteins containing an SSR and/or a PolyAA	12.22%	12.22%	12.29%	10.96%	12.95%
Genes involved in predicted Secondary Metabolite (SM) clusters	145	201	200	233	219
SM genes containing an SSR and/or a PolyAA	10	23	22	11	19
% of SM genes containing an SSR and/or a PolyAA	6.90%	11.44%	11.00%	4.72%	8.68%
Secreted proteins (SP)	1095	1057	1091	1280	1167
SP containing an SSR and/or a PolyAA	137	137	132	149	163
% of SP containing an SSR and/or a PolyAA	12.51%	12.96%	12.10%	11.64%	13.97%

Candidate effectors (CE)	236	219	235	276	270
CE containing an SSR and/or a PolyAA	12	16	13	11	11
% of CE containing an SSR and/or a PolyAA	5.08%	7.31%	5.53%	3.99%	4.07%
Cell wall degrading enzymes (CWDE)	196	205	190	212	190
CWDE containing an SSR and/or a PolyAA	13	19	16	17	17
% of CWDE containing an SSR and/or a PolyAA	6.63%	9.27%	8.42%	8.02%	8.95%
Elsinoë*-specific genes	1615	1601	1331	1894	1816
***Elsinoë*-specific genes containing an SSR and/or a PolyAA**	250	300	211	281	248
% of *Elsinoë*-specific genes containing an SSR and/or a PolyAA	15.48% (p=1.13710e-05)	18.74% (p=4.92263e-17)	15.85% (p=1.93040e-05)	14.84% (p=4.38589e-09)	13.66%

*Sequenced in the current study

**Predicted secreted CAZymes

***Unique among ten comparative species and no orthoMCL ID obtained

Table 5. Proportions of candidate virulence genes of the *Elsinoë* spp. which are within 20 Kb of an AT-rich region (Figure 26). P values are shown for statistically significant overrepresentation of genes within 20 Kb of an AT-rich region among the candidate virulence genes.

	E. australis JBS pathotype Hillston-2*	*E. australis* FL pathotype BRIP 52616a*	*E. australis* SO pathotype Arg-1	*E. fawcettii* BRIP 53147a*	*E. ampelina* YL-1
Total genes	9533	9215	9253	10080	9804
Total genes within 20 Kb of an AT-rich region	2216	654	326	348	4699
% of total proteins within 20 Kb of an AT-rich region	23.25%	7.10%	3.52%	3.45%	47.93%
Genes involved in predicted Secondary Metabolite (SM) clusters	145	201	200	233	219
SM genes within 20 Kb of an AT-rich region	26	42	12	0	92
% of SM genes within 20 Kb of an AT-rich region	17.93%	20.90% (p=1.28387e-10)	6.00%	0.00%	42.01%
Secreted proteins (SP)	1095	1057	1091	1280	1167
SP within 20 Kb of an AT-rich region	309	114	55	67	582
% of SP within 20 Kb of an AT-rich region	28.22% (p=2.79901e-05)	10.79% (p=1.77967e-06)	5.04% (p=0.00358)	5.23% (p=0.000266)	49.87%
Candidate effectors (CE)	236	219	235	276	270
CE within 20 Kb of an AT-rich region	63	29	16	15	150

% of CE within 20 Kb of an AT-rich region	26.69%	13.24% (*p*=0.00079)	6.81% (*p*=0.00887)	5.43%	55.56% (*p*=0.00654)
Cell wall degrading enzymes (CWDE)	196	205	190	212	190
CWDE within 20 Kb of an AT-rich region	71	28	13	14	114
% of CWDE within 20 Kb of an AT-rich region	36.22% (*p*=2.38834e-05)	13.66% (*p*=0.00059)	6.84% (*p*=0.01678)	6.60% (*p*=0.01518)	60.00% (*p*=0.00049)
Elsinoë-* specific genes	1615	1601	1331	1894	1816
***Elsinoë-* specific genes within 20 Kb of an AT-rich region**	463	97	47	64	862
% of *Elsinoë-* specific genes within 20 Kb of an AT-rich region	28.67% (*p*=1.64244e-08)	6.06%	3.53%	3.38%	47.47%

*Sequenced in the current study

**Predicted secreted CAZymes

***Unique among ten comparative species and no orthoMCL ID obtained

Table 6. Proportions of candidate virulence genes of the *Elsinoë* spp. on scaffolds < 20 Kb in length (Figure 27). P values are shown for statistically significant overrepresentation of genes found on scaffolds < 20 Kb in length among the candidate virulence genes.

	E. australis JBS pathotype Hillston-2*	*E. australis* FL pathotype BRIP 52616a*	*E. australis* SO pathotype Arg-1	*E. fawcettii* BRIP 53147a*	*E. ampelina* YL-1
Total genes	9533	9215	9253	10080	9804
Total genes on scaffolds < 20 Kb in length	67	8	7	130	0
% of total proteins on scaffolds < 20 Kb in length	0.70%	0.09%	0.08%	1.29%	0.00%
Genes involved in predicted Secondary Metabolite (SM) clusters	145	201	200	233	219
SM genes on scaffolds < 20 Kb in length	0	0	0	0	0
% of SM genes on scaffolds < 20 Kb in length	0.00%	0.00%	0.00%	0.00%	0.00%
Secreted proteins (SP)	1095	1057	1091	1280	1167
SP on scaffolds < 20 Kb in length	6	3	1	16	0
% of SP on scaffolds < 20 Kb in length	0.55%	0.28%	0.09%	1.25%	0.00%
Candidate effectors (CE)	236	219	235	276	270
CE on scaffolds < 20 Kb in length	1	0	0	7	0
% of CE on scaffolds < 20 Kb in length	0.42%	0.00%	0.00%	2.54%	0.00%
Cell wall degrading	196	205	190	212	190

enzymes (CWDE)					
CWDE on scaffolds < 20 Kb in length	1	2	0	0	0
% of CWDE on scaffolds < 20 Kb in length	0.51%	0.98% (p=0.01262)	0.00%	0.00%	0.00%
Elsinoë*-specific genes	1615	1601	1331	1894	1816
***Elsinoë*-specific genes on scaffolds < 20 Kb in length**	26	2	3	63	0
% of *Elsinoë*-specific genes on scaffolds < 20 Kb in length	1.61% (p=1.59040e-05)	0.12%	0.23%	3.33% (p=1.14938e-14)	0.00%

*Sequenced in the current study

**Predicted secreted CAZymes

***Unique among ten comparative species and no orthoMCL ID obtained

Table 7. Proportions of candidate virulence genes of the *Elsinoë* spp. located on the edge of a scaffold (Figure 28). P values are shown for statistically significant overrepresentation of genes found on the edge of a scaffold among the candidate virulence genes.

	E. australis JBS pathotype Hillston-2*	*E. australis* FL pathotype BRIP 52616a*	*E. australis* SO pathotype Arg-1	*E. fawcettii* BRIP 53147a*	*E. ampelina* YL-1
Total genes	9533	9215	9253	10080	9804
Total genes on the edge of a scaffold	253	73	40	252	26
% of total proteins on the edge of a scaffold	2.65%	0.79%	0.43%	2.50%	0.27%
Genes involved in predicted Secondary Metabolite (SM) clusters	145	201	200	233	219
SM genes on the edge of a scaffold	4	1	0	0	0
% of SM genes on the edge of a scaffold	2.76%	0.50%	0.00%	0.00%	0.00%
Secreted proteins (SP)	1095	1057	1091	1280	1167
SP on the edge of a scaffold	39	17	7	36	7
% of SP on the edge of a scaffold	3.56% (p=0.03338)	1.61% (p=0.003112)	0.64%	2.81%	0.60% (p=0.02852)
Candidate effectors (CE)	236	219	235	276	270
CE on the edge of a scaffold	14	3	1	14	4
% of CE on the edge of a scaffold	5.93% (p=0.00410)	1.37%	0.43%	5.07% (p=0.00946)	1.48% (p=0.00521)
Cell wall degrading enzymes** (CWDE)	196	205	190	212	190
CWDE on the edge of a scaffold	8	2	0	0	2
% of CWDE on the edge of a scaffold	4.08%	0.98%	0.00%	0.00%	1.05%

Elsinoë-specific genes***	1615	1601	1331	1894	1816
Elsinoë-specific genes on the edge of a scaffold	77	14	13	109	8
% of *Elsinoë*-specific genes on the edge of a scaffold	4.77% (*p*=5.80495e-08)	0.87%	0.98% (*p*=5.80495e-08)	5.76% (*p*=1.25758e-19)	0.44%

*Sequenced in the current study

**Predicted secreted CAZymes

***Unique among ten comparative species and no orthoMCL ID obtained

PLOS ONE

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

204

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

205

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
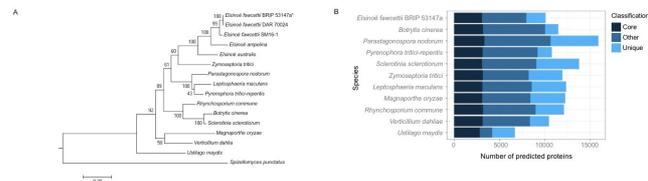
Fig 1. Species comparison. (A) Maximum likelihood phylogenetic tree of *E. fawcettii* BRIP 53147a with recently sequenced *Elsinoë* isolates and species included in the comparative study. The phylogenetic tree was inferred from a concatenated dataset including partial TEF1α and RPB2 regions. *Spizellomyces 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
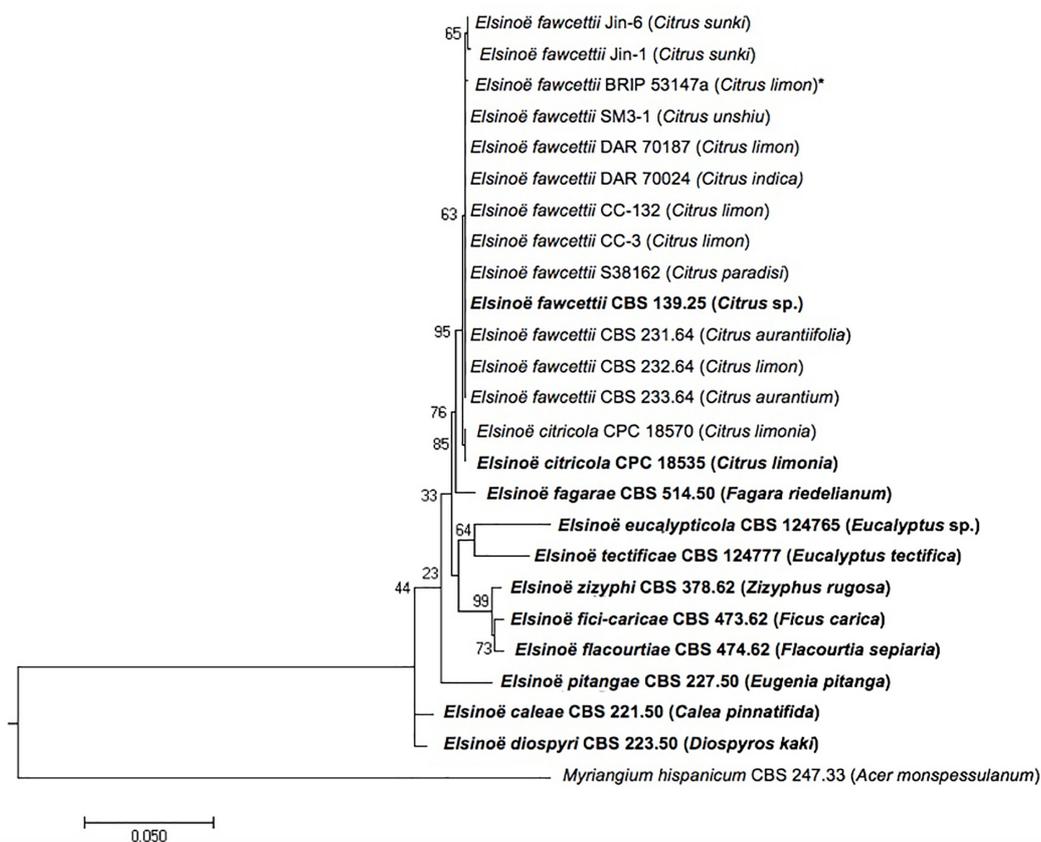
Fig 2. Maximum likelihood phylogenetic tree of *E. fawcettii* isolates and closely related species. The phylogenetic tree was inferred from a concatenated dataset including ITS and partial TEF1-α regions. *Myriangium hispanicum* was used as the outgroup. The branch length indicates the number of nucleotide 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

209

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)

210

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

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%

212

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

213

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].

214

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

https://doi.org/10.1371/journal.pone.0227396.t003

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

215
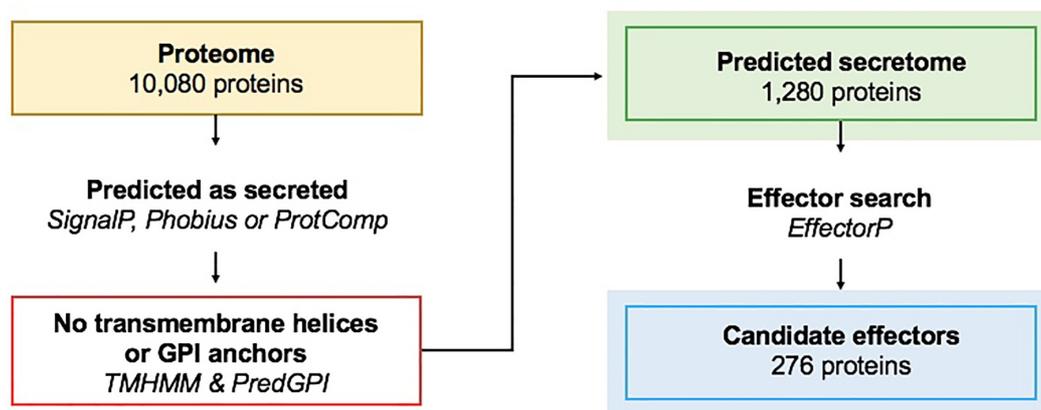
Fig 3. Pipeline for the discovery of the predicted secretome and candidate effectors. The secretome search started with the predicted proteins of a species, proteins were predicted as secreted using at least one of three tools, proteins with predicted transmembrane helices or GPI-anchors were removed. 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*)

217

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

https://doi.org/10.1371/journal.pone.0227396.t004

(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).

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

219
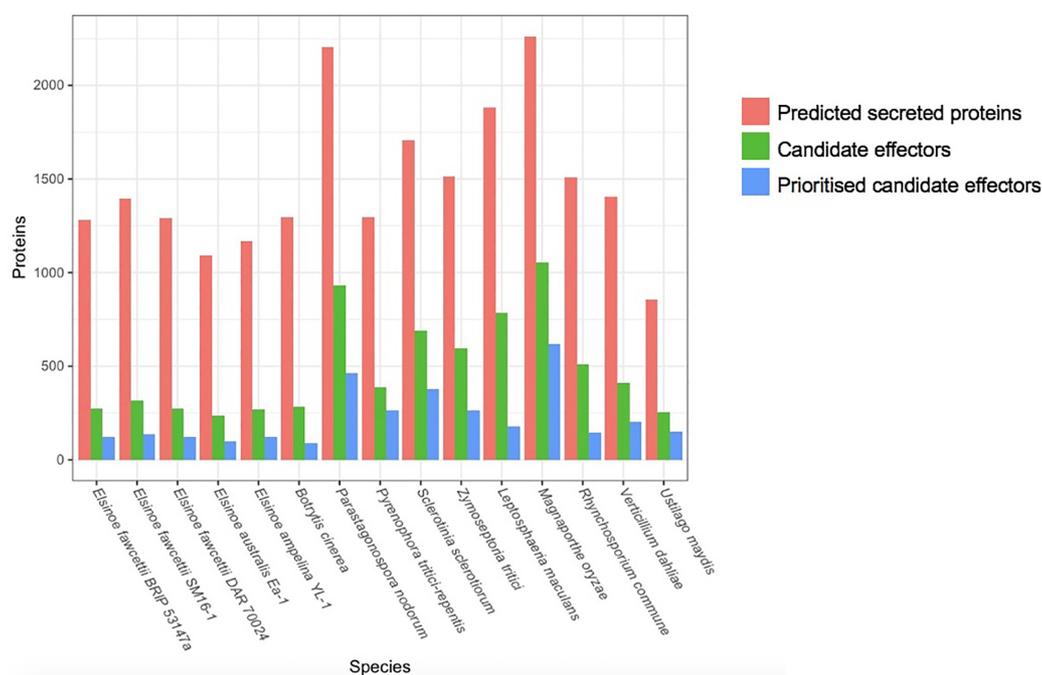
Fig 5. Comparison of numbers of secreted proteins, candidate effectors and prioritised candidate effectors among 15 fungal pathogens. Secreted proteins and candidate effectors were predicted using the pipeline in Fig 3. Prioritised candidate effectors were determined using features shown in Fig 4.

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%

220

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 *pgx*6 (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

https://doi.org/10.1371/journal.pone.0227396.t006

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,

222

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*)

223

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)			

https://doi.org/10.1371/journal.pone.0227396.t007

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%
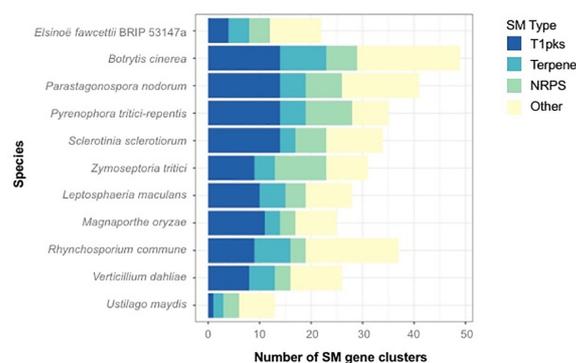
Fig 6. Comparison of numbers of predicted secondary metabolite gene clusters among 11 fungal species. Numbers of SM gene clusters, shown on the x axis, are divided into SM types; (I) Type I Polyketide synthase (T1PKS); (II) terpene; (III) non-ribosomal peptide synthetase (NRPS); and (IV) other, which 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.

https://doi.org/10.1371/journal.pone.0227396.g006

224

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

225

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.

227

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.

228

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 Phyto-pathol. 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 pathoge-nicity. 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–necro-trophic 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 Sus-ceptibility. 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 geno-mics 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 chal-lenges 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 *Stagonos-pora 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 spe-cies. 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 immu-nity. 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 diversifica-tion 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

229

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

230

65. Martinez JP, Ottum SA, Ali S, Francl LJ, Ciuffetti LM. Characterization of the ToxB gene from *Pyrenophora 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-pathogenic 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 evidence for the involvement of a specific polygalacturonase, P2c, in the invasion and spread of *Aspergillus 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 endopolygalacturonases 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*. Proceedings 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 *Penicillium 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 Interactions. 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 Biology. 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 Laboratory: 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. Bioinformatics. 2014; 30(15):2114–20. https://doi.org/10.1093/bioinformatics/btu170 PMID: 24695404

231

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–D85. 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

232

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

233

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).

234

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

235