Fungal Transcriptomics of Symbiotic Incompatibility of the endophyte *Epichloë bromicola* with Wheat

Tailong (Ben) Zhang

2019

School of Fundamental Sciences

Massey University

Supervisors:

Dr Linda Johnson, AgResearch Limited

Dr Christine Voisey, AgResearch Limited

Prof Rosie Bradshaw, Massey University

Dr Carl Mesarich, Massey University

Abstract

The fungal endophyte *Epichloë bromicola* forms mutualistic symbioses with grasses of the *Elymus dahuricus* complex. Artificial inoculation studies have shown that *E. bromicola* strains AR3060 and AR3018 are also able to colonise the wheat cultivar "Monad" (*Triticum aestivum*). Unlike the compatible and symptomless associations they form with their native hosts, colonisation of wheat by AR3060 causes host dwarfing but the endophyte is still able to undergo its life-cycle and be vertically transmitted via seed. AR3018 infection is less compatible and causes severe stunting in wheat, followed by plant death and/or the formation of endophyte-free tillers. Little is known about the molecular events associated with compatible and incompatible endophyte-host interactions.

In this study, transcriptome analysis of compatible versus incompatible interactions between AR3060 and AR3018 with their native and non-native hosts, showed that many genes predicted to encode effector-like proteins were highly up-regulated during interactions with host plants, and metal ion-binding and transmembrane transporter-related genes were significantly enriched in a gene ontology analysis. One of the most abundant transcripts *in planta* encodes a LysM-domain containing protein that has high identity to *Cladosporium fulvum* effector Ecp6 based on its secondary and tertiary structure.

Our results indicate that *Epichloë* endophytes may manipulate grass physiology via the production of effector proteins to produce an environment favourable for fungal infection. Through bioinformatics approaches, 27 *E. bromicola* candidate effectors with potential to regulate compatibility between *Epichloë* fungi and their hosts have been identified. Transient expression of effectors in non-host plants, combined with targeted gene overexpression/deletion studies, are now being conducted to explore their roles in regulating defence responses and symbiotic interactions between *Epichloë* and its novel wheat host.

Long term, this study aims to facilitate fully compatible interactions between *E. bromicola* and wheat to enable these fungi to confer biotic and abiotic stress protection to commercial wheat.

Acknowledgement

Firstly, I would like to thank my supervisors, Dr Linda Johnson, Prof Rosie Bradshaw, Dr Christine Voisey, and Dr Carl Mesarich for their guidance over these years. Especially, thank you Rosie for all the prompt responses; thank you Tina for the incredible writing suggestions; thank you Carl for your immense knowledge in effectors; and thank you, Linda, for everything.

Thanks to the lovely people at AgResearch, past and present. To Tash, Anouck, Wayne, Mike and Suli, who love to share their knowledge and skills. To Paul for the bioinformatics training and discussions. To Stuart, for being a good friend.

To my wife Iris and all of my family, here and gone, thank you for your love, understanding and support.

My special thanks to all the kind people that listened, cared and helped me while I was in my darkest days. Thank you, for saving my life.

Table of Contents

---

1. Introduction

1.1 *Epichloë* endophytes

Microorganisms play an essential role in almost every environment. They live freely or establish symbiotic relationships with other organisms that range from pathogenic to mutualistic. A microbe that lives internally and asymptomatically within a host plant for at least part of its life cycle is termed an endophyte (Wilson 1995). *Epichloë* species (Clavicipitaceae, Hypocreales, Ascomycota) are fungal endophytes that form systemic symbioses with cool-season grasses in the subfamily Pooideae, which includes many economically important grasses, e.g., *Festuca* spp. and *Lolium* spp. (Schardl et al. 2004, Spatafora et al. 2007, Leuchtmann et al. 2014). *Epichloë* species are predicted to be associated with 30% of cool-season grasses (Leuchtmann 1992). Given that grass-dominated habitats cover approximately 40% of the Earth's land surface, *Epichloë* species are considered to have substantial ecosystem value (Strömberg 2011, Saikkonen et al. 2016).

1.1.1 Growth and life cycle *in planta*

Systemic colonisation of grasses by *Epichloë* endophytes is closely correlated with host development. Endophyte hyphae invade the developing shoot apex of the seedling and ramify between host cells in the shoot apical meristem. The hyphae then invade leaf primordia, and from there are able to infect leaf blades and sheaths as they are developing, resulting in systemic colonization of aerial plant tissues (Clay 1990, Christensen et al. 2002). *Epichloë* endophytes remain in the intercellular spaces of the host and do not invade host cells. Fungal growth in meristematic tissues is characterized by the production of hyphal branches and apical extension. However, in the apoplast of growing tissues hyphae are long and straight with few branches. The endophytic hyphae within the growing culm, leaves and inflorescences elongate, together with the expanding plant cells, through intercalary extension, which is thought to be induced by mechanical stretching imposed on hyphae that are physically attached to expanding plant cells (Christensen et al. 2008, Voisey 2010).

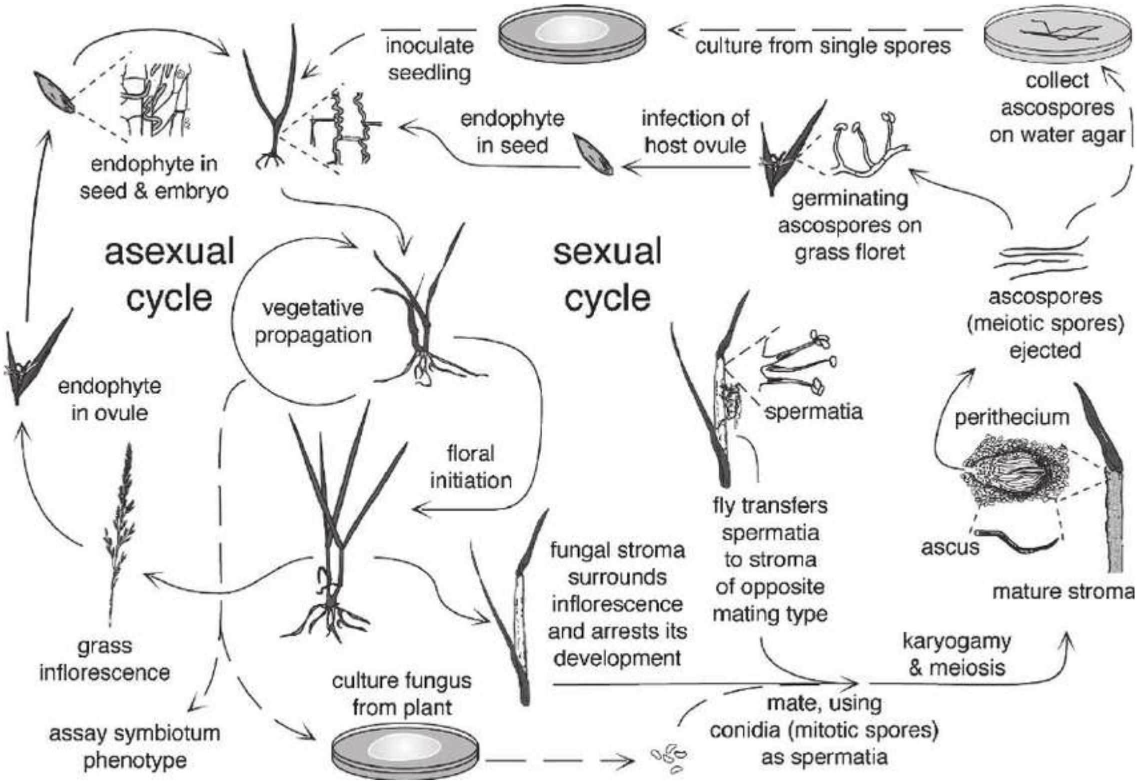


Figure 1. Asexual (left) and sexual (right) life cycles of *Epichloë festucae* (Schardl et al. 2009). Dashed arrows show experimental manipulations including isolation and inoculation.

Epichloë endophytes can be transmitted vertically and/or horizontally between plants (Figure 1). Asexual *Epichloë* species infect nearly all aerial plant tissues with no visible symptoms and transmit efficiently through the infection of host seeds (White 1993). In contrast, sexual *Epichloë* species are also able to transmit horizontally. This is an antagonistic symbiosis which causes disease symptoms known as ‘choke’ that occurs on the flowering tillers. On the affected tiller, the reproductive structures (receptive hyphae and sterigma) of *Epichloë* form on a stroma, suppressing floral development and seed production of the host (Kirby 1961, Chung and Schardl 1997). Conidia are produced on stromata and are transferred to stromata on other plants through a symbiotic fly vector (Bultman and White 1988). If the conidia are of opposite mating types, ascospores are formed and are ejected from the mature asci to initiate new plant infections. Ironically, many sexual *Epichloë* species can transmit vertically through seeds as well, however it is often prevented by the production of fungal stomata which surround and arrest host inflorescence (Moon et al. 2002, Tintjer et al. 2008, Schardl 2010).

1.1.2 Benefits of the symbiosis and its application in agriculture

The agricultural and ecological importance of *Epichloë* endophytes is well accepted by the scientific community. One of the most well-known indications of the agricultural significance of *Epichloë* endophytes in New Zealand is the necessity for most pastures to be endophyte-symbiotic for resistance against insects such as black beetle and Argentine stem weevil (Johnson et al. 2013a).

Endophyte presence in forage grasses can also be associated with mammalian toxicity caused by the toxic alkaloids they produce, e.g. ergot and lolitrem alkaloids (Clay and Schardl 2002). These mammalian-toxic strains are typically referred to as common-toxic or wild-type. *Epichloë* endophytes are therefore considered to be protectors of their grass host by providing protection against insects and mammalian herbivory through the production of alkaloid compounds (Clay et al. 1985, Bacon et al. 1986, Bush et al. 1997, Schardl et al. 2007). Some *Epichloë* fungi are also known to impart other significant protection against biotic and abiotic stresses to their hosts, such as disease resistance (Gwinn and Gavin 1992), drought tolerance (Arachevaleta et al. 1989, Elmi and West 1995, Kannadan and Rudgers 2008), and increased competitive ability and improved survival (West et al. 1993, Malinowski and Belesky 2000).

Large screening efforts have been undertaken to screen for *Epichloë* endophytes that are less toxic to livestock while maintaining their advantageous aspects (Johnson et al. 2013a). This has resulted in the development of several successful commercial grass seed products containing animal-safe non-toxic endophytes such as *E. festucae* var. *lolii* (formerly *Neotyphodium lolii*) strain AR1 in perennial ryegrass (Fletcher 1999) and *E. coenophiala* (formerly *N. coenophialum*) strain AR584 (MaxQ®) in the tall fescue variety "Jesup" (Bouton et al. 2002). The current economic value of *Epichloë*-grass associations in agriculture and turf grass industry is widely recognized in the USA and New Zealand (Hoveland 1993, Gundel et al. 2013). Novel endophyte strains are estimated to contribute approximately $200 million per annum to the New Zealand economy (Johnson et al. 2013a).

Recent advances in endophyte biology and discovery empowers us to utilize the fitness benefits that endophytic *Epichloë* species provide to their hosts in other species of agricultural significance (Gundel et al. 2013, Johnson et al. 2013a). Heat treatments or fungicides can be used to eliminate endophytes from host seeds, and artificial infection of endophyte-free plants (E-) with pure isolates from other endophyte-symbiotic plants (E+) can improve the biotic and abiotic resistance of new plant hosts (Leuchtmann and Clay 1993). In the meantime, recent achievements in bioinformatics and genetics have generated a huge knowledge of the functions of genes that may have biological or ecological significance. These technological tools and scientific knowledge have the potential to widen the possibilities of using selected protective

Epichloë endophytes in other crop plants such as cereals. To date, despite endophytes being isolated from some cultivars of wheat, no common endophyte has been found in modern wheat (Larran et al. 2016, Comby et al. 2017). Inoculation of *Epichloë* endophytes into wheat has been achieved by the Plant-Microbe Interactions team at AgResearch. However, the relationship between the endophyte and the novel wheat host is wholly or partly incompatible, resulting in stunted plants or elimination of the endophyte (Johnson et al., unpublished). Therefore, an understanding of the molecular mechanisms of compatibility underpinning the *Epichloë*-host interaction is required to fully utilize the benefits of this protective endophyte beyond forage grasses.

1.1.3 Host specificity of *Epichloë* from *Elymus* hosts

Associations formed between *Epichloë* and grasses appear to be host-specific. For instance, *E. festucae* var. *lolii* (formerly *Neotyphodium lolii*) specifically forms a symbiosis with perennial ryegrass (*Lolium perenne*) while *E. coenophiala* (formerly *N. coenophialum*) colonises tall fescue (*Lolium arundinaceum*). However, certain *Epichloë* endophytes derived from *Elymus* grasses were shown to be able to form synthetic symbiosis with non-host plants through artificial inoculation (Simpson et al. 2014).

Elymus L. is the largest genus of grasses in the wheat tribe, i.e., Poaceae: Hordeeae (Figure 2), which is considered to be the key gene pool in terms of genetic improvement of cereal crops (Lu 1994). It encompasses approximately 150 perennial species, many widespread across much of Eurasia, North America, and South America (Helfgott and Mason-Gamer 2004). *Elymus* species contain various combinations of five basic genomes that are donated by different genera (Dewey 1971, Jensen 1990, Torabinejad and Mueller 1993). The genomes are categorised as St (*Pseudoroegneria*), H (*Hordeum*), P (*Agropyron*), W (*Australopyrum*) and Y (unknown donor). All *Elymus* species have at least one St genome, which is therefore considered to be a fundamental genome (Dewey 1967).

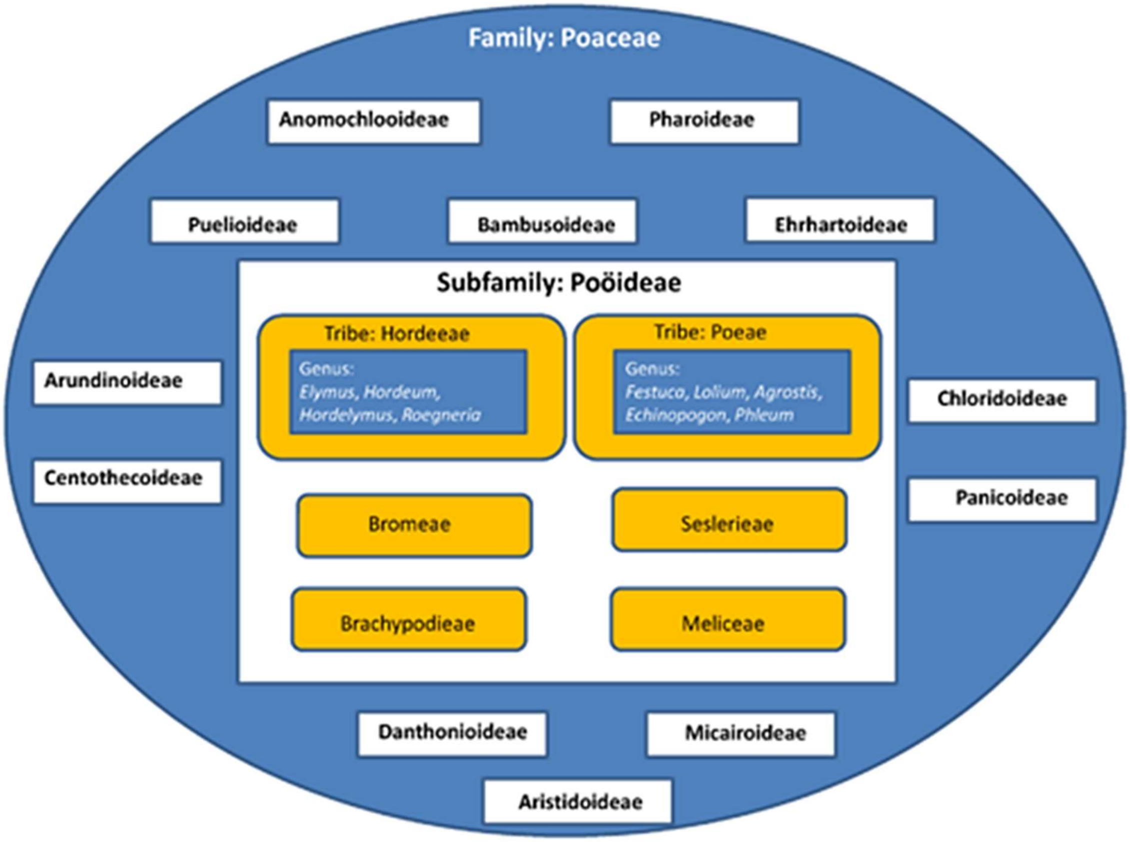


Figure 2. Subfamilies and genera within the true grass family Poaceae (Simpson et al. 2014). Subfamilies within the family Poaceae are shown in white boxes. Yellow boxes denote genera within the subfamily Poöideae. Blue boxes show several representative genera in the tribes Hordeeae and Poeae.

Hybridization and polyploidization are the major driving forces in the diversity and evolution of the genus *Elymus*. According to phylogenetic analysis, many *Elymus* species have multiple origins and also experienced recurrent hybridization between species with difference genomes (Liu et al. 2006). A recent report indicated that *Pseudoroegneria* has no endophytes; however, many *Epichloë* endophytes, and *E.*

bromicola specifically, are found in *Hordeum* and *Elymus* species from China (Card et al. 2014, Song and Nan 2015). Considering the diverse sub-genome combinations among *Elymus* species, it has been suggested that *Epichloë* endophytes may have spread between grass hosts during the process of plant hybridization, for instance, from *Hordeum* into *Elymus* (Song et al. 2015). In addition, molecular phylogenetic analyses of endophytes suggest that host jumps between different species and genera of Pooideae, including *Elymus* species, are common events (Iannone et al. 2011, Oberhofer and Leuchtmann 2012).

As shown in Figure 3, Simpson et al. (2014) have demonstrated that *E. bromicola* strain AR3018 can infect the 'Rahu' cultivar of rye (*Secale cereale*), and *E. bromicola* strain AR3060 is able to form a partly compatible association with the "Monad" cultivar of wheat (*Triticum aestivum*). *E. bromicola* strains AR3060 and AR3018 were sourced from their native plant hosts, *E. dahuricus* and *E. dahuricus* subspecies *excelsus*, respectively. Both *Elymus* hosts are from the *E. dahuricus* complex. The *E. dahuricus* complex contains six taxonomic entities: *E. dahuricus*, *E. excelsus* (*E. dahuricus* ssp. *excelsus*), *E. tangutorum*, *E. cylindricus* (*E. dahuricus* var. *cylindricus*), *E. purpuraristitatus* and *E. villifer* (Lu 1993), which have been recognized as separate species or as infraspecific taxa at various times due to subtle morphological differences. To date, *E. bromicola* AR3060 and AR3018 are the two endophyte strains (of a total of three, from a collection of over 100 strains) able to cross the species barrier and infect wheat. They therefore represent a great opportunity to better understand the mechanism associated with host-endophyte compatibility.

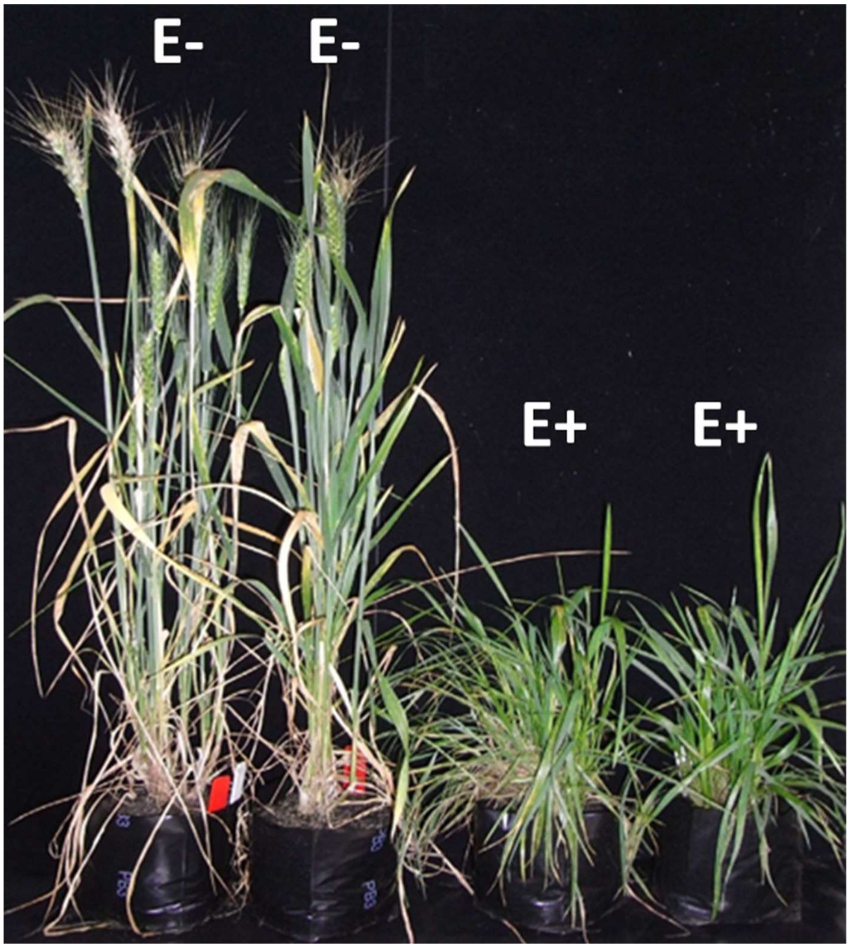


Figure 3. *E. bromicola* AR3060-symbiotic "Monad" wheat (Simpson et al. 2014). "Monad" wheat (*Triticum aestivum*) at six weeks. E-: endophyte-free. E+: *E. bromicola* AR3060-symbiotic.

1.2 Hexaploid bread wheat

Wheat is one of the major staple food sources in the world. It covers the largest agricultural land area among all food crops and global demand is increasing with world population growth and dietary change (Shewry and Hey 2015). However, wheat production in agriculture suffers from many diseases caused by fungi, bacteria and viruses. Among those, fungal pathogen incited diseases cost 15%–20% wheat yield losses per year (Figueroa et al. 2018). Major diseases that threaten wheat production include wheat rust caused by *Puccinia* species, blotch diseases caused by *Zymoseptoria tritici*, *Parastagonospora nodorum* and *Pyrenophora tritici-repentis*, and wheat blast caused by *Magnaporthe oryzae Triticum* (Castroagudín et al. 2016; Cruz and Valent 2017). Large amounts of chemicals are used as pesticides in crop lands every year for disease management. However, wheat diseases are caused by numerous different pathogen species and little is known about the molecular basis of each pathogen-wheat interaction, which can lead to rampant chemical usage and increased pesticide tolerance in some pathogen groups.

The overall complexity of wheat diseases needs to be appreciated. Besides that, there are other challenges that pose continuous threats to wheat yield gains as well, including various biotic and abiotic stresses such as insect damage and climate change (Figueroa et al. 2018). Therefore, protection for wheat needs to be considered in a more holistic perspective, in order to overcome multiple challenges in a way that is both comprehensive and environmentally friendly.

One key difficulty for understanding molecular processes underpinning pathogenicity and breeding in wheat is the wheat genome itself. Similar to *Elymus* grasses, wheat genomes contain copies of sets of chromosomes from each of its progenitors (El Baidouri et al. 2017). Less than half a million years ago, polyploidization occurred between *Triticum urartu* (genome: AA) and an *Aegilops speltoides* species (genome: BB), forming the first polyploid wheat *Triticum turgidum* ssp. *diccocoides* (AABB) genome (Appendix 1). Then there was another polyploidization event approximately 8,000 years ago, between *Triticum turgidum* ssp. *durum* (AABB) and *Aegilops tauschii* (DD), creating the common hexaploid bread wheat *Triticum aestivum* (AABBDD) which is currently the most widely cultivated wheat across the world (Huang et al. 2002, El Baidouri et al. 2017). Because of the six copies (AABBDD) of seven chromosomes (7*6=42 chromosomes total), wheat has a genome that is more than five times as large as the human genome. Years have been spent by researchers in a worldwide collaboration to obtain a high-quality sequenced wheat reference genome. Recently, the International Wheat Genome Sequencing Consortium (IWGSC) reported the achievement of a high-quality, fully-annotated modern wheat (*Triticum aestivum* L., AABBDD) genome containing detailed structural information for all chromosomes, which may greatly facilitate wheat genetics-related studies (Appels et al. 2018).

To reach better protection against biotic and abiotic stress for wheat, genetic modification efforts such as chromosome deletion, addition, and substitution with chromosomes from other species, have been conducted with the wheat genome. *Leymus racemosus* is a perennial grass, one of the wild relatives of wheat. Chromosomes from *L. racemosus* have been introduced into the wheat genome in several studies, mainly owing to the significant resistance of *L. racemosus* to *Fusarium* head blight, which is one of the most devastating wheat diseases in East Asia (Ban 2002). The chromosome introgression lines designated TACBOW0001-0017 (Tottori Alien Chromosome Bank of Wheat) showed resistance to leaf rust, tolerance to heat stress, early heading and large seeds, which is promising for further breeding and genetic studies (Kishii et al. 2004, Mohammed et al. 2014).

1.3 Plant immunity

1.3.1 Model of pathogenicity

Plants utilize a two-layer immune response that involves a combination of localized plasma membrane receptors and intracellular receptors (Dodds and Rathjen

2010, Asai and Shirasu 2015). Jones and Dangl (2006) have described a zigzag model of the plant immune system to illustrate competitive host-pathogen coevolution (Figure 4).

The initial layer of defence response is triggered by generally invariant molecules exposed by the microbe, referred to as pathogen-associated molecular patterns (PAMPs) or microbe-associated molecular patterns (MAMPs). Fungal cell wall components such as polysaccharides (e.g chitin) are examples of these molecules (Win et al. 2012). PAMPs are recognized through membrane-localized pattern recognition receptors (PRRs), which activate the first tier of defence reactions termed PAMP-triggered immunity (PTI). Most PRRs are receptor-like kinases with an extracellular receptor domain, a transmembrane domain, and a cytoplasmic kinase domain (Stotz et al. 2014).

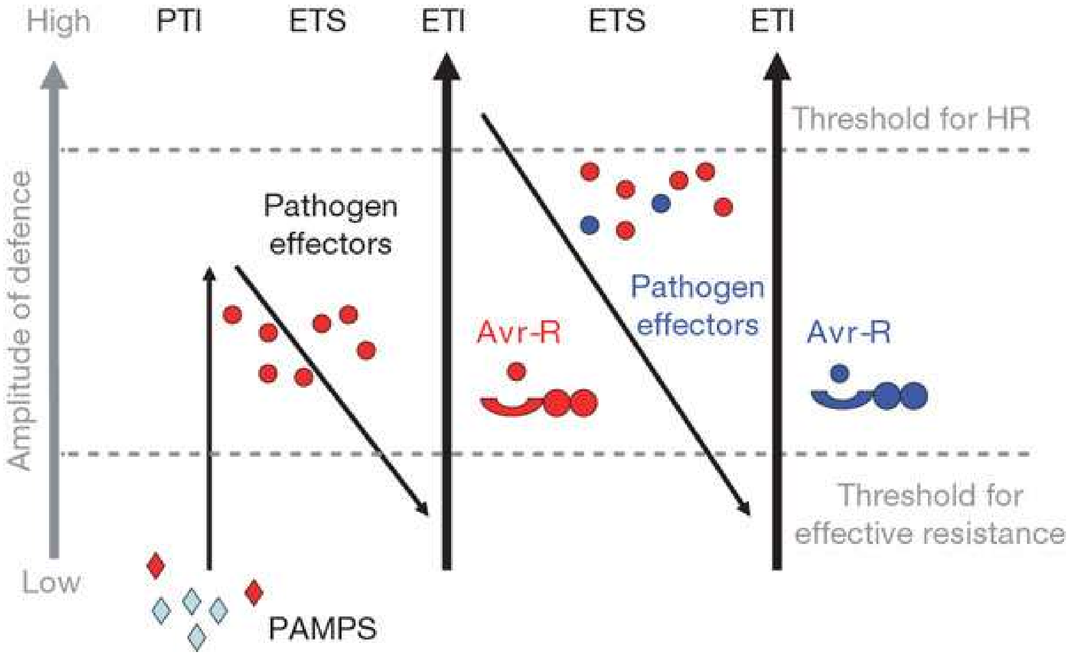


Figure 4. The zigzag model of the plant immune system. (Jones and Dangl 2006) PAMPS, Pathogen-associated molecular pattern; PTI, PAMP-triggered immunity; ETS, effector-triggered susceptibility; ETI, effector-triggered immunity; Avr, avirulence gene; R, resistance gene.

To successfully establish compatible infections fungi must be able to counteract PTI. To do so, adapted pathogens secrete effectors during the infection process that support host colonization, often by suppressing host immune responses (de Jonge et al. 2011, Giraldo and Valent 2013). Effectors can be effective in suppressing PTI, however, they can also be recognized by plant surveillance mechanisms, which induce an additional plant response termed effector-triggered immunity (ETI).

1.3.1 Effector-triggered immunity

Effectors that trigger an ETI response are usually recognized by plant resistance proteins (R proteins), many of which are nucleotide-binding leucine rich repeat (NLR) receptors (Cui et al. 2015). Plant NLRs are conserved intracellular proteins and usually contain a C-terminal leucine rich repeat (LRR) domain and a central NB-ARC (nucleotide-binding adaptor shared by APAF-1, R proteins, and CED-4) domain, therefore are also called NB-LRR, NBS-LRR or NB-ARC-LRR proteins (Qi and Innes 2013). Effector proteins can be recognized by NLR receptors either directly (receptor-binding mediated) or indirectly (accessory protein-involved) (Dodds and Rathjen 2010, Cui et al. 2015).

The activation of ETI often triggers a hypersensitive cell death response (HR) localized at the infection site and usually results in disease resistance. The ETI response is a gene-for-gene interaction, where the effector, coined Avr (for avirulence), is recognized by a host R protein which initiates a host resistance response (Jones and Dangl 2006, Dodds and Rathjen 2010). Effectors and receptors that trigger ETI responses are highly variable and often dispensable compared with the conserved molecules involved in PTI, which may be evidence of the antagonistic coevolution of the ETI components (Dodds and Rathjen 2010). To date, more than 80 effector proteins have been identified and characterized from fungi and oomycetes, 43 of which are encoded by Avr genes (Selin et al. 2016). In many cases, the cognate plant R-proteins associated with the specific Avr have also been identified (Stergiopoulos and de Wit 2009, Gururania et al. 2012, Ali et al. 2014).

1.4 Fungal Effectors

Effectors are proteins or small molecules that are secreted from plant-associated organisms into the host, altering host cell structure and function, thereby stimulating plant infection and/or triggering defence responses (Kamoun 2006, 2007). Effectors are not defined by similar structural properties, but rather distinguished by their function during plant-microbe interactions. Like pathogens, endophytes also develop close associations with host plants. During the process of such symbioses, plant receptors continue to perceive PAMPs/MAMPs. Therefore, endophytes must be able to counteract the immune responses of the plant host, often in a way that is similar to pathogens, such as through the use of effectors (Plett and Martin 2015).

Fungal pathogens utilise apoplastic effectors and/or cytoplasmic effectors to facilitate their infection, depending on their infection strategy. However, effectors can be recognised as new invasion patterns by plant immune receptors in extracellular spaces or inside plant cells (Sonah et al. 2016). In order to evade recognition, genomes of fungal pathogens are plastic, under selection pressure to develop novel effectors and/or modify present effectors by mutagenesis mechanisms (Kane et al. 2015).

1.4.1 Computational prediction of fungal effectors

In comparison to bacterial or oomycete effectors that have conserved domains, such as the type III secretion system or RXLR motif, fungal effectors are more difficult to computationally predict due to the lack of common sequence features.

Although most fungal effectors do not exhibit clear conservation of sequence features and structural folds, some loosely generalised effector properties have been observed within many experimentally discovered effectors. Therefore, those criteria have been widely used for prediction of effectors in various fungal genomes, such as a small size, the presence of a signal peptide, a high percentage of cysteines, differential expression and high *in planta* expression (Sperschneider et al. 2015). Based upon such generic properties, many computational tools are often utilised for prediction of fungal effectors including SignalP, TargetP, TMHMM and Big-PI.

However, most criteria that are normally used for computational prediction of effectors have exceptions. Many apoplastic effectors that are translocated into the intercellular space are small, secreted, cysteine-rich proteins (SSCPs), while cytoplasmic effectors delivered into the plant cell are often larger proteins with few cysteines (Stergiopoulos and de Wit 2009). Multiple disulfide bonds are therefore considered to be important to stabilise the structural folds of proteins in the intercellular space. However, the apoplastic effector AvrLm1 from the hemibiotrophic oilseed rape pathogen *Leptosphaeria maculans* contains only one cysteine (Gout et al. 2006), and conversely the cytoplasmic effectors AvrP4 and AvrP123 from the flax rust pathogen *Melampsora lini* are cysteine-rich and of a small size (Garnica et al. 2014).

The presence of a secretion signal at the N-terminus is considered to be the principal criterion for an effector. Therefore, SignalP, a computational tool that was designed to identify signal peptides, is commonly used as a first step in effector screening. However, this principal criterion is also not of necessity. This has been exemplified by two non-classically secreted effectors PsIsc1 and VdIsc1 from *Phytophthora sojae* and *Verticillium dahliae* (Liu et al. 2006). These two effector proteins are putative isochorismate synthases that are secreted via an unconventional secretion pathway and suppress salicylate-mediated plant immune responses (Liu et al. 2006).

Among all the evidence that can be used to identify putative effectors, an *in planta* induced differential expression pattern was considered to be the only universal feature shared by all identified fungal effectors (Sperschneider et al. 2015). However, there is one reported exception: the effector PIIN_08944 from *Piriformospora indica*. This effector was shown to be highly expressed in advance of direct contact with the host plant and was not differentially regulated during the colonization of *Arabidopsis* roots, indicating its induction is independent of a plant signal (Akum et al. 2015). Effectors like PIIN_08944 may be critical to the initiation of colonization therefore are highly expressed before host contact to neutralise the pre-invasive defence response.

Although fungal effectors are likely to share no conservation, there is an exception in *Fusarium oxysporum formae speciales* (ff. spp.) named ‘SIX’ (Secreted In Xylem) genes. Those genes often have transposable element markers in the upstream region which are highly conserved (Lievens et al. 2009). However, in most cases fungal effector prediction is complex due to absence of evident conservation. Therefore, various computational prediction methods often need to be employed in a combinatorial approach to screen effectors from fungal genomes. Additionally, in order to avoid the risk of excluding genuine effectors that do not meet certain criteria as well as reduce the candidate set, computational predictions may be refined by supplementary experiment data, for instance, *in planta* transcriptomics.

Transcriptome sequencing and analysis is deemed to be one of the most efficient methods for facilitating the identification of effectors. Two polymorphic candidate effectors were identified in *C. fulvum* using computational prediction combined with transcriptome comparison of virulent and avirulent strains. One of the candidate genes, *Avr5*, was later confirmed to encode the avirulence protein recognized by immune receptor Cf5 in tomato (Mesarich et al. 2014).

To date, around 100 fungal effectors have been identified through different approaches (Jones et al. 2018). With an increased number of fungal effectors that have been experimentally confirmed, machine learning techniques are likely to achieve higher accuracy as they keep improving due to the growth in training set data. EffectorP is the first machine learning program designed for computational prediction of fungal effectors (Sperschneider et al. 2016). However, it was trained by a small input of 58 experimentally supported fungal effectors, raising questions about its accuracy in actual use. Recently, EffectorP 2.0 was announced and achieved a higher accuracy with a larger training set of 94 fungal effectors (Sperschneider et al. 2018). Machine learning is also used in LOCALIZER and ApoplastP, for predicting subcellular localization in plant cells and localization to the plant apoplast, respectively, for both fungal proteins and plant proteins (Sperschneider et al. 2017 and 2018).

1.4.2 Effectors in symbiosis

Most reported fungal effectors are from plant pathogens. However, there are effectors or effector-like small secreted proteins (SSPs) that have been found in fungal symbionts that play important roles in mutualistic plant-fungal symbiosis as well.

P. indica is a root-colonizing fungal endophyte that has a broad host spectrum and confers beneficial traits to its host such as enhanced growth and stress resistance. *P. indica* actively suppresses host immunity which is triggered by various MAMPs (Jacobs 2011). Transcriptome analysis showed that around 10% of the genes induced at the early biotrophic stage of symbiosis encode effector-like proteins (Zuccaro et al. 2011). PIIN_08944 was identified as a *P. indica* effector that facilitates early colonisation stages of the mutualistic symbiosis (Akum et al. 2015). Like hemibiotrophic pathogens such as *Magnaporthe oryzae*, host cell death is required for further colonization at the later infection stages of *P. indica*. Considering that hemibiotrophic pathogens often utilize LysM effectors to suppress early host immune responses (de Jonge et al. 2010, Mentlak et al. 2012), the LysM domain containing effector repertoire found in *P. indica* is assumed to act similarly in immune suppression (Zuccaro et al. 2011).

Many gene knockout mutants of *Epichloë* spp. exhibit some degree of incompatibility with their host plants, e.g. Δ*noxA*, Δ*sakA*, Δ*proA* and Δ*sidN* (Tanaka et al. 2006, Eaton et al. 2010, Tanaka et al. 2013, Johnson et al. 2013b). Deletions of *noxA*, *sakA* and *proA* in *E. festucae* showed stunted growth and precocious senescence in host plants, in contrast to the normally asymptomatic phenotype of wild-type-infected plants (Tanaka et al. 2006, Eaton et al. 2010, Tanaka et al. 2013). A transcriptome analysis on the three knockout mutants demonstrated that genes encoding a subset of SSPs are significantly down-regulated compared with wild-type *E. festucae*, indicating that putative effector proteins may play an important role in maintaining compatible symbiosis (Eaton et al. 2015). SOLiD-SAGE in *E. festucae*-infected *Festuca rubra* revealed that 13 of the 20 most abundant *E. festucae* transcripts encode secreted proteins (Ambrose and Belange 2012). Furthermore, six of the 13 secreted proteins are cysteine-rich and are less than 210 amino acids in length, which match the characteristics of many well-known effector proteins (Ambrose and Belange 2012). However, functional characterisation of *Epichloë* effector candidates can be challenging. A recent study (Hassing et al. 2019) showed the growth of both perennial ryegrass *(L. perenne* cv Samson*)* and the *E. festucae* endophyte was not affected by the deletion or overexpression of four effector-like candidates, indicating functional redundancy of effector-like proteins in *Epichloë*. Alternatively, these proteins may be involved in some other biological process not related to the plant-fungal interaction.

The genomes of other fungal symbionts also encode many effector-like SSPs, such as arbuscular mycorrhizal (AM) and ectomycorrhizal (ECM) fungi (Martin et al. 2008, Martin et al. 2010, Plett and Martin 2011, Plett et al. 2011). The genome of the ECM fungus *Laccaria bicolor* encodes 278 cysteine-rich proteins that have a size < 300 amino acids, several of which are specifically induced in the symbiotic tissues. Among all of these mycorrhiza-induced small secreted proteins (MiSSPs), MiSSP7 was shown to be an effector that is necessary for the symbiotic relationship and localizes to the nucleus of poplar host cells to interact with PtJAZ6 protein, resulting in the repression of the jasmonate (JA) pathway (Plett et al. 2011, Plett et al. 2014). Likewise, an effector-like LysM-domain containing protein is highly expressed (>1000 fold) in symbiotic tissues of the ECM fungus *Tuber melanosporum* compared with free-living mycelium (Martin et al. 2010). In the AM fungus *Glomus intraradices*, a repetitive effector SP7 is suggested to down-regulate MAMP-induced plant defence responses by modulating the activity of a host ethylene-responsive transcription factor ERF19 (Kloppholz et al. 2011). Recently, the first CRN effector in symbionts was identified from the AM fungus *Rhizophagus irregularis*, showing a localization to plant nuclear bodies and having a crucial regulatory role for symbiosis development (Voß et al. 2018).

Collectively, these studies indicate that fungal secreted ‘effector’ molecules, often in the form of small secreted proteins, play essential roles in symbiotic associations between the fungus and the plant host, including mutualistic interactions. The molecular basis of establishing a compatible symbiosis between *E. bromicola* and host grasses remains to be elucidated.

1.5 Aim of this study

Wheat production is of outstanding importance in agriculture across the world. Conventional breeding and genetic modification of wheat have been intensively studied to achieve better disease resistance and higher crop yield. In addition, since mutualistic endophytes like *Epichloë* provide significant protection for the plant host against biotic and abiotic stresses, efforts have been made to discover *Epichloë* endophytes that form compatible associations with wheat. To date, despite several fungal and bacterial endophytes having been isolated from some wheat cultivars in certain areas, there is no evidence of a common endophyte-host association in modern wheat (Larran et al. 2016, Comby et al. 2017). However, *Epichloë* endophytes are naturally associated with many wild relatives of wheat, such as *Elymus* grasses, imparting benefits to the host plant (Song et al. 2016).

A recent study indicates that *Epichloë bromicola* strains AR3060 and AR3018 are able to form incompatible associations with the spring wheat cultivar 'Monad'. Colonisation of wheat by AR3060 causes host dwarfing, while infection with AR3018 causes a more severe phenotype comprising stunting, excessive tillering and plant death (Simpson et al. 2014, unpublished data). These strains both originate from the *Elymus dahuricus* complex and form classical compatible symbioses with their native hosts. However, it is unknown what the differences are with the molecular dialog or cross talk that occurs with infections of *Epichloë* with its natural host (*Elymus* sp), compared to with wheat, which gives rise to compatible and incompatible interactions respectively.

The aim of this study is to examine the transcriptomic changes in *Epichloë*-wheat and *Epichloë-Elymus* symbioses, and identify key symbiotic regulators in *E. bromicola*, especially fungal effector-like proteins. This will provide better understanding of the underlying mechanisms involved in regulating compatible or incompatible outcomes between *Epichloë* fungi and their native *Elymus* and novel wheat hosts respectively.

2. Material and Methods

2.1 Biological material

The strains used in this study were *Epichloë bromicola* AR3060 and AR3018, both originally isolated from western China and provided by AgResearch, Palmerston North, New Zealand (Appendix 2). The native host for AR3060 is *Elymus dahuricus* subsp. *excelsus* (*E. excelsus*), whereas the native host for AR3018 is *Elymus dahuricus*. The two endophyte strains produce different characterised alkaloids in their native hosts: AR3060 only produces ergovaline and AR3018 only produces peramine (unpublished data).

The wheat cultivars used in this study were “Monad”, a commercial spring bread wheat (*Triticum aestivum*); and “TACBOW0011” (TAC11), a wheat-*Leymus racemosus* chromosome substitution line in the background of wheat cultivar “Chinese Spring” (*Triticum aestivum*) (Kishii et al. 2004). Wheat and *Elymus* seeds were obtained from the Margot Forde Germplasm Centre, Palmerston North New Zealand.

2.2 Cultivation of organisms

2.2.1 Cultivation of *E. bromicola*

E. bromicola strains were grown for 8-12 days on potato dextrose (PD) agar in a growth chamber under light at 22˚C. For short term storage (<1 month), PD agar plates were wrapped with parafilm and stored at 4˚C. For long term storage, plugs from the edges of fungal colonies were ground, homogenized in PD broth, and stored in 15% (v/v) glycerol at -80˚C.

Potato dextrose medium: 2.4% (w/v) potato dextrose (Difco, France), containing 1.5% (w/v) agar for solid media.

2.2.2 Cultivation of *Elymus* and wheat plants

Mature endophyte-symbiotic Monad wheat and *Elymus* plants used in Experiment I (Chapter 3) were generated by artificial inoculation. Endophyte-free *Elymus* and Monad seeds were germinated on water agar in a growth room at 22˚C. Seedlings were kept in the dark for 5-8 days until their growth was sufficient for inoculation. After inoculation, seedlings were incubated in the dark for a week and in diurnal cycle for another week and then moved to root trainers with potting mix (Daltons Premium). Plants were maintained in ambient conditions in PC2 or PEQ2 glasshouses at AgResearch.

Seedlings of "Monad", "TAC11" and *E. excelsus* used in Experiment II (Chapter 4) were progenies germinated from seeds that were harvested from endophyte-symbiotic parent plants. The seeds were placed 3 cm under the well-watered potting mix in a growth room at 22˚C under constant low red light. Seedlings were then sampled for RNA-seq at the same growth stage, i.e., Zadoks decimal growth stage Z1.1 (Zadoks et al. 1974) with one leaf on main shoot and the second emerging. Plants were maintained in ambient conditions in a PC2 glasshouse at AgResearch.

Water agar (for seed germination): 3% (w/v) agar.

2.3 Plant manipulation

2.3.1 Seed sterilization

Monad and *Elymus* seeds were immersed in 50% (v/v) H_2SO_4 (reagent grade) for 15 min at room temperature and rinsed thoroughly in tap water. Subsequently seeds were soaked in 10% (v/v) Janola Liquid Bleach (containing 3% (w/v) sodium hypochlorite; Janola, New Zealand) and incubated in a laminar flow hood for 15 min. After rinsing thoroughly with sterile water, the seeds were air-dried on filter paper (Whatman, UK) and then stored in plates at 4°C.

2.3.2 Seedling inoculation

Monad and *Elymus* seedlings (5-8 days after germination on water agar) were inoculated with 6-10 day old *E. bromicola* cultures grown on PD agar. Firstly, a 2-3 mm longitudinal cut was made into the shoot apex of the seedling using a sterile scalpel. Then approximately 0.5 mm from the leading edge of a freshly-grown *E. bromicola* colony was placed into the wound.

2.3.3 Immunoblot detection of endophyte

To screen for endophyte infection, wheat and *Elymus* tillers were cut at the base of the pseudostem and printed on a nitrocellulose membrane (BDH, UK). The membrane was then incubated in blocking solution (20 mM Tris, 10 mM HCl, 50 mM NaCl, 0.5% non-fat milk powder, pH adjusted to 7.5) at 25°C with gentle shaking. After two hours, the membrane was moved to another blocking buffer containing 1:1000 dilution of rabbit polyclonal antibody (Sigma-Aldrich, USA) for four hours. The rabbit antibody was raised against endophyte mycelium. Then the membrane was washed three times with fresh blocking buffer and incubated with blocking buffer containing 1:2000 dilution of anti-rabbit alkaline phosphatase-conjugated secondary antibody (Sigma) with gentle shaking. After two hours, the Fast Red chromogen (Sigma) was used to develop the membrane for detection.

2.3.4 Microscopy detection of endophyte

For endophyte detection *in planta*, a 1-2 cm piece of the inner epidermis of the grass leaf sheath was peeled off and placed on a microscopic slide. The sheath tissue was then stained in aniline blue (0.05% aniline blue in lactic acid/glycerol/water 1:2:1) and observed using light microscopy (Olympus BX50) at 400x.

2.4 DNA Isolation

E. bromicola strains were grown for 8-12 days on PD agar. The growing edges of each colony were then harvested using a sterile scalpel, ground and resuspended in 150 µL of lysis buffer (100 mM Tris-HCl, 100 mM EDTA, 1% SDS, PH=8) in 1.5 mL tubes. DNA was extracted using the Fungal/Bacterial DNA MiniPrep kit (ZYMO Research, USA) according to the manufacturer's manual. The quality and the quantity of the genomic DNA was assessed using a NanoDrop 2000 UV-Vis Spectrophotometer.

2.5 RNA isolation

2.5.1 Isolation of RNA from *E. bromicola* in culture samples

Total RNA was isolated from *E. bromicola* mycelia using Trizol reagent (Invitrogen, USA) according to the manufacturer's instructions. Edges of *E. bromicola* colonies in PD agar (12 days old) were removed and placed into a 1.5 ml Eppendorf tube, flash-frozen with liquid nitrogen, then ground into a fine powder using a battery-powered drill with pestle tips which were pre-cooled with liquid nitrogen. 1.2 ml Trizol was added into each tube containing 250 ml frozen powder. The resulting mixture was used for total RNA extraction. In the RNA wash step, columns from an RNA clean and concentrator kit (ZYMO) were used for better yield and quality. Selected RNA samples were stored in -80°C and then shipped to BGI (Beijing Genomics Institute, China) for transcriptome sequencing.

2.5.2 Isolation of RNA from plant samples

Total RNA was isolated from plant tissues using Trizol reagent (Invitrogen) according to the manufacturer's protocol. Tillers of Monad wheat and *Elymus* were cut close to the base, flash-frozen, and ground into a fine powder. In the RNA wash step, columns from an RNA clean and concentrator kit (ZYMO) were used for better yield and quality. A NanoDrop 2000 UV-Vis Spectrophotometer, an Agilent 2100 Bioanalyzer and a standard agarose gel electrophoresis was used to assess the quantity, purity, and integrity of all RNA samples. Selected RNA samples were stored in -80°C and then shipped to BGI.

2.6 Genome annotation and refinement of *E. bromicola* gene models

AR3018 and AR3060 genomes were previously sequenced and provided by AgResearch. Genome annotation was conducted using the MAKER program. MAKER combined the gene prediction tools of FGENESH, SNAP, Genemark-ES and Augustus with transcript (Model 3 from *E. festucae* E2368) and protein (Swiss-prot) information. As part of this project, the annotations of the MAKER-derived genes were updated by searching the corresponding protein predictions against the Swiss-prot, Uniprot and Refseq databases using BLAST, and by protein functional domain analyses, taking only the top hit with an expect value below 1E-5 (a threshold set to find potential domains in the absence of a good hit). The foregoing annotation process was carried out by AgResearch bioinformatician Paul Maclean.

I then improved the quality of the existing AR3060 and AR3018 gene models using the Illumina RNA-seq data to supplement the alignment-based gene predictions. Briefly, reference-based assemblies of AR3060 and AR3018 RNA-seq reads were generated by Cufflinks and compared to the previous Maker predicted gene models to obtain precise exon/intron boundaries and start/stop sites (Table 1). Interproscan, signalP, targetP, TMHMM, and WoLF PSORT were then run on the protein predictions for further annotation and prediction of secretion.

2.7 Transcriptome analysis

2.7.1 Tissue sampling for transcriptome studies

Plant tissue samples used for analysis of *E. bromicola* strains AR3018 and AR3060 in mature wheat (Experiment I, Chapter 3) were collected from the pseudostems of six-week-old wheat and *Elymus* plants infected with AR3018 and AR3060. An axenic in-culture sample of AR3060 (grown in PDA) was also prepared. Five treatments were used for RNA extraction: AR3060 grown in axenic culture (3060), AR3018 in *Elymus dahuricus* (El3018), AR3060 in *Elymus dahuricus subsp. excelsus* (El3060), AR3018 in “Monad” wheat (Mo3018), and AR3060 in “Monad” wheat (Mo3060). Same sets of RNA samples were sent for SOLiD-SAGE and Illumina RNA-seq. Sampling and sequencing of Experiment I were conducted previously by the Plant-Microbe Interaction team, AgResearch Limited, raw sequencing data were supplied at the start of this project. Only one biological replicate was used to represent each treatment in this previous RNA-seq.

To further investigate gene expression changes at early time points in a more natural situation, vertically transmitted endophyte-symbiotic seeds were used in Experiment II, Chapter 4 (transcriptome analysis of progenies of *E. bromicola* strain

AR3060 infected wheat). “TACBOW0011” (TAC11), a wheat line that has recently been discovered to be compatible with AR3060, was added for sampling. Plant tissue samples were the aerial parts of the newly geminated seedlings at Zadoks decimal growth stage Z1.1 (appendix 3). Seedling samples were cut at the shoot apex region and the wound at the bottom of the seedling was used for immunoblot to confirm endophyte presence. An axenic in-culture sample of AR3060 (grown in PDB) was also prepared. For this transcriptome analysis of progenies, four treatments were selected: AR3060 grown in axenic culture (p3060), AR3060 in *E. excelsus* (pEl3060), AR3060 in “Monad” wheat (pMo3060), and AR3060 in “TAC11” wheat (pTAC3060). Three biological replicates were used to represent each treatment.

2.7.2 RNA sequencing

In transcriptome analysis of *E. bromicola* strains AR3018 and AR3060 in mature wheat (Experiment I, Chapter 3), RNA samples were sequenced by New Zealand Genomics Ltd. (http://www.nzgenomics.co.nz/) using two lanes of the Illumina sequencing platform (HiSeq 2500, version 3 chemistry). SOLiD-SAGE was conducted previously by SeqWright (US) on the SOLiD4 platform. Raw sequencing data were provided at the start of this project.

In transcriptome analysis of progenies of *E. bromicola* strain AR3060 infected wheat (Experiment II, Chapter 4), RNA samples with RIN (RNA Integrity Number) values above eight were used to construct separate cDNA libraries with fragment lengths of 130 bp (±25 bp). Then, paired-end sequencing was performed by BGI (Beijing Genomics Institute, China) using the Illumina sequencing platform HiSeq 4000. Paired-end reads with a length of 100 bp were generated. The quality of the reads was analysed using FastQC Version 0.10.1. Before mapping, the low-quality regions (regions with base quality of at least 20 retained), sequencing adapters and primers were trimmed from the reads using Flexbar Version 2.4.

2.7.3 Reference mapping and reference-based assembly

Illumina RNA-seq data were mapped independently onto the annotated genomes of *E. bromicola* strains AR3060 and AR3018 using STAR Version 2.5.0c. To improve the MAKER predicted gene models, a reference-based assembly of RNA-seq reads was generated using Cufflinks Version 2.2.1. The reference-based gene models were then compared to the MAKER predicted ones to assess the improvement in *E. bromicola* genome predictions (generally in the form of extended untranslated regions of genes) using Integrative Genomics Viewer (IGV) Version 2.31.

2.7.4 Identification and clustering of differentially-expressed (DE) genes

The expression level of *E. bromicola* genes was enumerated from the uniquely mapped read pairs from the reverse strand in mapping files using the feature Counts program of the SUBREAD package Version 1.5.0-p3. Differential gene expression analysis of the trimmed mean of M normalised counts was carried out in all comparisons with the R package EdgeR using common dispersion estimated from the deviance of the gene counts. Differentially-expressed (DE) genes were restricted to those with P value <0.05 and fold change >2 in at least one of the following pairwise comparisons: El3060 vs. Mo3060, El3060 vs. 3060, and Mo3060 vs. 3060. Perl scripts were designed to search and gather DE genes of each comparison using certain criteria. For each DE gene, Gene Ontology (GO) annotation was obtained with Blast2GO, Interproscan and KEGG. AriGO and REVIGO were then used to analyse and demonstrate the enrichment of GO terms in DE genes. The hierarchical method in Cluster 3.0 was used for gene expression patterns clustering in Chapter 3

(Experiment I). The Self Organizing Tree Algorithm (SOTA) method in T-MeV4.9 was used for expression pattern clustering in Chapter 4 (Experiment II).

2.8 Computational discovery and selection of candidate effectors

SSPs were filtered from *E. bromicola* genomes using criteria that include the presence of a secretion signal peptide (SP), protein size smaller than 300 amino acids in length, absence of a glycosylphosphatidylinositol (GPI) anchor, and absence of transmembrane helix domains (TMH). SignalP v3.0 and v4.1 were used for signal peptide prediction. Big-PI was used for predicting a possible GPI anchor. TMHMM was used for transmembrane domain prediction. For selecting *E. bromicola* effector candidates, more bioinformatic tools were used and are listed in Table 1.

Table 1. List of bioinformatic resources used in this study

Resources	Feature	Reference / Website
EfM3 gene models	Annotations, transcripts and protein sequences for *E. festucae* E2368 genome	Schardl et al. 2013 csbio-l.csr.uky.edu/ef2011
InterProScan	Provides functional information of protein sequences including family classification, functional domains and important sites	Jones et al. 2014 www.ebi.ac.uk/interpro
Uniprot	Universal Protein Resource, a database for protein sequences and functional annotation data	www.uniprot.org/uniprot
Refseq	Reference Sequences from assembled genomes in GenBank and transcript and protein records	www.ncbi.nlm.nih.gov/refseq
MAKER	Genome annotation pipeline program	www.yandell-lab.org/software/maker.htmL
Cufflinks	Comprehensive software for RNA-seq de novo assembly, normalisation, differential expression calculation	cole-trapnell-lab.github.io/cufflinks
EdgeR	R package for normalisation and differential expression of read counts	Robinson et al. 2010 bioconductor.org/packages/release/bioc/htmL/edgeR.htmL
TM4 MeV	Multiple Experiment Viewer, designed to analyse, visualise, and stratify data from RNA-seq and microarray studies	mev.tm4.org/#/welcome
Cluster 3.0	An open-source gene expression clustering software	de Hoon et al. 2004 bonsai.hgc.jp/~mdehoon/software/cluster/software.htm
Java TreeView	Viewer of clustering results from Cluster 3.0	jtreeview.sourceforge.net
AgriGO	Web tool for Gene Ontology analysis, particularly for agricultural species	Du et al. 2010 bioinfo.cau.edu.cn/agriGO/index.php

Resources	Feature	Reference / Website
REVIGO	Summarise and visualise GO terms list in various forms, e.g. tag clouds, scatterplots, and interactive graphs	Supek et al. 2011 revigo.irb.hr
WEGO	Web Gene Ontology Annotation Ploting to compare, plot and visualise GO analysis results	Ye et al. 2006 wego.genomics.org.cn/cgi-bin/wego/index.pl
SignalP	Secretion signal prediction tool	Petersen et al. 2011 www.cbs.dtu.dk/services/SignalP
TargetP	Subcellular location prediction tool for eukaryotic proteins	Emanuelsson et al. 2000 www.cbs.dtu.dk/services/TargetP
WoLF PSORT	Subcellular location prediction tool	www.genscript.com/wolf-psort.htmL
TMHMM	Transmembrane helix prediction tool	www.cbs.dtu.dk/services/TMHMM
I-TASSER	Protein 3D structure and function predication server	Zhang 2008 zhanglab.ccmb.med.umich.edu/I-TASSER
Big-PI	GPI-anchor prediction web server	Eisenhaber et al. 1998 mendel.imp.ac.at/gpi/gpi_server.htmL
EffectorP	Online sever to predict fungal effector proteins in secretomes using machine learning method	Sperschneider et al. 2015 effectorp.csiro.au
SecretomeP	Protein non-classical secretion prediction tool	Bendtsen et al. 2004 www.cbs.dtu.dk/services/SecretomeP
PSIPRED	Protein secondary structure prediction	bioinf.cs.ucl.ac.uk/psipred
HHpred	Protein secondary structure homology search tool	Söding et al. 2005 toolkit.tuebingen.mpg.de/hhpred
PDB	Protein Data Bank, archive of experimentally-determined structures information of proteins and nucleic acids	www.rcsb.org/pdb/home/home.do
TM-align	Protein structure comparison and scoring tool	Zhang and Skolnick 2005 zhanglab.ccmb.med.umich.edu/TM-align/
MEME	Novel recurring motif discover tool	meme-suite.org/tools/meme

3. Transcriptome experiment I: *E. bromicola* in mature plant host

3.1 Fungal transcriptomics of *E. bromicola*-symbiotic mature wheat

E. bromicola endophytes AR3060 and AR3018 originate from the *E. dahuricus* complex and form compatible mutualistic symbioses with their native hosts. Recent studies show that these strains can colonize the spring wheat cultivar “Monad” (Simpson et al. 2014, unpublished data). However, the novel associations between *E. bromicola* and wheat are wholly or partly incompatible resulting in stunted or dwarfed plants for AR3018 and AR3060 respectively (Simpson et al. 2014). To investigate the molecular events underpinning the incompatibility of *E. bromicola*-wheat symbiosis, transcriptomes of AR3060 in culture plus AR3060 and AR3018 in mature *Elymus* and wheat plants were sequenced using SOLiD-SAGE and Illumina RNA-seq (sampling and sequencing were conducted previously, raw sequencing data were supplied at the start of this project).

3.1.1 Transcriptome sequencing and mapping of reads to genome references

The first step of the transcriptome analysis required mapping of the previous SOLiD-SAGE and RNA-seq data onto existing genome references from the corresponding strains. However, due to the short-read nature of SOLiD–SAGE in the past and some technical difficulties, libraries of SOLiD-SAGE showed a much lower mapping rate than RNA-seq (Appendix 4). Therefore, SOLiD-SAGE was only used as complimentary and validation in this study. In addition, only one replicate was sequenced for this RNA-seq experiment, which is a major limitation of Experiment I. However, this did not hinder transcriptomics being an effective tool to filter for effector-like gene candidates.

Five RNA-seq libraries generated approximately 133 million clean reads per sample after quality filtering (Table 2). Fungal reads were correspondingly mapped onto AR3018 and AR3060 genomes provided by AgResearch. As the *E. bromicola* genomes were previously assembled from Illumina reads, they were more fragmented compared with the wheat reference genome obtained from the International Wheat Genome Sequencing Consortium (http://plants.ensembl.org/*Triticum_aestivum*). Therefore, many rRNA, mitochondrial and other highly conserved sequences from the AR3060 in culture sample were found to map to the wheat genome, comprising 15% of the fungal reads. Due to the lack of an available *Elymus* genome, *Elymus* and wheat reads were both mapped onto the same wheat reference genome. Expectedly, most of AR3060 reads in culture mapped to the AR3060 genome, and less than 3.22% of the *in planta* mapped reads mapped to the genome because of low hyphal biomass *in planta*. The fungal biomass has been estimated to be less than 2% in host plants according to relative quantities of total DNA (Young et al. 2005, Rasmussen et al. 2007).

While 2.57% of AR3018-symbiotic *Elymus* mapped to the AR3018 genome, only 4278 fungal reads from the AR3018-symbiotic "Monad" wheat (Mo3018) transcriptome mapped to the AR3018 genome, comprising less than 0.01% of total mapped reads in that sample. Considering the stunted phenotype of wheat caused by AR3018 infection, such low fungal reads may indicate that the wheat host plant was dying or becoming endophyte free. Therefore, most of the following bioinformatic analyses were based on AR3060 samples only. A transcriptome study or quantitative reverse transcription polymerase chain reaction (qRT-PCR) of AR3018-symbiotic wheat at an earlier stage (<4-week-old) should be conducted in order to fully understand fungal gene expression changes in AR3018 during interactions with the highly incompatible wheat host.

Table 2. Summary of RNA-seq read alignments to reference genomes

Sample[a]	Filtered reads[b]	Mapped reads[c]	%	Endophyte reads[d]	%	Wheat reads[e]	%
Mo3018	126,783,732	71,283,143	56	4,278	0.01	71,278,865	99.99
El3018	115,173,660	32,387,984	28	833,827	2.57	31,554,157	97.43
Mo3060	116,477,686	64,793,004	56	522,165	0.81	64,270,839	99.19
El3060	124,951,778	36,284,132	29	1,168,393	3.22	35,115,739	96.78
3060	179,344,598	174,106,465	97	147,312,351	84.61	26,794,114	15.39

[a] RNA samples of AR3060 grown in axenic culture (3060), AR3018-symbiotic *Elymus dahuricus* (El3018), AR3060-symbiotic *Elymus dahuricus subsp. excelsus* (El3060), AR3018-symbiotic “Monad” wheat (Mo3018), and AR3060-symbiotic “Monad” wheat (Mo3060).

[b] Total number of reads passing quality control.

[c] Reads mapping to genome reference of either AR3018, AR3060, or wheat, and the corresponding percentage (%) of filtered reads.

[d] Reads mapping to AR3018 or AR3060 genome reference, and the corresponding percentage (%) of mapped reads.

[e] Reads mapping to wheat genome reference, and the corresponding percentage (%) of mapped reads.

3.1.2 Expression of *E. bromicola* genes in mature plant hosts vs in culture

Many *E. bromicola* transcripts were extremely abundant *in planta* compared with the in culture sample (Figure 5). The 20 most highly expressed AR3060 transcripts in *Elymus* comprised nearly 50% of the total fungal reads of native host associations. Conversely, the 20 most expressed AR3060 transcripts in culture collectively comprise just 18.7% of the total reads. Among the six most highly expressed endophyte genes in culture, five of them are heat shock proteins and have similar expression levels. Other genes highly expressed in culture encode a clock-controlled protein, elongation factor 1, glucosidases, and cell wall-related enzymes. Unlike the in culture sample, among the top 20 most highly expressed fungal genes *in planta*, only six genes had annotations that could be found in various databases. The most abundant transcript in both *Elymus* and wheat interactions was *gigA*, a gene encoding a cyclic oligopeptide of unknown function and present across many different *Epichloë* species (Johnson et al. 2015). Transcripts corresponding to a serine/threonine protein kinase, a glycoside hydrolase, a LysM domain-containing protein, a subtilisin-like protease, and a glucose-repressible protein were also highly abundant in AR3060-symbiotic host plants. Very high expression of specific genes *in planta* versus in culture is a result of intensive induction of fungal genes after colonisation of host plants. The genes are presumed to regulate processes that are required for fungal growth *in planta*, maintenance of compatibility, or potentially the production of bioactive metabolites (e.g. *gigA*), but this remains to be determined. Details of the 20 most highly expressed genes in culture, *Elymus* and Monad of RNA-seq Experiment I are listed in Appendix 7.

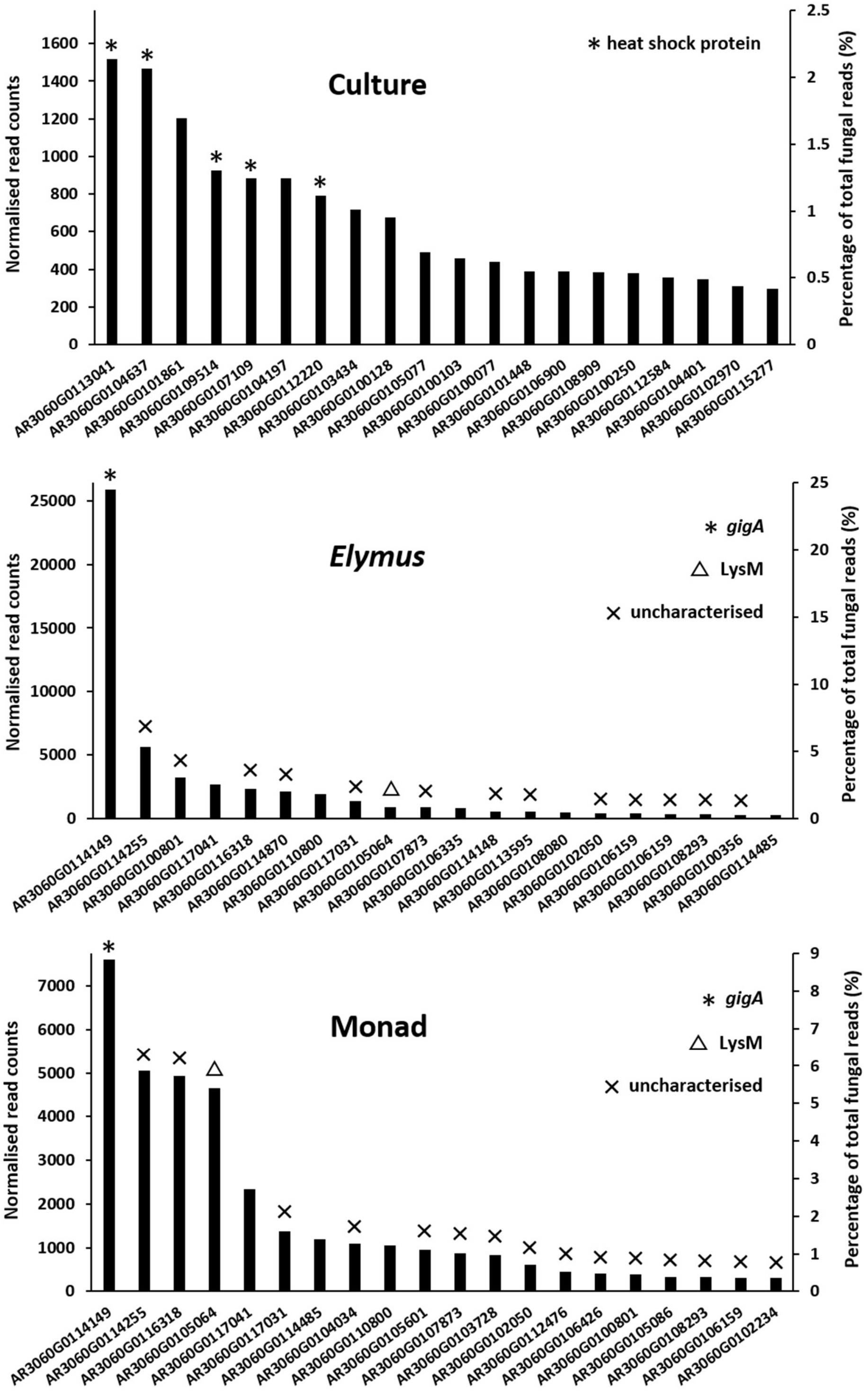


Figure 5. The 20 most highly expressed *E. bromicola* AR3060 genes in mature plant hosts and in culture. Primary vertical axis (left) shows normalised RNA-seq counts of fungal reads as an indication of gene expression level. Secondary vertical axis (right) shows the proportion of the transcript in total reads in each sample. Culture: AR3060 grown in axenic culture. *Elymus*: AR3060-symbiotic *Elymus dahuricus* subsp. *excelsus*. Wheat: AR3060-symbiotic "Monad" wheat.

3.1.3 Fungal genes differentially expressed in mature plant hosts

The next objective was to identify *E. bromicola* genes that are differentially regulated in culture versus *in planta*, and also those that were differentially expressed in compatible (mutualistic) versus incompatible (antagonistic) interactions with different host plants. Based on AR3060 RNA-seq data, 853 genes were found to be differentially expressed in at least one of the following pairwise comparisons: Mo3060 vs. 3060, El3060 vs. 3060, and Mo3060 vs El3060 (Figure 6 and 7). These genes were either highly induced *in planta*, or differentially expressed in the *Elymus* and wheat hosts. Among all differentially expressed genes, nearly 90% (764 out of 853) were *in planta*-induced fungal genes in which approximately 65% (496 out of 764) were not equally represented in the two hosts, suggesting *in planta*-induced expression profiles are host-specific. Only 254 endophyte genes were significantly affected by the change of host, among which 96 and 158 genes were upregulated and downregulated in wheat compared to *Elymus*, respectively (Figure 7). Due to the lack of RNA-seq reads of AR3018 in *Elymus*, SOLiD-SAGE data were used to analyse differential gene expression between AR3018 samples (Supplementary Table 1).

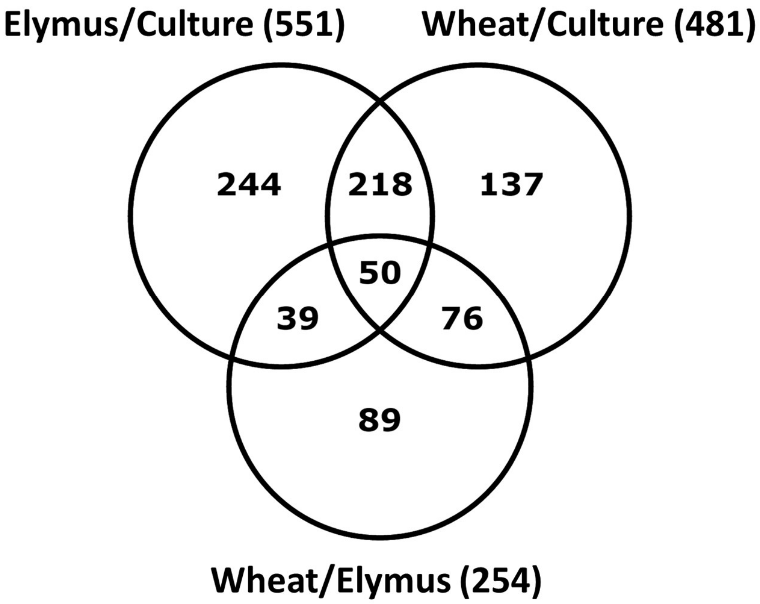


Figure 6. Venn diagram of differentially expressed fungal genes across three comparisons. Differentially expressed genes in AR3060 with P value <0.05 and fold change >2 in at least one of the pairwise comparisons were detected by R package EdgeR using common dispersion estimated from the deviance of the read counts. Culture: AR3060 grown in axenic culture. Elymus: AR3060 symbiotic with *Elymus excelsus*. Wheat: AR3060 with "Monad" wheat.

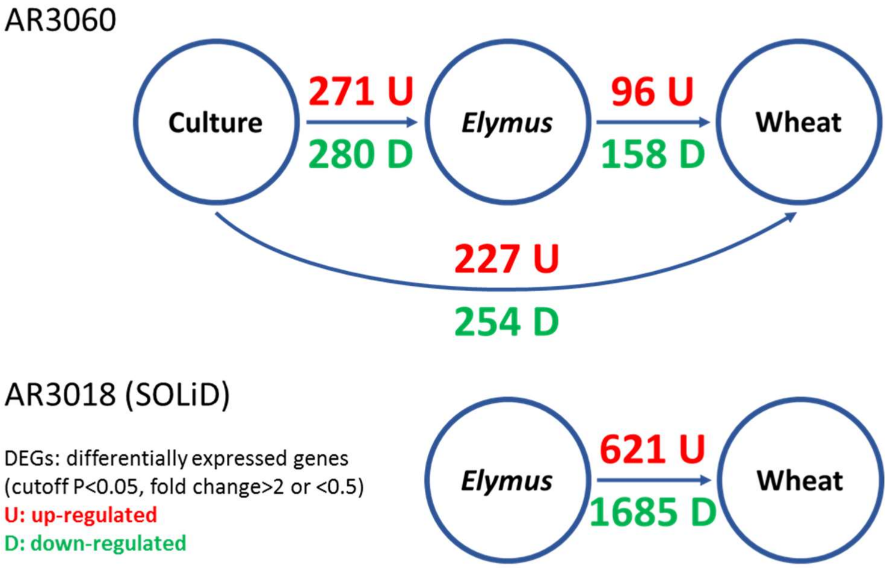


Figure 7. Distribution of differentially expressed fungal genes. Differentially expressed fungal genes across AR3060 and AR3018 samples detected by R package EdgeR using common dispersion estimated from the deviance of the read counts. Culture: AR3060 grown in axenic culture. *Elymus*: AR3060 or AR3018 in native *Elymus* hosts. Wheat: AR3060 or AR3018 in "Monad" wheat. RNA-seq data were used in AR3060 samples and SOLiD-SAGE data in AR3018 samples**.**

The complex *in planta* gene regulations were clearly demonstrated in the hierarchical expression clustering of 853 AR3060 differentially expressed genes (Figure 8). Those genes have been grouped into six clusters based on their distinct expression patterns. Clusters I and V are similar groups of endophyte genes that were expressed in culture and only in one of the two plant hosts. Clusters II and IV showed strong differences in fungal gene expression patterns between *Elymus* and wheat hosts, indicating a substantial change in plant-endophyte interactions. Furthermore, many of the genes in these clusters code for secreted proteins of unknown function. Cluster III includes the common *in planta* induced genes shared by both *Elymus* and wheat, overlapping with the 268 genes shared by El3060/3060 and Mo3060/3060 in the Venn diagram (Figure 6). Cluster VI is a group of endophyte genes that are relatively highly expressed in culture but low *in planta*. Thus, they are supposedly not related to the colonisation of plant hosts and were therefore not included in the following effector mining process.

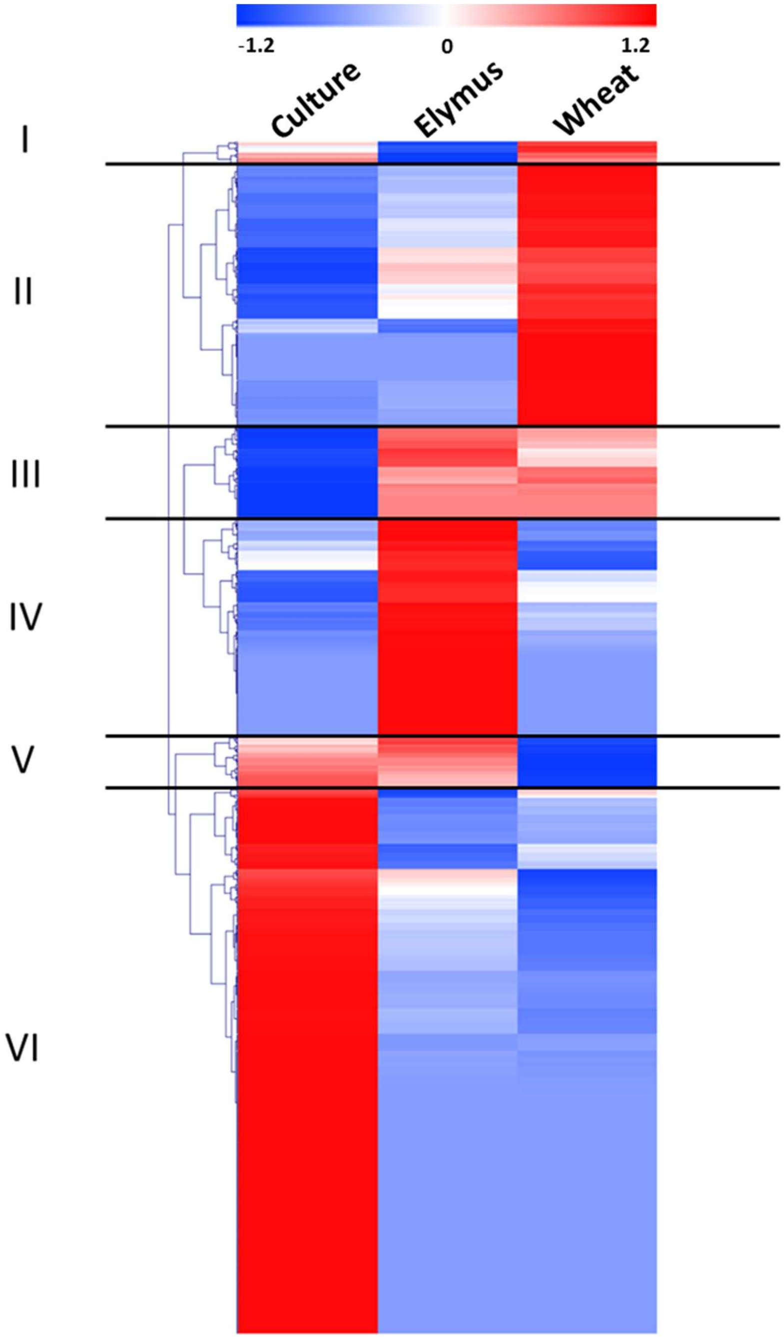


Figure 8. Hierarchical expression clustering of 853 differentially expressed genes in AR3060. Differentially expressed genes in AR3060 with a P value <0.05 and fold change >2 in at least one of three pairwise comparisons were detected by R package EdgeR. Genes have been grouped into six clusters based on their expression patterns. Culture: AR3060 grown in axenic culture. *Elymus*: AR3060 in *Elymus dahuricus* subsp. *excelsus*. Wheat: AR3060 in "Monad" wheat. Gene expression values have been normalized in MeV4.9 using the mean and the standard deviation of the row of the matrix to which the normalised read count value belongs, using formula: Value = [(Value) – Mean(Row)]/[Standard deviation(Row)].

3.1.4 Gene ontology analysis of fungal DE genes in mature plant hosts

Gene ontology (GO) enrichment analysis was conducted for the 254 AR3060 genes differentially expressed between *Elymus* and wheat to better interpret their function (Figure 9). Around 46% (118 out of 254) of the differentially expressed genes could be assigned with GO annotations using Blast2GO. Annotations were then further improved by combining GO annotations obtained via different approaches, e.g. InterProScan, KEGG, and Blast2GO. Approximately 65% (166 out of 254) of differentially expressed genes fell into 20 Molecular Function GO terms after annotation improvement. Six highly similar GO groups were related to transmembrane transporter activity and two were related to ion binding activity. The significant enrichment of such sets of genes in the dataset suggests their importance in maintaining a compatible mutualistic symbiosis.

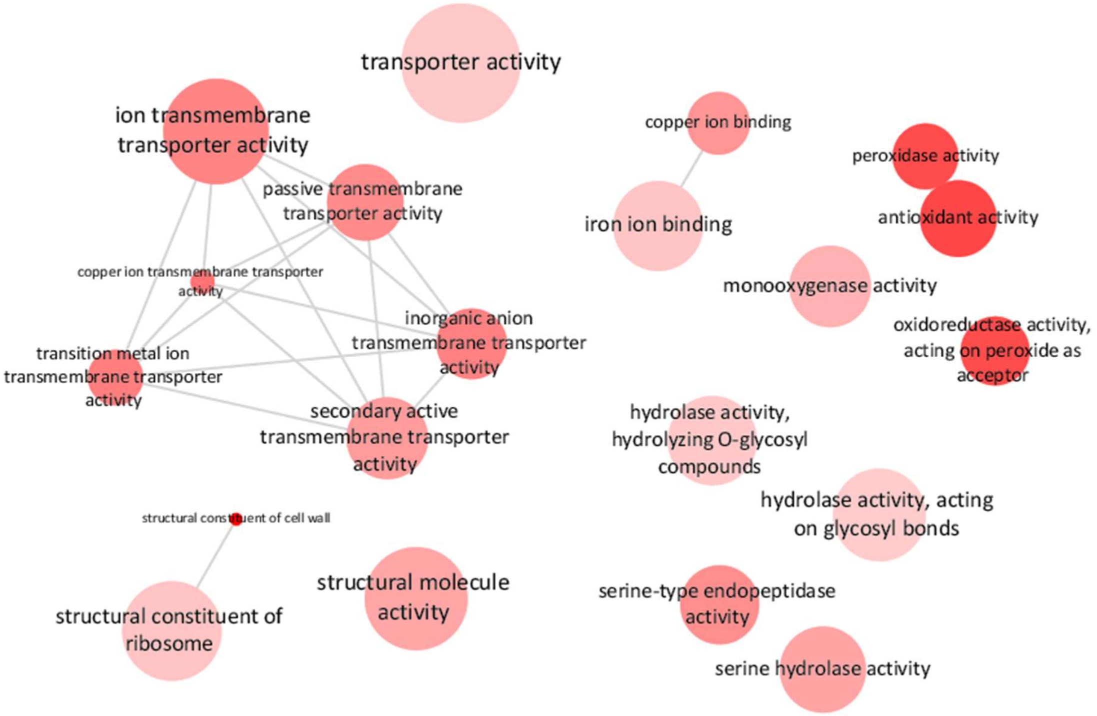


Figure 9. Fungal gene ontology groups differentially expressed *in planta*. REVIGO interactive map of GO terms under Molecular Function category enriched in differentially expressed genes between AR3060-symbiotic wheat and *Elymus*. The sizes of the red circles indicate the relative frequencies of GO terms. Darker shading indicates lower P values. Highly similar GO terms are linked by lines with arbitrary length.

3.2 *E. bromicola* effector identification

To date, only three effectors have been functionally characterised in mutualistic fungi, SP7, MiSSP7, and PIIN_08944 (Plett et al. 2011, Kloppholz et al. 2011, Akum et al. 2015). In transcriptome analysis of *E. bromicola* in *Elymus* and wheat hosts, the expression of a large number of secreted proteins with unknown function were found to be significantly induced *in planta*. Here the aim was to computationally identify and select effector candidates in *E. bromicola* genomes through combined bioinformatics gene prediction and transcriptome analysis approaches.

3.2.1 Effector mining pipeline

Many fungal effectors previously identified from plant pathogenic or symbiotic fungi have been found to be small, secreted, cysteine-rich proteins (SSCPs) with no conserved sequence features (Lo Presti et al. 2015). Therefore, protein size, secretion signals, and percentage of cysteines were often used as the criteria for fungal effector identification. Typically, the criteria for effector discovery are defined as proteins less than 150-200 amino acids (AA), with a percentage of cysteines greater than 3-4 percent (Saunders et al. 2012, Ma et al. 2010). Using these criteria to screen the genomes of AR3018 and AR3060 resulted in the recovery of 255 and 174 SSPs (<200AA, presence of a signal peptide (SP+), no transmembrane helix (TMH-), no GPI anchor (GPI-)) respectively. Among those, 98 and 74 are SSCPs (cysteine >3%). However, only 9 and 8 of the SSCPs are highly expressed *in planta* (normalised counts >100), indicating the limitation of using traditional criteria.

A bioinformatics effector mining pipeline was developed to identify candidate effector genes from the *E. bromicola* genomes using the RNA-seq data (Figure 10). Since AR3060 formed a more compatible association with wheat and generated much more Illumina data than AR3018, only AR3060 was selected for effector mining and for the further Experiment II. Firstly, *E. bromicola* gene predictions were improved by reference-based RNA-seq assembly and the genome annotation was updated (see chapter 2.6). Secondly, arbitrary thresholds of normalised counts were set to select effector candidates that are highly expressed *in planta* or only induced *in planta* based on RNA-seq data. Of all predicted proteins in AR3060 genome, 875 were predicted to have a secretion signal. Among those, 655 were predicted to be secreted out of the cell (not bound to cell membrane or intracellular organelles). Only 94 showed high *in planta* expression above the threshold in AR3060 genome. Finally, considering exceptions in each filter steps, the effector candidates were segregated according to the four following categories:

1) ***In planta*-induced candidate effector:**

   in culture normalised counts <1 (raw counts <42), *in planta* normalised counts >100 (in at least one plant sample), SP+, TMH-, GPI-

2) **Plant-signal independent candidate effector:**

   in culture normalised counts ≥1 (raw counts ≥42), *in planta* normalised counts >100 (in at least one plant sample), SP+, TMH-, GPI-

3) **Non-classically secreted candidate effector:**

   in planta normalised counts >100, SP-, TMH-, GPI-, SecretomeP+

4) **Time-point-specific candidate effector:**

   in planta normalised counts <100, oomycete effector motifs (RXLR and LXLFLAK motif, WY domain)

For candidate effectors that may be secreted via a non-classical secretion pathway, a web-based tool called SecretomeP was used for *ab initio* predictions of unconventional protein secretion (see Table 1). Candidates that have common functional domains or homologs (for instance cytochrome P450, elongation factor) were removed based on hits in various databases (Swiss-prot, Uniprot, Refseq, Pfam). The priority of effector candidates for further deletion/overexpression studies was calculated based on gene expression level, differential expression patterns, and protein domain and secondary structure homology. Since some fungal effectors are not small and cysteine-rich, and not all small secreted cysteine-rich proteins function as effectors, protein size and cysteine count were ignored to avoid bias, but were considered in the priority selection.

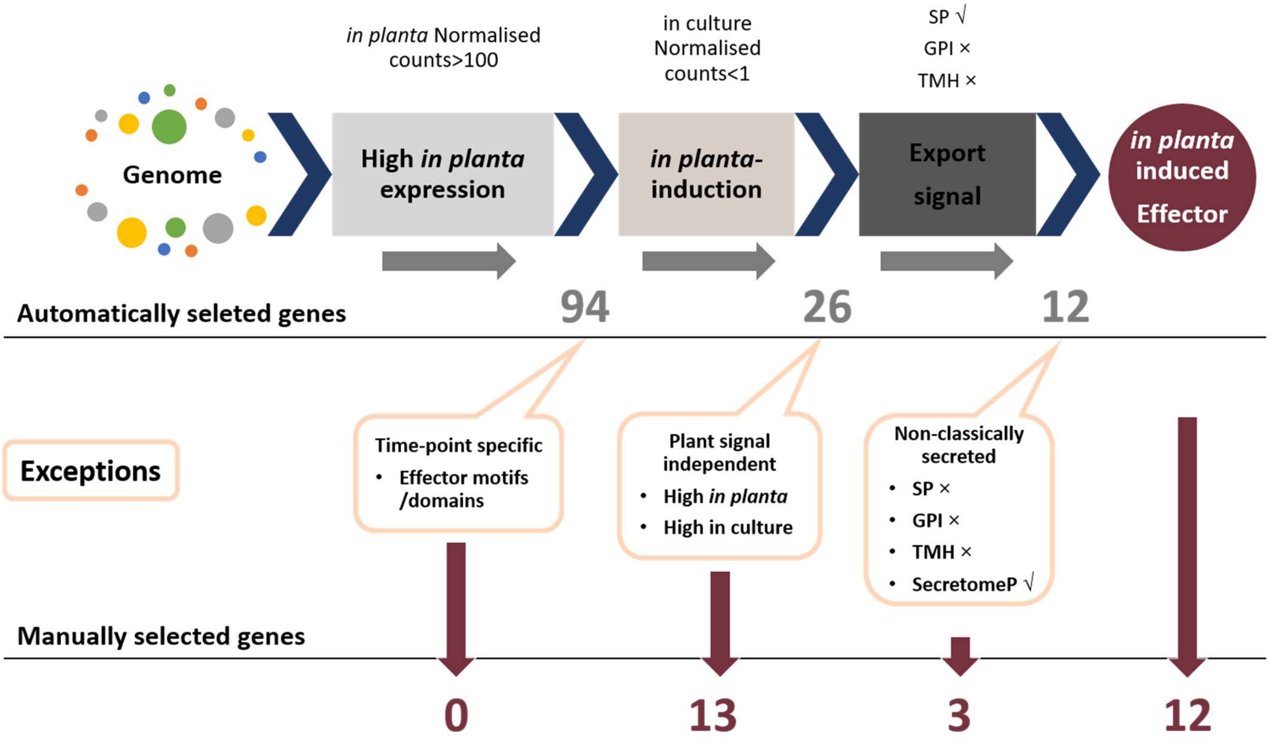


Figure 10. Bioinformatics pipeline designed for the selection of candidate effector genes in *E. bromicola*. The genome of *E. bromicola* strain AR3060 was used for effector mining. Genes were filtered using sequence features, and expression level *in planta* or in culture based on RNA-seq data. Hollow callout (orange outline) indicates exceptions in each step. SP, signal peptide predicted by SignalP. GPI, glycosylphosphatidylinositol anchor predicted by Big-PI. TMH, transmembrane helix predicted by TMHMM. SecretomeP, non-classical secretion predicted by SecretomeP.

In total 28 candidate effector genes from the AR3060 genome were selected according to the aforementioned criteria, including 12 induced *in planta*, 13 that were plant-signal independent, and 3 non-classically secreted putative effectors (Figure 10). One exception is EbCE16 as an *in planta*-induced candidate effector—although its normalised counts was lower than 100 in AR3060-symbiotic plants, it was found to have a homologue in AR3018 that is highly expressed *in planta*. The annotations of all candidate genes, along with their expression profile are shown in Table 3. Most (21) of the candidate effector genes were proteins of uncharacterised function. Annotated candidates include a LysM domain-containing protein, a ribonuclease, a pectinesterase, a subtilisin-like protease, a chitinase, and surprisingly, a killer toxin KP4-like protein. Killer toxin KP4 is a single polypeptide encoded by a symbiotic RNA virus within the cells of corn smut fungus *Ustilago maydis* (Park et al. 1994). The same KP4 annotation can be confirmed in more than one databases using BLAST, and homology search via HHpred also shows a similar secondary protein structure.

To visualise the expression patterns of effector candidates, a hierarchical expression clustering of 137 genes with normalized read counts greater than 100 in AR3060 was carried out (Figure 11). Most of the candidates fell into clusters II, III, and IV except AR3060EbCE16, which has an AR3018 homolog that was highly expressed in *Elymus*.

Table 3. Effector candidates selected from AR3060 genome and transcriptome

	RNA-seq					SOLiD-SAGE								
	Normalised Counts [b]			Monad/Elymus		Normalised Counts		Monad/Elymus						
Gene [a]	Culture	Monad	Elymus	Log2 FC [c]	DE [d]	Monad	Elymus	Log2 FC	DE	Feature [e]	Cys [f]	Length [g]	SP [h]	Location [i]
EbCE1	0	5042	5661	-0.17	N					Uncharacterised	9	110	Y	extr
EbCE2	0	4924	2308	1.09	N	9379	4466	1.52	Y	Uncharacterised, *Epichloe* unique	2	61	Y	extr
EbCE3	0	78	2143	-4.78	Y	51	1882	-4.82	Y	Uncharacterised	6	179	Y	extr
EbCE4	0	4658	905	2.36	Y	735	647	0.80	N	LysM domain protein	18	316	Y	extr
EbCE5	0	865	898	-0.05	N	742	911	0.20	N	Uncharacterised	7	144	Y	cyto
EbCE6	0	48	559	-3.54	Y	19	1867	-6.11	Y	Uncharacterised	8	287	Y	extr
EbCE7	0	612	417	0.55	N	1172	1139	0.25	N	Uncharacterised	6	98	Y	extr
EbCE8	0	41	355	-3.11	Y					Ribonuclease/ribotoxin	4	155	Y	extr
EbCE9	1	840	195	2.11	Y	1517	455	2.34	Y	Uncharacterised	4	235	Y	extr
EbCE10	0	1083	104	3.37	Y	2350	411	2.84	Y	Uncharacterised, *Epichloe* unique	4	104	Y	extr
EbCE11	0	213	24	3.14	Y					Uncharacterised, *Epichloe* unique	10	169	Y	extr
EbCE12	0	187	8	4.62	Y	66	16	2.71	Y	Uncharacterised	4	169	Y	extr
EbCE13	1	225	45	2.32	Y	105	44	1.67	Y	Pectinesterase	4	334	Y	extr
EbCE14	40	10	205	-4.35	Y	25	623	-4.21	Y	Uncharacterised, *Penicillium* homolog	0	271	Y	cyto
EbCE15	2	214	194	0.14	N	63	270	-1.03	Y	Uncharacterised	6	115	Y	extr
EbCE16	0	2	77	-5.20	Y					Uncharacterised, *Epichloe* unique	3	193	Y	extr
EbCE17	2	961	217	2.15	Y	21	33	0.03	N	Uncharacterised/UPF0047 domain	6	270	Y	extr
EbCE18	31	316	410	-0.38	N					Uncharacterised	4	155	Y	extr
EbCE19	5	86	853	-3.31	Y					Subtilisin-like protease	4	351	Y	extr
EbCE20	3	14	162	-3.53	Y					Killer toxin Kp4-like	10	131	Y	extr
EbCE21	2	256	41	2.65	Y	227	42	2.88	Y	Uncharacterised	4	237	Y	extr
EbCE22	5	454	44	3.35	Y	850	96	3.58	Y	Uncharacterised	2	151	Y	extr
EbCE23	17	7596	25865	-1.77	Y	6380	39564	-2.09	Y	Uncharacterised, *Epichloe* unique, nc25	0	70	Y	extr
EbCE24	11	1201	298	2.01	Y					Chitinase	5	454	Y	extr
EbCE25	63	254	223	0.18	N					Hydrolase	6	367	Y	extr
EbCE26	10	332	138	1.27	N	456	255	1.07	Y	Uncharacterised	3	188	N	cyto
EbCE27	18	113	50	1.18	N					Uncharacterised, *Epichloe* unique	1	321	N	cyto
EbCE28	198	105	78	0.43	N	1	1	-0.95	N	Uncharacterised, *Epichloe* unique	1	91	N	cyto

[a] All candidate genes have no transmembrane helix as predicted by TMHMM, no GPI anchor as predicted by Big-PI.

[b] Normalised read counts of candidate genes in RNA-seq Experiment I and previous SOLiD-SAGE. Culture: *Epichloë bromicola* AR3060 grown in axenic culture. *Elymus*: AR3060 symbiotic with native host *Elymus excelsus*. Monad: AR3060 with "Monad" wheat. [c] FC, fold change. [d] DE, differentially expressed (yes or no).

[e] Features are generated based on BLAST hits from databases (Swiss-prot, Uniprot, Refseq, Pfam) and HHpred search results.

[f] Cys, number of cysteines. [g] Length, amino acids (length of the encoded protein product). [h] SP, signal peptide. [i] Location was predicted by WoLF PSORT and TargetP.

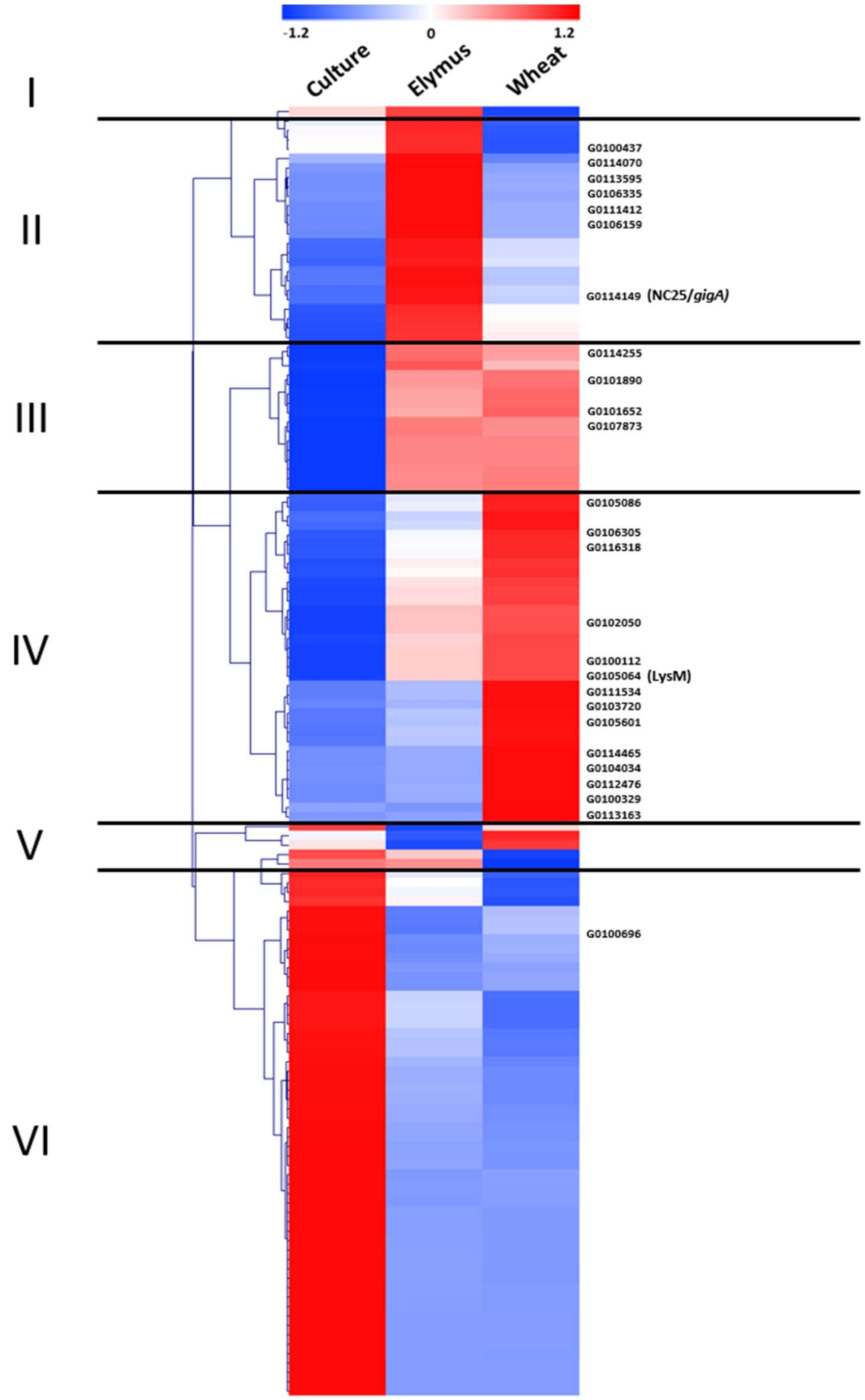


Figure 11. Hierarchical expression clustering of 137 genes in AR3060 with normalized read counts greater than 100 showing expression patterns of effector candidates. Genes have been grouped into six clusters based on their expression patterns. Candidate effectors in this study have been marked on the right. Culture: AR3060 grown in axenic culture. *Elymus*: AR3060 in *Elymus dahuricus* subsp. *excelsus*. Wheat: AR3060 in "Monad" wheat. Gene expression values have been normalized as read counts in MeV4.9 using formula: Value = [(Value) – Mean(Row)]/[Standard deviation(Row)].

3.2.2 Identification of *E. bromicola* putative effectors in other fungi

HHpred is a powerful tool for protein secondary structure prediction and homology searching (Söding et al. 2005). Nucleotide and amino acid sequence homology searches were performed using BLAST, and protein secondary structure homology search using HHpred for every highly expressed putative effector gene in AR3060 genome. A total of 72 putative effector genes were selected by their high *in planta* expression (*in planta* normalised counts >100 in at least one plant sample) and sequence features (SP+, TMH-, GPI-). However, a single putative LysM effector (EbCE4) was the only candidate with similar features to an identified effector protein.

3.2.3 Bioinformatics analysis of a putative LysM effector

The LysM domain-containing candidate effector EbCE4 is small (316AA), secreted, cysteine-rich (~6%), and has two LysM domains. By homology searching in all sequenced *Epichloë* species via the website of Genome Projects at University of Kentucky (Schardl et al., 2013) using blastn, blastp and tblastn, only two EbCE4 homologues, one in *E. bromicola* AL04262, one in *E. festucae* Fl1 were found (Figure 12). AR3018 genome also possesses an EbCE4 homolog with approximately 63% amino acid sequence identity. However, identity between EbCE4 and its homolog in AL04262 and Fl1 are approximately 45% and 39%, respectively.

Using HHpred, the secondary structure of EbCE4 was found to be highly similar to many LysM domain-containing proteins including *C. fulvum* LysM effector Ecp6, the rice chitin receptor OsCEBiP, *Thermus thermophilus* endopeptidase P60_tth, and *Pteris ryukyuensis* chitinase-A (PrChi-A) (Figure 13). Among those Ecp6 was the top hit with a 99.6% probability and a E-value = 1E-14. Chitin is a common PAMP that can trigger plant immunity as it is the major component of most fungal cell walls. The

LysM effector Ecp6 promotes colonization of tomato by the pathogen *C. fulvum* by isolating chitin oligosaccharides released from fungal cell walls to evade detection (de Jonge et al. 2010). In further analysis of the tertiary structure of EbCE4 and Ecp6, it was found that they share an extremely similar structural fold as well (Figure 14), indicating EbCE4 may function similarly as a chitin-binding effector in *Epichloë*-host interactions.

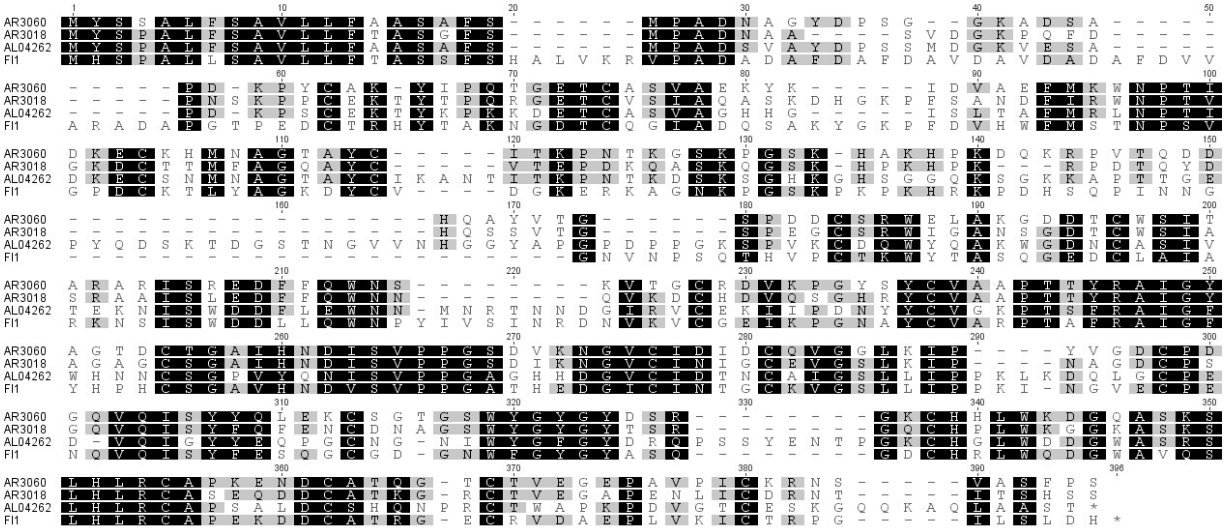


Figure 12. Multiple sequence alignment of EbCE4 and its homologues in three *Epichloë* strains. Amino acid sequences of EbCE4 from *E. bromicola* AR3060, and homologues from *E. bromicola* AR3018 (AR3018G0100684A1), *E. bromicola* AL04262 (maker-contig01824-fgenesh-gene-0.39) and *E. festucae* Fl1 (EFM3.029340) are shown in the alignment. The alignment was generated by ClustalW using the Blosum matrix. The shading intensity indicates frequency of occurrence of an amino acid in the specific position. Black: very frequent. Grey: not as frequent.

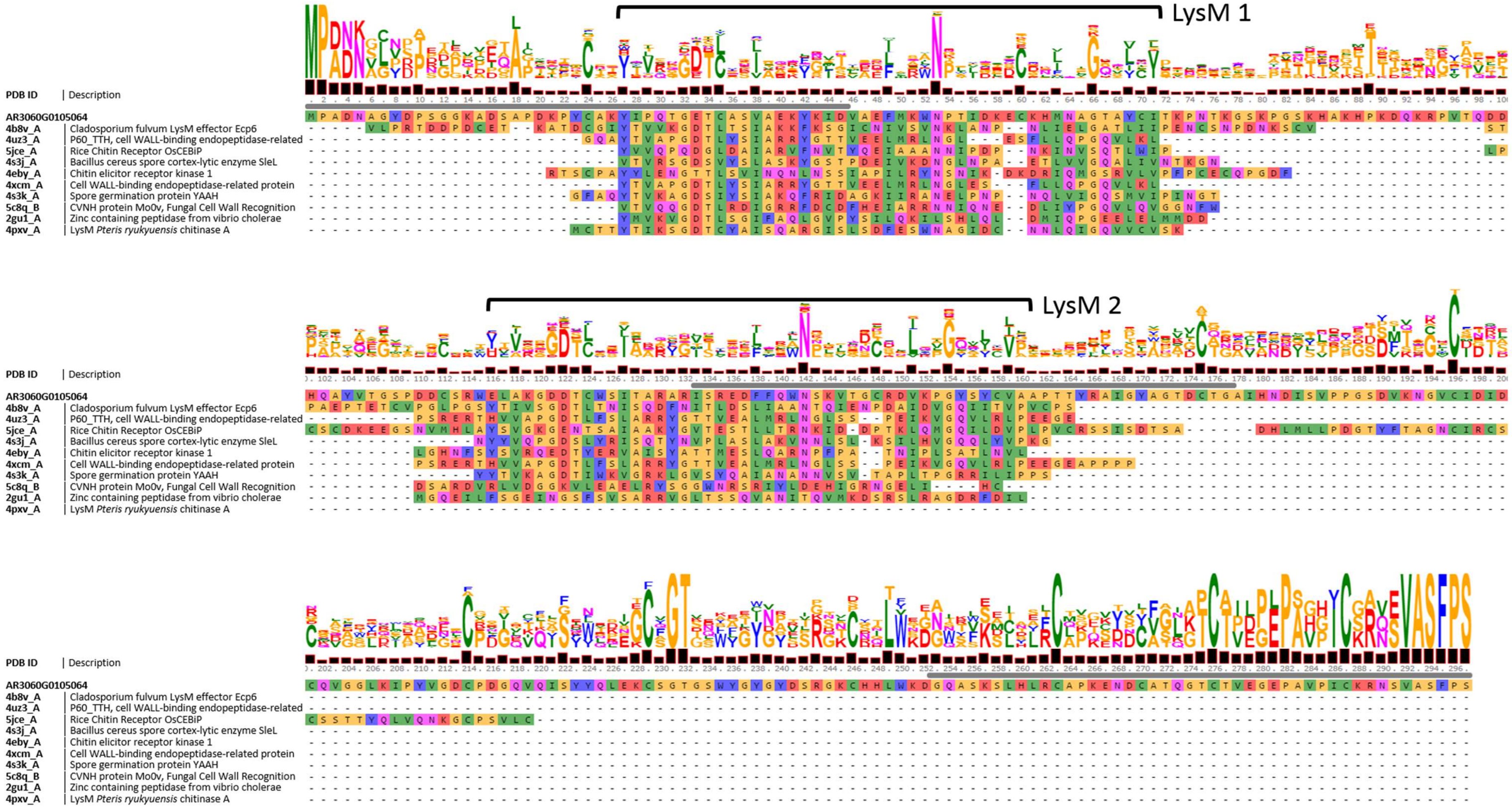


Figure 13. Amino acid sequence alignment of EbCE4 with secondary structure homologous proteins in the Protein Database (PDB). Sequences were the top hits from HHpred homology search. Clustal X color scheme was used. Two LysM domains in EbCE4 are marked as LysM1 and LysM2.

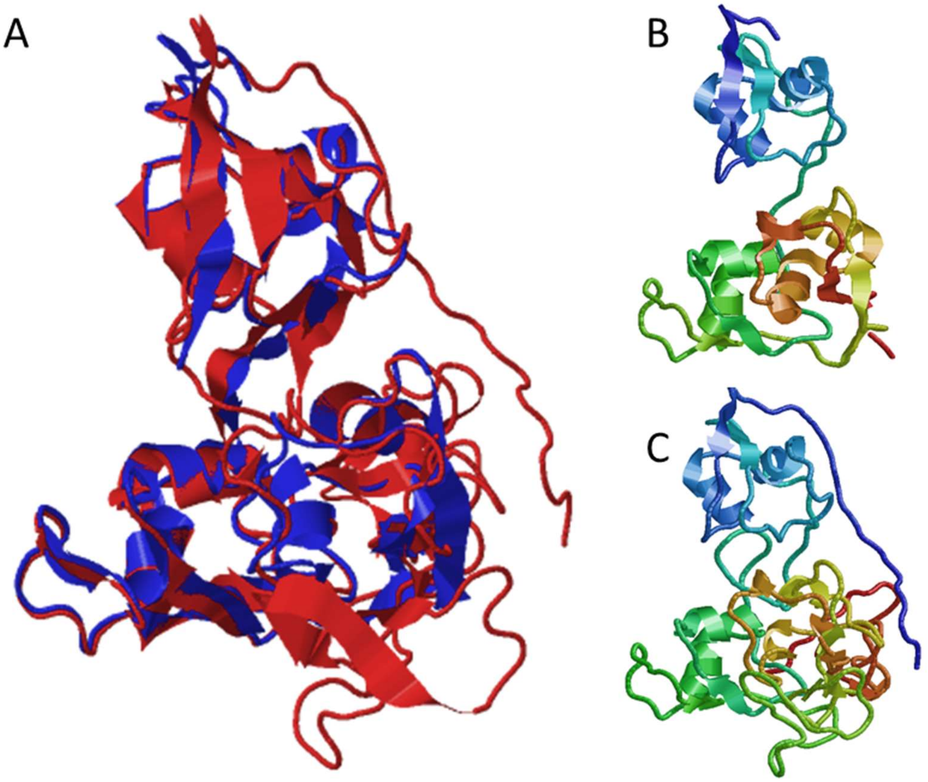


Figure 14. 3D structure alignment of EbCE4 and Ecp6 protein. The putative LysM effector EbCE4 and Ecp6 have a highly similar structural fold. A, pairwise structure alignment using the TM-align server with a TM-score = 0.94227. TM-score has value in (0,1], where 1 indicates a perfect match between two structures. EbCE4 is in red and Ecp6 in blue. B, 3D structure of EbCE4 predicted by I-TASSER. C, Crystal structure of Ecp6 from the Protein Data Bank (PDB: 4B8V).

4. Transcriptome experiment II: *Epichloë bromicola* in progeny seedlings

4.1 Fungal transcriptomics of progenies of *E. bromicola* infected wheat

Epichloë bromicola endophyte AR3060 forms a compatible mutualistic symbiosis with its native host *Elymus excelsus*. This particular *Epichloë* strain can also colonise spring wheat "Monad" and "TAC11" through artificial inoculation, and is vertically transmitted to the next generation by infecting wheat seeds. However, the infected "Monad" plants are dwarfed with delayed plant development, and both infected "Monad" and "TAC11" wheat produce shrivelled seeds with low germination rates (Simpson et al. 2014, unpublished data). Moreover, despite the seed from the inoculated "Monad" wheat being able to germinate, the progeny plants are stunted and die before flowering stage. To examine gene expression differences in the interactions AR3060 forms with these three different hosts, and to further understand the *E. bromicola*-wheat incompatibility at an early stage in a more natural condition, transcriptomes of progeny seedlings geminated from *Elymus* and wheat seeds infected with AR3060 were sequenced using Illumina RNA-seq.

4.1.1 Transcriptome sequencing and mapping of reads to genome references

In this transcriptome study, three replicates were used for each seedling and mycelium RNA sample. In total, 21 samples were sent to BGI for transcriptome sequencing, including 9 samples of AR3060-symbiotic seedlings ("Monad" ×3, "TAC11" ×3, and *Elymus* ×3), 9 samples of endophyte-free seedlings as negative controls ("Monad" ×3, "TAC11" ×3, and *Elymus* ×3), and 3 samples of AR3060 grown in axenic culture. Due to the commonly low biomass (<2%) of *Epichloë* endophytes in the infected host plants (Young et al. 2005, Rasmussen et al. 2007), 20G bases were sequenced for each endophyte-symbiotic plant samples in order to achieve better fungal genome coverage, in comparison, 12Gb were sequenced for endophyte-free plants and 4Gb for in culture samples.

Nine AR3060-symbiotic plant RNA-seq libraries generated approximately 66~70 million clean reads per sample after quality filtering, while AR3060 in culture generated around 15 million reads (Table 4). Fungal reads were then mapped onto the AR3060 genome and "Monad" wheat reads mapped onto the wheat reference genome obtained from the International Wheat Genome Sequencing Consortium (IWGSC). Due to the lack of an available *Elymus* and *Leymus* genome ("TAC11" has a chromosome addition from *Leymus racemosus*), *Elymus* and "TAC11" reads were also both mapped onto the same wheat reference genome.

As shown in Table 4, over 90% of AR3060 reads mapped to the AR3060 genome for the first two AR3060 in culture biological replicates (p3060-1 and p3060-2), but the third replicate had only 75.45% of clean reads mapped, which may have been caused by culture contamination. Expectedly, less than 1% of all the filtered *in planta* reads were successfully mapped because of low hyphal biomass *in planta*. However, while about 0.5% - 1% of reads from endophyte-symbiotic *Elymus* and "TAC11" mapped to the fungal genome, less than 500 fungal reads from the AR3060-symbiotic "Monad" wheat (pMo3060-1, pMo3060-2 and pMo3060-3) transcriptome mapped to the AR3060 genome, comprising less than 0.001% of total mapped reads in the corresponding samples. Considering the stunted phenotype of progenies of AR3060-symbiotic "Monad" wheat, such low fungal reads may indicate that the fungus was not able to successfully colonise the plant. Therefore, "Monad" samples are not included in the following analysis.

The AR3060 gene models used in this study were predicted by the MAKER program. Of about 8000 predicted genes, more than 6000 were detected from the transcriptome of AR3060-symbiotic "TAC11" and *Elymus* seedlings. Among three "TAC11" replicates, pTAC3060-3 has almost no fungal reads, indicating that this sample is endophyte-free yet produced a false-positive result in an immunoblot test for endophyte infection confirmation. On the other hand, *Elymus* sample pEl3060-3 showed a surprisingly high amount of endophyte reads (3.58%), which is rather uncommon. Therefore, these two samples (pTAC3060-3 and pEl3060-3) were not used for the following transcriptomic analysis (Table 4).

Table 4. Summary of RNA-seq read alignments to reference genomes

Sample[a]	Filtered reads[b]	Mapped reads[c]	%	AR3060 reads[d]	%	AR3060 genes[e]
pMo3060-1	66,769,321	46629495	73%	72	0.0001%	6
pMo3060-2	67,037,132	45841577	71%	490	0.0007%	254
pMo3060-3	66,811,482	45554808	71%	469	0.0007%	268
pTAC3060-1	**67,091,624**	**45144244**	**70%**	**361,049**	**0.54%**	**6403**
pTAC3060-2	**67,057,810**	**45258955**	**70%**	**416,503**	**0.62%**	**6330**
pTAC3060-3	66,817,218	44985106	70%	16	0.00002%	6
pEl3060-1	**70,269,895**	**41707242**	**62%**	**501,266**	**0.71%**	**5418**
pEl3060-2	**70,067,031**	**41475829**	**62%**	**730,730**	**1.04%**	**6532**
pEl3060-3	70,135,770	34744878	52%	2,513,298	3.58%	5483
p3060-1	**15,665,828**	**14119154**	**98%**	**14,106,895**	**90.05%**	**8321**
p3060-2	**15,693,387**	**14250727**	**99%**	**14,239,657**	**90.74%**	**8437**
p3060-3	15,262,403	11542032	90%	11,516,199	75.45%	8382

[a] RNA samples of AR3060-symbiotic "Monad" wheat (pMo3060), AR3060-symbiotic "TACBOW0011" wheat (pTAC3060), AR3060-symbiotic *Elymus dahuricus subsp. excelsus* (pEl3060), and AR3060 grown in axenic culture (p3060). Darken lanes are replicates used for RNA-seq analysis.

[b] Total number of reads passing quality control.

[c] Reads mapping to genome of either AR3060 or wheat, and the corresponding percentage (%) of filtered reads.

[d] Reads mapping to AR3060 genome reference, and the corresponding percentage (%) of mapped reads.

[e] Number of unique AR3060 genes detected from each sample.

4.1.2 The most highly expressed *E. bromicola* genes in progeny seedlings

The expression profile of *E. bromicola* genes in progeny seedlings is very similar to that in mature plant hosts in many ways. The in culture samples showed rather evenly distributed gene expression across large sections of the genome, while *in planta* samples had many genes with extremely abundant expression (Figure 15). The top 20 most highly expressed AR3060 transcripts in *Elymus* comprised more than 42% of the total fungal reads in this native host association. On the other hand, the top 20 fungal transcripts in culture collectively comprise just 7.9% of the total reads.

The most highly expressed endophyte genes in culture and *in planta* of Experiment II have many genes in common with Experiment I. For example, several heat shock proteins, a clock-controlled protein and a cell wall mannoprotein, were again present in the top 20 fungal transcripts in culture. Besides these there were other genes that were highly expressed in culture in both Experiment I and II, including, elongation factor 1, exonuclease III, ammonium transporter, histone H3, glutamine synthetase and ADP/ATP carrier protein. Similarly, for the mature plant host samples in Experiment I, most genes that were highly expressed in progeny seedling samples in Experiment II also lacked similarity with functionally annotated genes. Among the top 20 most highly expressed fungal genes, 14 were shared between Experiment I and II for *Elymus* samples, and 12 for wheat samples, but only 6 had functional annotations. *GigA* (Johnson et al. 2015) still had the highest gene expression in both *Elymus* and TAC11 interactions in Experiment II, indicating it likely has a critical role in *E. bromicola*-plant interactions. Also similar to Experiment I, the putative LysM effector EbCE4 (AR3060G0105064) was highly expressed in progeny seedlings. However, it showed higher expression in TAC11 wheat compared with Monad wheat. EbCE4 was the second highly expressed fungal gene in TAC11 progeny seedlings with almost half the read counts of the extremely abundant *gigA*. Other transcripts that were abundant in *Elymus* progeny seedlings putatively encode a CAMK (Ca^{2+}/calmodulin-dependent protein kinase) family protein, an endo-1,6-beta-glucanase, a heat shock 70 kDa protein (Hsp70) and a protein disulfide-isomerase (PDI). Apart from the genes that were shared between *Elymus* and TAC11 (*gigA*, EbCE4, CAMK, Hsp70 and endo-1,6-beta-glucanase), fungal genes that were most highly expressed in TAC11 also include another Hsp70, an elongation factor 1 and a chitinase. Details of the 20 most highly expressed genes in culture, *Elymus* and TACBOW wheat of Experiment II are listed in Appendix 8.

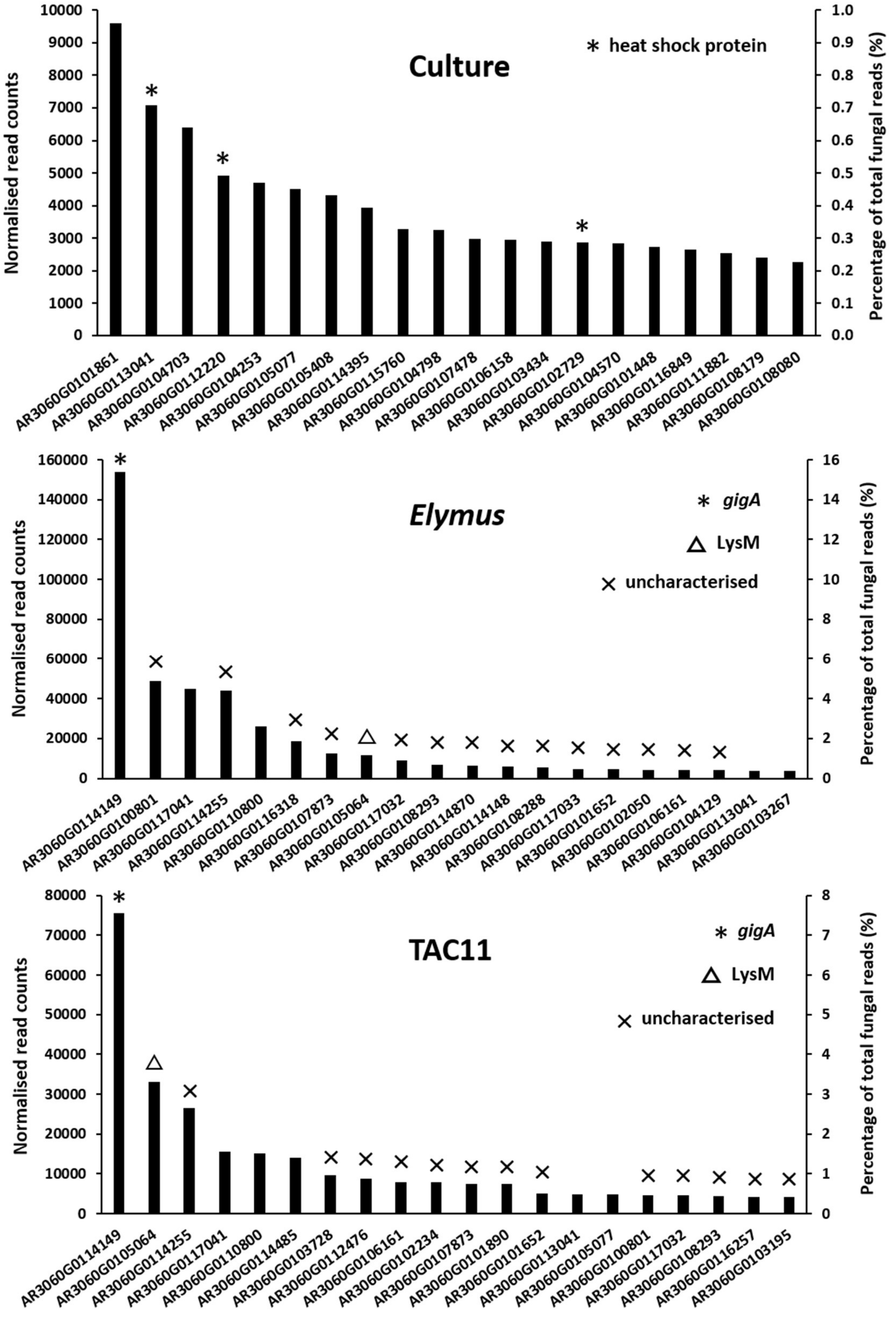


Figure 15. The 20 most highly expressed *E. bromicola* AR3060 genes in progeny seedlings and in culture. Primary vertical axis (left) shows normalised RNA-seq counts of fungal reads as an indication of gene expression level. Secondary vertical axis (right) shows the proportion of the transcript in total reads in each samples. Culture: AR3060 grown in axenic culture. *Elymus*: AR3060-symbiotic *Elymus excelsus*. TAC11: AR3060-symbiotic "TACBOW0011" wheat.

4.1.3 Fungal DE genes in progeny seedlings

E. bromicola genes that are differentially expressed in three samples (culture, native *Elymus* host, and TAC11 wheat) were identified by the R package EdgeR using common dispersion estimated from the deviance of the read counts. Three comparisons were analysed, including *Elymus* vs culture (pEl3060/p3060), TAC11 vs culture (pTAC3060/p3060), and TAC11 vs *Elymus* (pTAC3060/pEl3060). In total, 2382 fungal genes were found to be differentially expressed in at least one comparison (Figure 16). These genes were either expressed differentially in culture vs *in planta*, or in native *Elymus* host vs novel wheat host. Of all 2382 fungal DE genes, 2353 genes (nearly 99%) were considered to be *in planta*-induced. Among *in planta*-induced DE genes, approximately 68% (1594 out of 2353) were not shared by *Elymus* and TAC11, indicating a host-specific gene expression pattern. Among those genes expressed at similar levels in culture and *in planta*, only 29 genes were specifically differentially expressed between TAC11 and *Elymus*. However, there were less genes (~17%, 402 out of 2382) that were significantly affected by the change of plant host in progeny seedlings compared with mature plants (~30%, 254 out of 853), suggesting a less dramatic gene expression profile change in progeny seedlings of TAC11.

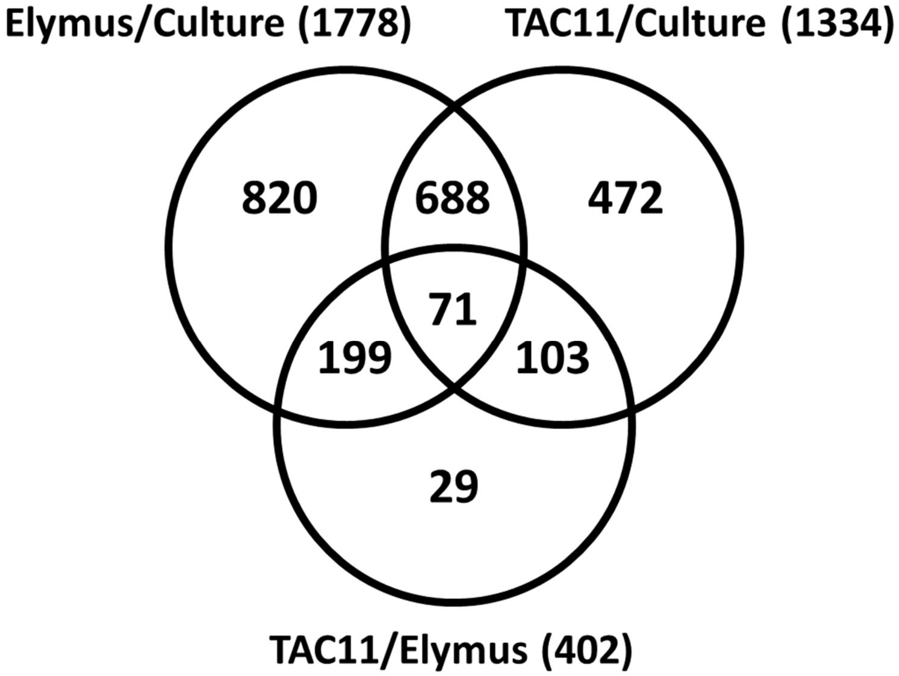


Figure 16. Venn diagram of differentially expressed fungal genes across three comparisons. Differentially expressed genes in AR3060 with P value <0.05 and fold change >2 in at least one of the pairwise comparisons were detected by R package EdgeR using common dispersion estimated from the deviance of the read counts.
Culture: AR3060 grown in axenic culture. Elymus: AR3060 symbiotic with *Elymus excelsus*.
TAC11: AR3060 symbiotic with "TACBOW0011" wheat.

The 2382 endophyte DEGs across the three comparisons were clustered and analysed using the Self Organizing Tree Algorithm (SOTA) in MeV4.9 (Figure 17), which provided better gene expression clustering than hierarchical or K-mean methods for this large gene set. DEGs were grouped into six clusters based on their distinct gene expression patterns, demonstrating different patterns of gene regulation for different sets of genes. Clusters I and V are similar groups of a relatively small number of endophyte genes which were expressed in culture and also expressed in one of the two progeny hosts. Clusters II, III, and IV all showed a clear *in planta*-induced expression pattern. However, genes in Cluster IV were not differentially expressed between TAC11 and *Elymus*. Therefore, Clusters II and III that have a strong expression pattern difference between TAC11 and *Elymus* host could contain many genes of interest related to the novel *E. bromicola*-TAC11 interaction. Many genes in Cluster II and III have no homology to characterised ones and some of them encode small secreted proteins. Cluster VI is a group of endophyte genes that are relatively highly expressed in culture but low in both plant hosts. Thus they are supposedly not related to the colonisation of plant hosts and were therefore not included in the following revision of the effector candidate list. Information of all genes in both RNA-seq Experiment I and II are combined and listed in Supplementary Table 2.

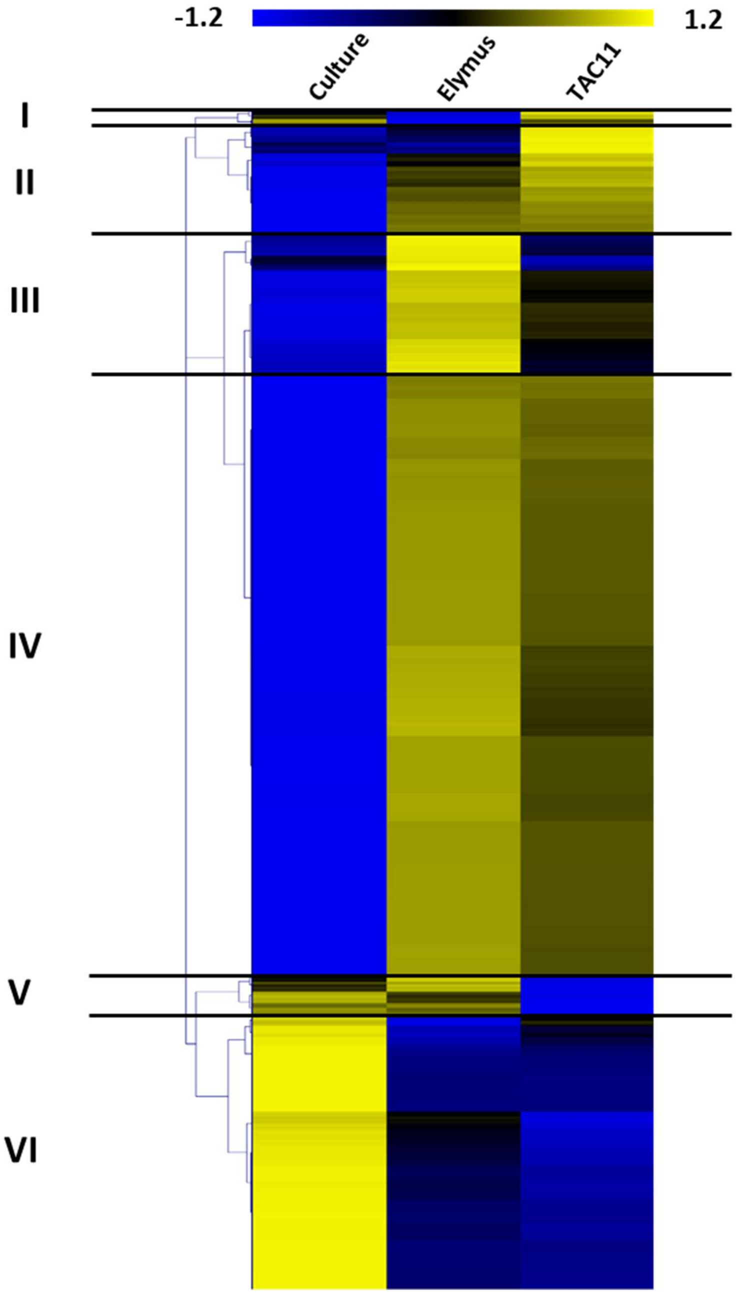


Figure 17. Expression clustering of 2382 differentially expressed fungal genes in progenies of AR3060-symbiotic *Elymus* and "TAC11" wheat. Differentially expressed genes in AR3060 with a P value <0.05 and fold change >2 in at least one of three pairwise comparisons were detected by R package EdgeR. Genes have been grouped into six clusters based on their expression patterns using the Self Organizing Tree Algorithm (SOTA). Gene expression counts were normalized in MeV4.9 using the mean and the standard deviation of the row of the matrix to which the normalised read count value belonged, using the formula: Value = [(Value) – Mean(Row)]/[Standard deviation(Row)]. Culture: AR3060 grown in axenic culture. *Elymus*: AR3060-symbiotic *Elymus excelsus*. TAC: AR3060-symbiotic "TAC11" wheat. Blue: low expression. Yellow: high expression.

4.1.4 Gene ontology analysis of fungal DE genes in progeny seedlings

For an overall understanding of the functions of the DEGs, gene ontology (GO) enrichment analysis was conducted for the 402 fungal DE genes in TAC11 vs *Elymus* (pTAC3060/pEl3060). Around 44% (175 out of 402) of the differentially expressed genes could be assigned with GO annotations using Blast2GO. Annotations were then further improved by combining GO annotations obtained via different approaches including InterProScan, KEGG, and Blast2GO. After the improvement, approximately 52% (208 out of 402) of DEGs were assigned with GO annotations and they fell into 22 Biological Process GO terms and 20 Molecular Function GO terms. The 10 GO terms with the most significant enrichment are listed in the table of Appendix 5. The DEGs were significantly enriched in secondary metabolic process, carbohydrate catabolic process and alcohol metabolic process under the Biological Process category (Appendix 6). They were also enriched in serine peptidase/hydrolase activity, endopeptidase activity, and several other hydrolase activities under Molecular Function category (Figure 18). Similar to *E. bromicola* in mature plant hosts, DEGs in progeny seedlings were also enriched in several GO groups related to transmembrane transporter activity and ion binding activity.

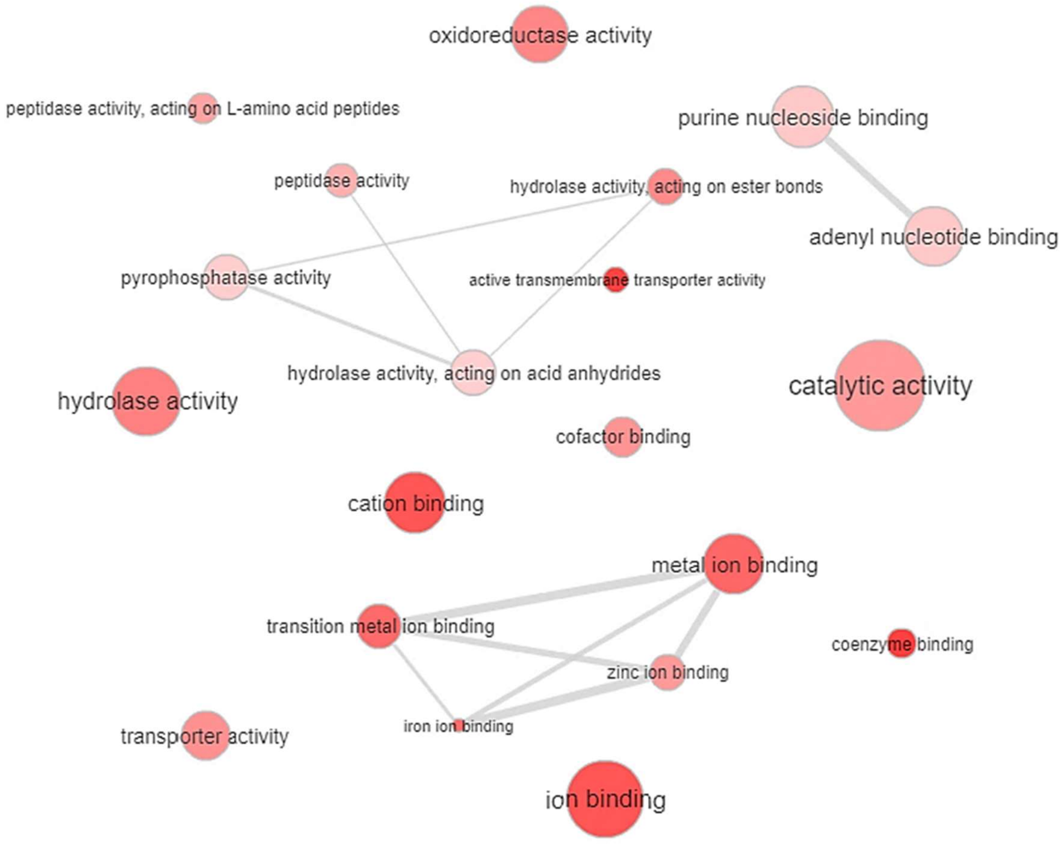


Figure 18. Fungal gene ontology groups differentially expressed *in planta*. REVIGO interactive map of GO terms under Molecular Function category enriched in differentially expressed genes between AR3060-symbiotic "TAC11" wheat and *Elymus*. The sizes of the red circles indicate the relative frequencies of GO terms. Darker shading indicates lower P values. Highly similar GO terms are linked by lines with arbitrary length.

4.2 Validation and revision of *E. bromicola* effector candidates

Apart from providing better understanding of overall fungal transcriptional changes in *E. bromicola*-symbiotic plants during early stages of colonisation, the transcriptome data of Experiment II provided an opportunity to validate the gene expression of *E. bromicola* candidate effectors (EbCEs). It also generated a more reliable assembly of *E. bromicola* reads for inspecting the accuracy of these gene model predictions, since Experiment II has more biological replicates and higher sequencing coverage of the genome. Therefore, a reference-based assembly of AR3060 RNA-seq reads in this Experiment II was generated and compared to the previous gene models in Experiment I. Moreover, a new AR3060 genome sequence obtained using PacBio became available during the project. This was also used to examine gene model predictions of EbCEs. After assessment, EbCE11, EbCE18, EbCE26, EbCE27 and EbCE28 were removed from the candidate list because they had no Maker predictions in the new PacBio sequenced genome and/or not much coverage of these regions by RNA-seq reads of Experiment II.

Using gene expression data from Experiment II, *E. bromicola* genes were filtered again based on the same criteria of the effector prediction process outlined in Chapter 3 (Experiment I), resulting in an addition of four new candidates named EbCE29, EbCE30, EbCE31 and EbCE32. EbCE29 is an Endo-b-1,4-glucanase that was among the most highly *in planta*-expressed fungal genes in both RNA-seq experiments. EbCE30 was found to have around 64% identity to a Glycosidase from *Neosartorya fumigata* (*Aspergillus fumigatus*). EbCE 31 and EbCE32 were uncharacterised small proteins. In the end, 27 candidate fungal effectors were identified in *E. bromicola* AR3060, 18 of which have no characterised function/domain. The annotations of all candidate genes, along with their expression profile in both RNA-seq experiments are shown in Table 5.

Table 5. Revised *E. bromicola* effector candidate list with gene expression in transcriptome experiments I and II

		Mature plant (RNA-seq I) [b]			Progeny seedling (RNA-seq II)							
Candidate [a]	**Gene ID**	**Culture**	***Elymus***	**Monad**	**Culture**	***Elymus***	**TAC11**	**Feature [c]**	**Cys [d]**	**Length [e]**	**SP [f]**	**Location [g]**
EbCE1	AR3060G0114255	0	5661	5042	27	44144	26468	Uncharacterised	9	110	Y	extr
EbCE2	AR3060G0116318	0	2308	4924	0	18748	2451	Uncharacterised, *Epichloe* unique	2	61	Y	extr
EbCE3	AR3060G0114870	0	2143	78	0	6273	504	Uncharacterised	6	179	Y	extr
EbCE4	AR3060G0105064	0	905	4658	5	11648	33173	LysM domain protein	18	316	Y	extr
EbCE5	AR3060G0107873	0	898	865	12	12609	7571	Uncharacterised	7	144	Y	cyto
EbCE6	AR3060G0113595	0	559	48	13	347	4	Uncharacterised	8	287	Y	extr
EbCE7	AR3060G0102050	0	417	612	0	4327	605	Uncharacterised	6	98	Y	extr
EbCE8	AR3060G0106159	0	355	41	19	187	0	Ribonuclease/ribotoxin	4	155	Y	extr
EbCE9	AR3060G0103728	1	195	840	42	2407	9684	Uncharacterised	4	235	Y	extr
EbCE10	AR3060G0104034	0	104	1083	4	541	2815	Uncharacterised, *Epichloe* unique	4	104	Y	extr
EbCE12	AR3060G0113163	0	8	187	7	117	1294	Uncharacterised	4	169	Y	extr
EbCE13	AR3060G0100112	1	45	225	16	601	3295	Pectinesterase	4	334	Y	extr
EbCE14	AR3060G0100437	40	205	10	5	998	192	Uncharacterised, *Penicillium* homolog	0	271	Y	cyto
EbCE15	AR3060G0101890	2	194	214	2	2835	7566	Uncharacterised	6	115	Y	extr
EbCE16	AR3060G0104410	0	77	2	52	49	0	Uncharacterised, *Epichloe* unique	3	193	Y	extr
EbCE17	AR3060G0105601	2	217	961	20	1159	3875	Uncharacterised/UPF0047 domain	6	270	Y	extr
EbCE19	AR3060G0106335	5	853	86	253	2722	341	Subtilisin-like protease	4	351	Y	extr
EbCE20	AR3060G0111412	3	162	14	214	229	34	Killer toxin Kp4-like	10	131	Y	extr
EbCE21	AR3060G0111534	2	41	256	27	324	1782	Uncharacterised	4	237	Y	extr
EbCE22	AR3060G0112476	5	44	454	1215	875	8848	Uncharacterised	2	151	Y	extr
EbCE23	AR3060G0114149	17	25865	7596	61	153970	75639	Uncharacterised, *Epichloe* unique, nc25	0	70	Y	extr
EbCE24	AR3060G0114485	11	298	1201	1193	2875	14065	Chitinase	5	454	Y	extr
EbCE25	AR3060G0101652	63	223	254	160	4388	5142	Hydrolase	6	367	Y	extr
EbCE29 [h]	AR3060G0110800	9	1922	1043	390	25905	15145	Endo-b-1,4-glucanase	5	431	Y	extr
EbCE30	AR3060G0108298	5	257	177	12	1956	2454	Glycosidase	6	349	Y	extr
EbCE31	AR3060G0117034	24	1374	1370	2	3612	2077	Uncharacterised	3	357	Y	extr
EbCE32	AR3060G0105029	0	19	139	16	149	2	Uncharacterised	4	307	N	extr

[a] All candidate genes have no transmembrane helix as predicted by TMHMM, no GPI anchor as predicted by Big-PI.

[b] Normalised read counts of candidate genes in RNA-seq Experiment I and II. Culture: *Epichloë bromicola* AR3060 grown in axenic culture. *Elymus*: AR3060 symbiotic with native host *Elymus excelsus*. Monad: AR3060 with “Monad” wheat. TAC11: AR3060 with “TACBOW0011” wheat.

[c] Features are generated based on BLAST hits from databases (Swiss-prot, Uniprot, Refseq, Pfam) and HHpred search results.

[d] Cys, number of cysteine. [e] Length, amino acids (length of the encoded protein product). [f] SP, signal peptide. [g] Location was predicted by WoLF Psort and TargetP.

[h] EbCE11, EbCE18, EbCE26, EbCE27 and EbCE28 were removed from the first version list. EbCE29- EbCE32 were new candidates.

5. Discussion

5.1 Transcriptomics gives insights into *Epichloë*-host compatibility

Epichloë endophyte typically form mutualistic symbioses with specific grasses and impart benefits to its host (a compatible symbiosis). However, *E. bromicola* strain AR3060 can colonize wheat through artificial inoculation, but resulting in stunted or dwarfed plants (an incompatible symbiosis). The aim of this study was to investigate the molecular events underpinning compatible and incompatible symbioses and potentially find gene targets that could affect such an outcome.

Transcriptomics is a great tool to look into the fungal and plant gene expression changes during the interaction of *Epichloë* endophytes and their plant hosts. Through analysing the large amount of data generated from the two RNA-seq experiments, fungal gene expression patterns have been identified and validated. Although RNA-seq Experiment I had only one replicate, it has still generated much useful information, most of which are consistent with the more robust Experiment II. The outstanding feature of the *in-planta* expression patterns of *E. bromicola* genes is the exceptional abundance of a few fungal transcripts and several genes that comprise a large percentage of the total fungal reads. For instance, the 20 most highly expressed AR3060 transcripts in *Elymus* comprised nearly 50% of the total fungal reads in that sample. Such abundance has also been observed by Ambrose and Belange (2012). Using SOLiD-SAGE, they identified differentially expressed plant genes and some surprisingly abundant fungal genes. NC25 (later named *gigA*) was their fourth most highly expressed fungal gene and NC12 had the highest expression (Johnson et al. 2003; Johnson et al. 2015; discussed more below in page 69). Similar to this study, most of the fungal transcripts that are most abundant *in-planta* were secreted proteins with unknown function. An antifungal protein gene (EbCE14) that has a homolog in *Penicillium* showed up among the highly expressed fungal genes. Interestingly, the second most expressed fungal tag in their study was also for an antifungal protein that is similar to genes in *Penicillium* and *Aspergillus* spp., suggesting its relevance in disease resistance provided to the plant host. Such an antifungal protein may be secreted by the endophyte during colonization preventing infection of an alternative, possibly pathogenic fungus.

Recent transcriptomic studies of *Epichloë*-grass symbioses demonstrated that endophyte genes are expressed significantly differently in different plant parts (Schmid et al. 2016; Nagabhyru et al. 2019). Highly expressed genes of actively elongating hyphae in growing plant tissues were related to hyphal growth. In contrast, highly expressed genes in mature host tissues were related to synthesis of antiherbivore compounds (Schmid et al. 2016). Although pathways related to secondary metabolism were enriched in GO analysis, antiherbivore compound-related genes were not found among the highly expressed fungal genes in the current study, which may be the result of whole plant sampling in RNA-seq experiment I. However, the concordance between RNA-seq experiment I and II confirmed the expression profiles of most fungal genes and, therefore, demonstrated the reliability of the transcriptomic data. RNA-seq experiment II was the unique transcriptome profiling of a group of *Epichloë*-plant symbioses. It used progeny of *Epichloë*-symbiotic *Elymus* and wheat plants, and endophyte-free *Elymus* seeds collected from nature. So, unlike other studies, there will be no inoculation wounds on the plant and no chemical treatment needed to generate endophyte-free native host plants. Such associations provided more accurate comparison between samples.

5.2 The effector armoury of *E. bromicola* might facilitate colonization

Pathogens utilise effectors to supress host defence responses. In order to establish a compatible symbiotic association with the host plant, *Epichloë* endophytes also need to avoid detection or modulate the host defence responses. Although many genes that encode effector-like small secreted proteins have been found in the genomes of mutualistic symbionts, only several were confirmed to be effectors that facilitate host defence suppression (Akum et al. 2015; Wawra et al. 2016). Little is known regarding how mutualistic symbionts like *Epichloë* species handle plant host immune responses. However, transcriptomic studies showed major expression/up-regulation of large number of small secreted proteins of *Epichloë in-planta*, indicating the mutualistic endophyte may alter/avoid the host defence system in a similar way to fungal pathogens.

While most effector-like proteins are of unknown function and have little similarity with other effectors, some are well characterised, such as LysM-domain containing proteins. Some of them bind chitin in the cell wall as Avr4 in *Cladosporium fulvum* (van Esse et al. 2007) and some of them isolate chitin oligomers in the intercellular space as Ecp6 (de Jonge et al. 2010). There are also LysM proteins in *Colletotrichum higginsianum* that participate in both chitin-triggered immunity suppression and appressorium development. The function of the LysM candidate effector (EbCE4) in this study may be similar to Ecp6 as they have significant similarity in predicted secondary and 3D structures. A recent microscopy study of *E. festucae* using a chitin-binding hyphal stain showed that only the septa of endophytic hyphae were stained (Scott et al. 2018). There may be proteins like EbCE4 that modify or mask chitin in endophytic hyphae of *Epichloë* endophytes to avoid detection of the plant defence system.

The other gene of high interest is *gigA*. The striking feature of *gigA* is its high expression. It was the most highly expressed fungal gene in every plant sample in this study. This gene was previously found and named NC25, along with another secreted protein NC12, in *E. coenophiala*-symbiotic tall fescue (Johnson et al. 2003). It encodes a protein that is secreted and cleaved post-translationally to form a cyclic oligopeptide (Johnson et al. 2015). However, the function of the cyclic oligopeptide is still unknown. Given its extremely high expression *in-planta*, it is likely to be involved in some fundamental process (e.g., hyphal growth, defence suppression) which is critical to the fungal-plant interaction. It is also possible that *gigA* plays a bioprotective role that is similar to EAS and LTM genes, and has no impact on the symbiosis.

Among all effectors predicted through the effector mining pipeline, only a few have known function assigned. Aside from *gigA* and the LysM-domain protein EbCE4, there is a putative ribotoxin, EbCE8. Fungal ribotoxins are a set of ribonucleolytic proteins that are secreted by fungi (Lacadena et al. 2007). They specifically recognize and cut a universally conserved sequence of rRNA gene, resulting in cell death. Recently, a ribotoxin-like effector CSEP0064/BEC1054 has been identified in *Blumeria graminis*, a biotrophic fungal pathogen (Pennington et al. 2019). It is highly expressed *in-planta* and increases susceptibility to infection. Although the exact function of this effector is unknown, it is very likely that it counters endogenous plant RNases like ribosome-inhibiting proteins (RIPs) which are used by the plant defence system to trigger apoptosis in order to stop the spread of the fungal infection. EbCE8 may function similarly by suppressing the host RIPs to facilitate colonization.

EbCE13 is predicted to be a pectinesterase. Pectinesterases are cell wall active enzyme that are usually used by all land plants and some fungi to modify plant cell walls. Since *Epichloë* endophytes grow in the intercellular spaces of the plant host, it is reasonable to have pectinesterases that alter plant cell walls along with hyphae growth. However, in turn, it is also likely to be recognized by the plant defence system.

EbCE24 is identified as a chitinase. The chitin in fungal cell wall is a common pathogen-associated molecular patterns (PAMPs) that can be recognized by plant chitin receptors. EbCE24 may be required for the establishment of fungal-plant symbiosis as it is indicated for mycorrhiza-specific chitinase in *Medicago truncatula* (Salzer et al. 2007).

Although several of the candidate effectors identified in this study have annotations that are often seen in pathogen effectors, functional characterization of the 27 candidate effectors *in-planta* will still be difficult, as most of them are of unknown function, and may be redundant in function.

5.3 Computational mining for fungal effectors, challenging yet promising

Running genomic data through a computational mining pipeline is one of the most efficient ways to identify new fungal effectors for further characterisation. However, fungal effectors are notoriously difficult to predict as they normally share no common sequence feature. General criteria including small size, presence of a signal peptide, high percentage of cysteines can be used for mining but there are still many challenges.

Firstly, erroneous genome assembly and annotations greatly affect effector prediction. Effectors are often located in repeat-rich regions which make it reasonably difficult to achieve a complete and accurate assembly, thereby leading to missing data and errors in the identification of effectors from genome sequence (Thomma et al.

2016). In order to minimise the impacts of assembly and gene annotation errors in this project the first step was to inspect the quality of the *E. bromicola* genomes and to assess the accuracy of the gene annotations. Manual annotations of several randomly picked highly abundant fungal transcripts and inspection of transcript distribution in IGV revealed incorrect exon/intron boundaries, an absence of gene prediction in some transcribed regions, and missing start/stop codons for certain transcripts (data not shown). To improve the quality of the gene models, the alignment-based gene predictions of AR3060 and AR3018 were supplemented with the Illumina RNA-seq data. Reference-based assemblies of AR3060 and AR3018 RNA-seq reads were generated and compared to the previous Maker predicted gene models to obtain precise exon/intron boundaries and start/stop sites. After reassessment, 1766 and 438 gene models were corrected or newly added into the AR3060 and AR3018 genomes, respectively.

Secondly, the criteria used for effector mining are full of exceptions. Cysteines are considered crucial for maintaining a protein structure in the intercellular space but the apoplastic effector AvrLm1 from *Leptosphaeria maculans* contains only one cysteine (Gout et al. 2006). There are effectors that do not have a signal peptide and are non-classically secreted (Liu et al. 2006). There also is an effector that is highly expressed without the contact of the plant host and not differentially expressed during infection (Akum et al. 2015). Therefore, it is critical to set proper mining steps and thresholds for fungal effector prediction. In order to include as many effector candidates as possible while still keeping criteria strict, the mining steps have been carefully modified in this study. The widely used protein size filter and cysteine percentage filter have been removed and it is the first endophyte effector mining pipeline that put exceptions into consideration at every filtering steps. After running through automatic steps which are mainly based on transcriptomic data, genes that did not pass the next step were selected again through three exception steps with the aim of three rare types of effectors: plant-signal independent candidate effectors (not differentially expressed), non-classically secreted candidate effectors (no signal peptide), and effectors with known effector motifs.

The last but not least problem with common effector mining pipelines and automated servers is the lack of supporting gene expression data and protein structure information. In this study, transcriptomic data helped narrow the candidate list from hundreds of genes to around 30, as many effector-like secreted proteins showed no expression *in-planta*; and the lysM domain containing putative effector EbCE4 proved to be a homologue of Ecp6 through secondary and 3D structure comparison even though they have a low amino acid similarity. The combination of genomic, transcriptomic and structural data is crucial to narrow down and prioritize candidates. With fast increased availability of next-generation sequencing and rapidly advanced computational prediction tools, gene expression data and secondary and tertiary structure information should be integrated to increase the accuracy of effector identification.

5.4 Future work

Although the stunted and dwarfed plants of *E. bromicola*-symbiotic wheat have been observed, the microscopic phenotype of hyphae and plant cells of the incompatible interactions has not been documented. A microscopy study can be used to evaluate the progression of incompatibility of *E. bromicola* in its novel wheat hosts and compare its growth in wheat against the native host, *Elymus*.

For the candidate effectors filtered in this study, an agro-infiltration assay is currently being performed to induce transient expression of the fungal gene in *Nicotiana benthamiana* and *N. tabacum* leaves. Candidate genes that cause necrosis and production of reactive oxygen species (ROS) on the infiltration spot will be selected as it may indicate the induction of plant defence responses. Gene deletion or overexpression studies can be conducted to confirm the function of EbCEs *in planta*. *E. bromicola* AR3060 mutants will be characterised by inoculating recombinant strains into the native *Elymus* host and the novel wheat hosts (Monad and TAC11). The effects of targeted gene deletion on the phenotype of the strains in axenic culture and in the native and novel hosts can be compared against wild type strains using microscopy techniques such as differential interference contrast, transmission electron microscopy and confocal microscopy. The effects of the mutant strains on host plants will be established through detailed examination of host morphology, growth and development, and physiology.

For candidate effectors that were positive in functional characterisation, protein terminal GFP fusions and confocal microscopy can further be used to visualise their localisation during *Epichloë*-wheat interactions, and yeast two hybrid can be conducted for screening the plant protein that interacts with the fungal effector.

The story of an endophyte-plant interaction always has two sides. While this study has been focused on the endophyte side, a future analysis of the plant transcriptomic expression changes using the current RNA-seq data can be a good start for understanding the reaction on the plant side. However, a de novo approach needs to be designed for plant transcriptome analysis as there is no genome available for *Elymus* species or 'TACBOW0011' wheat.

5.5 Conclusion

E. bromicola AR3060 and AR3018 are two of the only three endophyte strains from a collection of over 100 strains that are able to cross the species barrier and infect a novel wheat host. They offer a great opportunity to better understand the mechanism underlying *Epichloë*-wheat compatibility. Transcriptome sequencing of the *E. bromicola*-symbiotic plants was used in this study to investigate gene expression changes in the fungal symbiont with different hosts.

RNA-seq of mature host plants inoculated with *E. bromicola* AR3060 showed an intensive induction of fungal gene expression *in planta*, including many uncharacterised secreted proteins. The most expressed fungal gene in every plant host was *gigA* that encodes a cyclic oligopeptide of unknown function that is present across many *Epichloë* species. In this study, the first *E. bromicola* effector mining pipeline was designed to identify candidate fungal genes that are involved in *Epichloë*-wheat compatibility using transcriptomic and genomic data, generating a list of 28 *E. bromicola* candidate effectors (EbCEs). Many abundant fungal transcripts *in-planta* are consistent with other *Epichloë* studies and also validated by the second RNA-seq experiment on progeny of *E. bromicola*-symbiotic plants. Combining two RNA-seq experiments, 27 effector candidates have been computationally identified and validated. Most of the candidates have no homology with any other characterised proteins, which is common for fungal effectors. The annotated candidates are a LysM domain-containing protein, and a predicted ribonuclease, pectinesterase, subtilisin-like protease, and chitinase, etc.

The key finding here is that *E. bromicola*, one of the *Epichloë* species that are generally considered mutualistic symbionts, have a set of putative effectors similar to fungal pathogens. Considering the high level of *in-planta* expression, dramatic expression changes between compatible and incompatible hosts, and the structures of some candidate effector proteins, those EbCEs are highly likely the key targets regulating a compatible symbiosis. This study has set a foundation for future functional characterisation of *Epichloë* effectors and it will be the first step towards the whole picture of the *Epichloë*-wheat compatibility.

6. References

Ali S, Laurie JD, Linning R, Cervantes-Chavez JA, Gaudet D, Bakkeren G (2014) An immunity-triggering effector from the Barley smut fungus *Ustilago hordei* resides in an Ustilaginaceae-specific cluster bearing signs of transposable element-assisted evolution. PLoS Pathog. 10: e1004223.

Akum FN, Steinbrenner J, Biedenkopf D, Imani J, Kogel KH (2015) The *Piriformospora indica* effector PIIN_08944 promotes the mutualistic Sebacinalean symbiosis. Front Plant Sci 6: 906.

Ambrose KV, Belanger FC (2012) SOLiD-SAGE of endophyte-symbiotic red fescue reveals numerous effects on host transcriptome and an abundance of highly expressed fungal secreted proteins. PLoS One 7: e53214.

Appels R, Eversole K, Feuillet C, Keller B, Rogers J, Stein N, Pozniak CJ, Choulet F, Distelfeld A, Poland J, Ronen G (2018) Shifting the limits in wheat research and breeding using a fully annotated reference genome. Science 361(6403): eaar7191.

Arachevaleta M, Bacon CW, Hoveland CS, Radcliffe DE (1989) Effect of the tall fescue endophyte on plant response to environmental stress. Agron J 81: 83–90.

Asai S, Shirasu K (2015) Plant cells under siege: plant immune system versus pathogen effectors. Curr Opin Plant Biol 28: 1–8.

Bacon CW, Lyons PC, Porter JK, Robbins JD (1986) Ergot toxicity from endophyte-symbiotic grasses: a review. Agron J 78: 106–116.

Ban T (2002) Physiological and ecological study of *Fusarium* head blight, and resistance in wheat (in Japanese). Plant Prot 56: 58–63.

Becker Y, Eaton CJ, Brasell E, May KJ, Becker M, Hassing B, Cartwright GM, Reinhold L, Scott B (2015) The fungal cell-wall integrity MAPK cascade is crucial for hyphal network formation and maintenance of restrictive growth of Epichloe festucae in symbiosis with Lolium perenne. MPMI 28(1): 69–85.

Bendtsen JD, Jensen LJ, Blom N, Von Heijne G, Brunak S (2004) Feature-based prediction of non-classical and leaderless protein secretion. Protein Eng Des Sel 17(4): 349–356.

Bouton JH, Latch GCM, Hill NS, Hoveland CS, McCann MA, Watson RH, Parish JA, Hawkins LL, Thompson FN (2002) Reinfection of tall fescue cultivars with non-ergot alkaloid-producing endophytes. Agronomie 94: 567–574.

Bultman TL, White JF (1988) “Pollination” of a fungus by a fly. Oecol 75: 317–319.

Bush LP, Wilkinson HH, Schardl CL (1997) Bioprotective alkaloids of grass fungal-endophyte symbioses. Plant Physiol 114: 1–7.

Clay K, Hardy TN, Hammond Jr AM (1985) Fungal endophytes of grasses and their effects on an insect herbivore. Oecol 66: 1–5.

Card SD, Faville MJ, Simpson WR, Johnson RD, Voisey CR, de Bonth AC, Hume DE (2014) Mutualistic fungal endophytes in the Triticeae - survey and description. FEMS Microbiol Ecol 88: 94–106.

Castroagudín VL, Moreira SI, Pereira DAS, Moreira SS, Brunner PC, Maciel JLN, Crous PW, McDonald BA, Alves E, Ceresini PC (2016) *Pyricularia graminis-tritici* sp. nov., a new *Pyricularia* species causing wheat blast. Persoonia 37: 199–216.

Christensen MJ, Bennett RJ, Schmid J (2002) Growth of *Epichloë* /*Neotyphodium* and p-endophytes in leaves of *Lolium* and *Festuca* grasses. Mycol Res 106: 93–106.

Christensen MJ, Bennett RJ, Ansari HA, Koga H, Johnson RD, Bryan GT, Simpson WR, Koolaard JP, Nickless EM, Voisey CR (2008) *Epichloë* endophytes grow by intercalary hyphal extension in elongating grass leaves. Fungal Genet Biol 45: 84–93.

Chung KR, Schardl CL (1997) Sexual cycle and horizontal transmission of the grass symbiont, *Epichloë typhina*. Mycol Res 101: 295–301.

Clay K (1990) Fungal endophytes of grasses. Annu Rev Ecol Syst 21: 275–298.

Clay K, Schardl C (2002) Evolutionary origins and ecological consequences of endophyte symbiosis with grasses. Am Nat 160: s99–s127.

Comby M, Gacoin M, Robineau M, Rabenoelina F, Ptas S, Dupont J, Profizi C, Baillieul F (2017) Screening of wheat endophytes as biological control agents against *Fusarium* head blight using two different in vitro tests. Microbiol Res 202: 11-20.

Cruz CD, Valent B (2017) Wheat blast disease: danger on the move. Trop Plant Pathol 42: 210–222.

Cui H, Tsuda K, Parker JE (2015) Effector-triggered immunity: from pathogen perception to robust defense. Annu Rev Plant Bio 66: 487–511.

de Hoon MJ, Imoto S, Nolan J, Miyano S (2004) Open source clustering software. Bioinformatics 20(9): 1453–1454.

de Jonge R, Bolton MD, Thomma BP (2011) How filamentous pathogens co-opt plants: the ins and outs of fungal effectors. Curr Opin Plant Biol 14: 400–406.

de Jonge R, van Esse HP, Kombrink A, Shinya T, Desaki Y, Bours R, van der Krol S, Shibuya N, Joosten MH, Thomma BP (2010) Conserved fungal LysM effector Ecp6 prevents chitin-triggered immunity in plants. Science 329: 953–955.

Dewey DR (1967) Synthetic hybrids of *Agropyron scribneri* × *Elymus juncea*. Bull Torrey Bot Club 94: 388–395.

Dewey DR (1971) Synthetic hybrids of *Hordeum bogdanii* with *Elymus canadensis* and *Sitanion hystrix*. Am J Bot 58: 902–908.

Dobin A, Gingeras TR (2015) Mapping RNA-seq reads with STAR. Curr Protoc Bioinform 51:11.14.1–11.14.19.

Dodds PN, Rathjen JP (2010) Plant immunity: towards an integrated view of plant-pathogen interactions. Nat Rev 11: 539–548.

Du Z, Zhou X, Ling Y, Zhang Z, Su Z (2010) agriGO: a GO analysis toolkit for the agricultural community. Nucleic Acids Res 38(Web Server issue): W64–W70.

Eaton CJ, Cox MP, Ambrose B, Becker M, Hesse U, Schardl CL, Scott B (2010) Disruption of signaling in a fungal-grass symbiosis leads to pathogenesis. Plant Physiol 153: 1780–1794.

Eaton CJ, Dupont PY, Solomon P, Clayton W, Scott B, Cox MP (2015) A core gene set describes the molecular basis of mutualism and antagonism in *Epichloë* spp. MPMI 28: 218–231.

Eisenhaber B, Bork P, Eisenhaber F (1998) Sequence properties of GPI-anchored proteins near the omega-site: constraints for the polypeptide binding site of the putative transamidase. Protein Eng 11(12): 1155-1161.

Elmi AA, West CP (1995) Endophyte infection effects on stomatal conductance, osmotic adjustment and drought recovery of tall fescue. New Phytol 131: 61–67.

El Baidouri M, Murat F, Veyssiere M, Molinier M, Flores R, Burlot L, Alaux M, Quesneville H, Pont C, Salse J (2017) Reconciling the evolutionary origin of bread wheat (*Triticum aestivum*). New Phytol 213: 1477-1486.

Emanuelsson O, Nielsen H, Brunak S, Von Heijne G (2000) Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. J Mol Biol 300(4): 1005–1016.

Fabro G, Steinbrenner J, Coates M, Ishaque N, Baxter L, Studholme DJ, Körner E, Allen RL, Piquerez SJ, Rougon-Cardoso A, Greenshields D (2011) Multiple candidate effectors from the oomycete pathogen *Hyaloperonospora arabidopsidis* suppress host plant immunity. PLoS Pathog 7(11): e1002348.

Figueroa M, Hammond-Kosack KE, Solomon PS (2018) A review of wheat diseases—a field perspective. Mol Plant Pathol 19: 1523-1536.

Fletcher LR (1999) “Non-toxic" endophytes in ryegrass and their effect on livestock health and production. Grassl Res Pract 7: 133–139.

Garnica DP, Nemri A, Upadhyaya NM, Rathjen JP, Dodds PN (2014) The ins and outs of rust haustoria. PLoS Pathog 10: e1004329.

Gibriel HA, Thomma BP, Seidl MF (2016) The age of effectors: genome-based discovery and applications. Phytopathology 106(10): 1206–1212.

Giraldo MC, Valent B (2013) Filamentous plant pathogen effectors in action. Nat Rev Microbiol 11: 800–814.

Gout L, Fudal I, Kuhn ML, Blaise F, Eckert M, Cattolico L, Balesdent MH, Rouxel T (2006) Lost in the middle of nowhere: the *AvrLm1* avirulence gene of the Dothideomycete *Leptosphaeria maculans*. Mol Microbiol 60: 67–80.

Gundel PE, Pérez LI, Helander M, Saikkonen K (2013) Symbiotically modified organisms: non-toxic fungal endophytes in grasses. Trends Plant Sci 18: 420–427.

Gururania MA, Venkatesh J, Upadhyaya CP, Nookaraju A, Pandey SK, Park SW (2012) Plant disease resistance genes: current status and future directions. Physiol Mol Plant Pathol 78: 51–65.

Gwinn KD, Gavin AM (1992) Relationship between endophyte infestation level of tall fescue seed lots and *Rhizoctonia zeae* seedling disease. Plant Dis 76: 911–914.

Hane JK, Williams AH, Taranto AP, Solomon PS, Oliver RP (2015) Repeat-induced point mutation: a fungal-specific, endogenous mutagenesis process. Genetic Transformation Systems in Fungi vol 2: 55-68.

Harris MO, Friesen TL, Xu SS, Chen MS, Giron D, Stuart JJ (2014) Pivoting from *Arabidopsis* to wheat to understand how agricultural plants integrate responses to biotic stress. J Exp Bot 66(2):513-31.

Hassing B, Winter D, Becker Y, Mesarich CH, Eaton CJ, Scott B (2019) Analysis of *Epichloë festucae* small secreted proteins in the interaction with *Lolium perenne*. PloS one 14(2).

Helfgott DM, Mason-Gamer RJ (2004) The evolution of North American *Elymus* (Triticeae, Poaceae) allotetraploids: evidence from phosphoenolpyruvate carboxylase gene sequences. Syst Bot 29: 850–861.

Huang S, Sirikhachornkit A, Su X, Faris J, Gill B, Haselkorn R, Gornicki P (2002) Genes encoding plastid acetyl-CoA carboxylase and 3-phosphoglycerate kinase of the *Triticum/Aegilops* complex and the evolutionary history of polyploid wheat. Proc Natl Acad Sci 99(12): 8133-8.

Iannone LJ, White JF Jr, Giussani LM, Cabral D, Novas MV (2011) Diversity and distribution of *Neotyphodium*-infected grasses in Argentina. Mycol Prog 10: 9–19.

Jacobs S, Zechmann B, Molitor A, Trujillo M, Petutschnig E, Lipka V, Kogel KH, Schäfer P (2011) Broad-spectrum suppression of innate immunity is required for colonization of Arabidopsis roots by the fungus *Piriformospora indica*. Plant Physiol 156: 726–740.

Jensen KB (1990) Cytology and taxonomy of *Elymus kengii*, *E. grandiglumis*, *E. alatavicus* and *E. batalinii* (Triticeae: Poaceae). Genome 33: 668–673.

Johnson LJ, Johnson RD, Schardl CL, Panaccione DG (2003) Identification of differentially expressed genes in the mutualistic association of tall fescue with *Neotyphodium coenophialum*. Physiol Mol Plant Pathol 63 (6): 305–317.

Johnson LJ, de Bonth ACM, Briggs LR, Caradus JR, Finch SC, Fleetwood DJ, Fletcher LR, Hume DE, Johnson RD, Popay AJ, Tapper BA, Simpson WR, Voisey CR, Card SD (2013a) The exploitation of epichloae endophytes for agricultural benefit. Fungal Divers 60: 171–188.

Johnson LJ, Koulman A, Christensen M, Lane GA, Fraser K, Forester N, Johnson RD, Bryan GT, Rasmussen S (2013b) An extracellular siderophore is required to maintain the mutualistic interaction of *Epichloë festucae* with *Lolium perenne*. PLoS Pathog 9: e1003332.

Johnson RD, Lane GA, Koulman A, Cao M, Fraser K, Fleetwood DJ, Voisey CR, Dyer JM, Pratt J, Christensen M, Simpson WR (2015) A novel family of cyclic oligopeptides derived from ribosomal peptide synthesis of an in planta-induced gene, *gigA*, in *Epichloë* endophytes of grasses. Fungal Genet Biol 85: 14–24.

Jones DA, Bertazzoni S, Turo CJ, Syme RA, Hane JK (2018) Bioinformatic prediction of plant–pathogenicity effector proteins of fungi. Curr Opin Microbiol 46: 43-9.

Jones JD, Dangl JL (2006) The plant immune system. Nature 444: 323–329.

Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, Pesseat S. (2014) InterProScan 5: genome-scale protein function classification. Bioinformatics 30(9): 1236–1240.

Kannadan S, Rudgers JA (2008) Endophyte symbiosis benefits a rare grass under low water availability. Funct Ecol 22: 706–713.

Kirby EJ (1961) Host-parasite relations in the choke disease of grasses. Trans Brit Mycol Soc 44: 493–503.

Kamoun S (2006) A catalogue of the effector secretome of plant pathogenic oomycetes. Ann Rev Phytopathol 44: 41–60.

Kamoun S (2007) Groovy times: filamentous pathogen effectors revealed. Curr Opin in Plant Biol 10: 358–365.

Kishii M, Yamada T, Sasakuma T, Tsujimoto H (2004) Production of wheat–*Leymus racemosus* chromosome addition lines. Theor Appl Genet 109(2): 255-60.

Kloppholz S, Kuhn H, Requena N (2011) A secreted fungal effector of *Glomus intraradices* promotes symbiotic biotrophy. Curr Biol 21: 1204–9.

Lacadena J, Álvarez-García E, Carreras-Sangrà N, Herrero-Galán E, Alegre-Cebollada J, Garcia-Ortega L, Oñaderra M, Gavilanes JG, Martínez del Pozo Á (2007) Fungal ribotoxins: molecular dissection of a family of natural killers. FEMS Microbiol Rev 31(2): 212-37.

Larran S, Simon MR, Moreno MV, Siurana MS, Perelló A (2016) Endophytes from wheat as biocontrol agents against tan spot disease. Biological control 92: 17-23.

Leuchtmann A. 1992. Systematics, distribution and host specificity of grass endophytes. Nat Toxins 1:150–162.

Leuchtmann A, Clay K (1993) Nonreciprocal compatibility between *Epichloë typhina* and four host grasses. Mycologia 85: 157–163.

Leuchtmann A, Bacon CW, Schardl CL, White JF, Tadych M (2014) Nomenclatural realignment of *Neotyphodium* species with genus *Epichloë*. Mycologia 106(2):202–15.

Lievens B, Houterman PM, Rep M (2009) Effector gene screening allows unambiguous identification of *Fusarium oxysporum* f. sp. *lycopersici* races and discrimination from other *formae speciales*. FEMS Microbiol Lett 300:201-215.

Liu Q, Ge S, Tang H, Zhang X, Zhu G, Lu BR (2006) Phylogenetic relationships in *Elymus* (Poaceae: Triticeae) based on the nuclear ribosomal internal transcribed spacer and chloroplast *trnL-F* sequences. New Phytol 170: 411–420.

Liu T, Song T, Zhang X, Yuan H, Su L, Li W, Xu J, Liu S, Chen L, Chen T, Zhang M (2014) Unconventionally secreted effectors of two filamentous pathogens target plant salicylate biosynthesis. Nat Commun 5: 4686.

Lo Presti L, Lanver D, Schweizer G, Tanaka S, Liang L, Tollot M, Zuccaro A, Reissmann S, Kahmann R (2015) Fungal effectors and plant susceptibility. Annu Rev Plant Biol 66: 513–545.

Löve Á (1984) Conspectus of the Triticeae. Feddes Repert 95: 425–521.

Lu BR (1994) The genus *Elymus* L. in Asia. Taxonomy and biosystematics with special reference to genomic relationships. In: Wang RRC, Jensen KB, Jaussi C, eds. Proceedings of the 2nd International Triticeae Symposium, Logan, UT, USA. Logan, UT, USA: The Utah State University Press 219–233.

Ma LJ, Van Der Does HC, Borkovich KA, Coleman JJ, Daboussi MJ, Di Pietro A, Dufresne M, Freitag M, Grabherr M, Henrissat B, Houterman PM (2010) Comparative genomics reveals mobile pathogenicity chromosomes in *Fusarium*. Nature 464(7287): 367–373.

Ma L, Lukasik E, Gawehns F, Takken FL (2012) The use of agroinfiltration for transient expression of plant resistance and fungal effector proteins in *Nicotiana benthamiana* leaves. Plant Fungal Pathogens: Methods and Protocols 61–74.

Malinowski DP, Belesky DP (2000) Adaptations of endophyte-symbiotic cool-season grasses to environmental stresses: mechanisms of drought and mineral stress tolerance. Crop Sci 40: 923–940.

Martin F, Aerts A, Ahrén D, Brun A, Danchin EG, Duchaussoy F, Gibon J, Kohler A, Lindquist E, Pereda V, Salamov A (2008) The genome of *Laccaria bicolor* provides insights into mycorrhizal symbiosis. Nature 452: 88–92.

Martin F, Kohler A, Murat C, Balestrini R, Coutinho PM, Jaillon O, Montanini B, Morin E, Noel B, Percudani R, Porcel B (2010) Périgord black truffle genome uncovers evolutionary origins and mechanisms of symbiosis. Nature 464:1033–1038.

Mentlak TA, Kombrink A, Shinya T, Ryder LS, Otomo I, Saitoh H, Terauchi R, Nishizawa Y, Shibuya N, Thomma BP, Talbot NJ (2012) Effector-mediated suppression of chitin-triggered immunity by *Magnaporthe oryzae* is necessary for rice blast disease. Plant Cell 24: 322–335.

Mesarich CH, Griffiths SA, van der Burgt A, Ökmen B, Beenen HG, Etalo DW, Joosten MH, de Wit PJ (2014). Transcriptome sequencing uncovers the Avr5 avirulence gene of the tomato leaf mold pathogen *Cladosporium fulvum*. MPMI 27(8): 846–857.

Moon CD, Miles CO, Järlfors U, Schardl CL (2002) The evolutionary origins of three new *Neotyphodium* endophyte species from grasses indigenous to the Southern Hemisphere. Mycologia 94: 694–711.

Nagabhyru P, Dinkins RD, Schardl CL (2019) Transcriptomics of *Epichloë*-grass symbioses in host vegetative and reproductive stages. MPMI 32(2): 194-207.

Oberhofer M, Leuchtmann A (2012) Genetic diversity in epichloid endophytes of *Hordelymus europaeus* suggests repeated host jumps and interspecific hybridizations. Mol Ecol 21: 2713–2726.

Park CM, Bruenn JA, Ganesa C, Flurkey WF, Bozarth RF, Koltin Y (1994) Structure and heterologous expression of the Ustilago maydis viral toxin KP4. Mol Microbiol 11(1): 155–64.

Pennington HG, Jones R, Kwon S, Bonciani G, Thieron H, Chandler T, Luong P, Morgan SN, Przydacz M, Bozkurt T, Bowden S (2019) The fungal ribonuclease-like effector protein CSEP0064/BEC1054 represses plant immunity and interferes with degradation of host ribosomal RNA. PLoS pathogens 15(3): e1007620.

Petersen TN, Brunak S, von Heijne G, Nielsen H (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. Nature methods 8(10): 785–786.

Plett JM, Martin F (2011) Blurred boundaries: lifestyle lessons from ectomycorrhizal fungal genomes. Trends Genet 27: 14–22.

Plett JM, Kemppainen M, Kale SD, Kohler A, Legue V, Brun A, Tyler BM, Pardo AG, Martin F (2011) A secreted effector protein of *Laccaria bicolor* is required for symbiosis development. Curr Biol 21: 1197–1203.

Plett JM, Daguerre Y, Wittulsky S, Vayssières A, Deveau A, Melton SJ, Kohler A, Morrell-Falvey JL, Brun A, Veneault-Fourrey C, Martin F (2014) Effector MiSSP7 of the mutualistic fungus *Laccaria bicolor* stabilizes the Populus JAZ6 protein and represses jasmonic acid (JA) responsive genes. Proc. Natl Acad Sci USA 111: 8299–8304.

Plett JM, Martin F (2015) Reconsidering mutualistic plant–fungal interactions through the lens of effector biology. Curr Opin in Plant Biol 26: 45–50.

Qi D, Innes RW (2013) Recent advances in plant NLR structure, function, localization, and signaling. Front Immunol 4: 348.

Rasmussen S, Parsons AJ, Bassett S, Christensen MJ, Hume DE, Johnson LJ, Johnson RD, Simpson WR, Stacke C, Voisey CR, Xue H (2007) High nitrogen supply and carbohydrate content reduce fungal endophyte and alkaloid concentration in *Lolium perenne*. New Phytol 173(4): 787–797.

Robinson MD, McCarthy DJ and Smyth GK (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26(1): 139–140.

Saikkonen K, Young CA, Helander M, Schardl CL (2016) Endophytic *Epichloë* species and their grass hosts: from evolution to applications. Plant Mol Biol 90: 665–675.

Salzer P, Bonanomi A, Beyer K, Vögeli-Lange R, Aeschbacher RA, Lange J, Wiemken A, Kim D, Cook DR, Boller T (2007) Differential expression of eight chitinase genes in *Medicago truncatula* roots during mycorrhiza formation, nodulation, and pathogen infection. MPMI 13(7): 763-77.

Sánchez-Vallet A, Saleem-Batcha R, Kombrink A, Hansen G, Valkenburg DJ, Thomma BP, Mesters JR (2013) Fungal effector Ecp6 outcompetes host immune receptor for chitin binding through intrachain LysM dimerization. Elife 2: e00790.

Saunders DG, Win J, Cano LM, Szabo LJ, 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.

Selin C, de Kievit TR, Belmonte MF, Fernando WD (2016) Elucidating the role of effectors in plant-fungal interactions: Progress and Challenges. Front Microbiol 7: 600.

Schardl CL, Leuchtmann A, Spiering MJ (2004) Symbioses of grass with seedborne fungal endophytes. Annu Rev Plant Biol 55: 315–340.

Schardl CL, Grossman RB, Nagabhyru P, Faulkner JR, Mallik UP (2007) Loline alkaloids: currencies of mutualism. Phytochem 68: 980–996.

Schardl CL, Balestrini R, Florea S, Zhang D, Scott B. (2009) *Epichloë* endophytes: clavicipitaceous symbionts of grasses. In Plant Relationships (pp. 275–306). Springer Berlin Heidelberg.

Schardl CL (2010) The Epichloae, symbionts of the grass subfamily Poöideae. Ann Mo Bot Gard 97: 646–665.

Schardl CL, Young CA, Hesse U, Amyotte SG, Andreeva K, Calie PJ, Fleetwood DJ, Haws DC, Moore N, Oeser B, Panaccione DG, Schweri KK, Voisey CR, Farman ML, Jaromczyk JW, Roe BA, O'Sullivan DM, Scott B, Tudzynski P, An Z, Arnaoudova EG, Bullock CT, Charlton ND, Chen L, Cox M, Dinkins RD, Florea S, Glenn AE, Gordon A, Güldener U, Harris DR, Hollin W, Jaromczyk J, Johnson RD, Khan AK, Leistner E, Leuchtmann A, Li C, Liu J, Liu J, Liu M, Mace W, Machado C, Nagabhyru P, Pan J, Schmid J, Sugawara K, Steiner U, Takach J, Tanaka E, Webb JS, Wilson EV, Wiseman JL, Yoshida R, Zeng Z. (2013). Plant-symbiotic fungi as chemical engineers: multi-genome analysis of the Clavicipitaceae reveals dynamics of alkaloid loci. PLoS Genet 9(2): e1003323.

Schmid J, Day R, Zhang N, Dupont PY, Cox MP, Schardl CL, Minards N, Truglio M, Moore N, Harris DR, Zhou Y (2016) Host tissue environment directs activities of an *Epichloë* endophyte, while it induces systemic hormone and defense responses in its native perennial ryegrass host. MPMI 30(2): 138-49.

Scott B, Green K, Berry D. The fine balance between mutualism and antagonism in the *Epichloë festucae*–grass symbiotic interaction (2018) Curr Opin Plant Biol 44: 32-8.

Shewry PR, Hey SJ (2015) The contribution of wheat to human diet and health. Food Energy Secur 4(3): 178-202.

Simpson WR, Faville MJ, Moraga RA, Williams WM, Mcmanus MT, Johnson RD (2014) *Epichloë* fungal endophytes and the formation of synthetic symbioses in Hordeeae (= Triticeae) grasses. J Syst Evol 52(6): 794–806.

Sonah H, Deshmukh RK, Bélanger RR (2016) Computational prediction of effector proteins in fungi: opportunities and challenges. Front Plant Sci 7: 126.

Song H, Nan Z (2015) Origin, divergence, and phylogeny of asexual *Epichloë* endophyte in *Elymus* species from western China. PLoS One 10: e0127096.

Song H, Song QY, Li XZ, Nan ZB (2015) Are *Epichloë* endophytes specific to *Elymus* grass hosts? Genet Mol Res 14: 17463–71.

Song H, Nan Z, Song Q, Xia C, Li X, Yao X, Xu W, Kuang Y, Tian P, Zhang Q (2016) Advances in research on *Epichloë* endophytes in Chinese native grasses. Front Microbiol. 7: 1399.

Söding J, Biegert A, Lupas AN (2005) The HHpred interactive server for protein homology detection and structure prediction. Nucleic Acids Res 33(suppl 2): W244–W248.

Spatafora JW, Sung GH, Sung JM, Hywel-Jones NL, White JF (2007). Phylogenetic evidence for an animal pathogen origin of ergot and the grass endophytes. Mol Ecol 16(8): 1701–1711.

Sperschneider J, Dodds PN, Gardiner DM, Manners JM, Singh KB, Taylor JM (2015) Advances and challenges in computational prediction of effectors from plant pathogenic fungi. PLoS Pathog, 11(5): e1004806.

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

Sperschneider J, Catanzariti AM, DeBoer K, Petre B, Gardiner DM, Singh KB, Dodds PN, Taylor JM (2017) LOCALIZER: subcellular localization prediction of both plant and effector proteins in the plant cell. Sci Rep 7: 44598.

Sperschneider J, Dodds PN, Singh KB, Taylor JM (2018) ApoplastP: prediction of effectors and plant proteins in the apoplast using machine learning. New Phytol 217: 1764-1778.

Sperschneider J, Dodds PN, Gardiner DM, Singh KB, Taylor JM (2018) Improved prediction of fungal effector proteins from secretomes with EffectorP 2.0. Mol Plant Pathol 19: 2094-2110.

Stergiopoulos I, de Wit PJ (2009) Fungal effector proteins. Ann Rev Phytopathol 47: 233–263.

Strömberg CA (2011) Evolution of grasses and grassland ecosystems. Annu Rev Earth Planet Sci 39: 517–544.

Stotz HU, Mitrousia GK, de Wit PJ, Fitt BD (2014) Effector-triggered defence against apoplastic fungal pathogens. Trends Plant Sci 19: 491–500.

Supek F, Bošnjak M, Škunca N, Šmuc T (2011) REVIGO summarizes and visualizes long lists of gene ontology terms. PloS one 6(7): e21800.

Tanaka A, Christensen MJ, Takemoto D, Park P, Scott B (2006) Reactive oxygen species play a role in regulating a fungus-perennial ryegrass mutualistic association. Plant Cell 18: 1052–1066.

Tanaka A, Cartwright GM, Saikia S, Kayano Y, Takemoto D, Kato M, Tsuge T, Scott B (2013) ProA, a transcriptional regulator of fungal fruiting body development, regulates leaf hyphal network development in the *Epichloë festucae–Lolium perenne* symbiosis. Mol Microbiol 90: 551–568.

Tintjer T, Leuchtmann A, Clay K (2008) Variation in horizontal and vertical transmission of the endophyte *Epichloë elymi* infecting the grass *Elymus hystrix*. New Phytol 179: 236–246.

Thomas WJ, Thireault CA, Kimbrel JA, Chang JH (2009) Recombineering and stable integration of the *Pseudomonas syringae* pv. *syringae* 61 *hrp/hrc* cluster into the genome of the soil bacterium *Pseudomonas fluorescens* Pf0-1. The Plant J 60(5): 919–928.

Thomma BP, Seidl MF, Shi-Kunne X, Cook DE, Bolton MD, van Kan JA, Faino L (2016) Mind the gap; seven reasons to close fragmented genome assemblies. Fungal Genet Biol 90: 24–30.

Torabinejad J, Mueller RJ (1993) Genome constitution of the Australian hexaploid grass, *Elymus scabrus* (Poaceae: Triticeae). Genome 36: 147–151.

Torto TA, Li S, Styer A, Huitema E, Testa A, Gow NA, Van West P, Kamoun S (2003) EST mining and functional expression assays identify extracellular effector proteins from the plant pathogen *Phytophthora*. Genome Res 13(7): 1675-85.

van Esse HP, Bolton MD, Stergiopoulos I, de Wit PJ, Thomma BP (2007) The chitin-binding Cladosporium fulvum effector protein Avr4 is a virulence factor. Mol Plant-Microbe Interact 20: 1092-1101.

Voisey CR (2010) Intercalary growth in hyphae of filamentous fungi. Fungal Biol Rev 24: 123–131.

Voß S, Betz R, Heidt S, Corradi N, Requena N (2018) RiCRN1, a Crinkler Effector From the Arbuscular Mycorrhizal Fungus *Rhizophagus irregularis*, Functions in Arbuscule Development. Front Microbiol 9: 2068.

Wawra S, Fesel P, Widmer H, Timm M, Seibel J, Leson L, Kesseler L, Nostadt R, Hilbert M, Langen G, Zuccaro A (2016) The fungal-specific β-glucan-binding lectin FGB1 alters cell-wall composition and suppresses glucan-triggered immunity in plants. Nat Commun 7: 13188.

West CP, Izekor E, Turner KE, Elmi AA (1993) Endophyte effects on growth and persistence of tall fescue along a water-supply gradient. Agron J 85: 264–270.

White JF Jr., Morgan-Jones G, Morrow AC (1993) Taxonomy, life cycle, reproduction and detection of *Acremonium* endophytes. Agric Ecosyst Environ 44: 13–37.

Wilson D (1995) Endophyte—the evolution of a term, and clarification of its use and definition. Oikos 73: 274–276.

Win J, Chaparro-Garcia A, Belhaj K, Saunders DG, Yoshida K, Dong S, Schornack S, Zipfel C, Robatzek S, Hogenhout SA, Kamoun S (2012) Effector biology of plant-associated organisms: concepts and perspectives. Cold Spring Harb Symp Quant Biol 77: 235–247.

Ye J, Fang L, Zheng H, Zhang Y, Chen J, Zhang Z, Wang J, Li S, Li R, Bolund L, Wang J (2006) WEGO: a web tool for plotting GO annotations. Nucleic Acids Res 34(suppl 2): W293–W297.

Young CA, Bryant MK, Christensen MJ, Tapper BA, Bryan GT, Scott B (2005) Molecular cloning and genetic analysis of a symbiosis-expressed gene cluster for lolitrem biosynthesis from a mutualistic endophyte of perennial ryegrass. Mol Genet Genomics 274(1): 13–29.

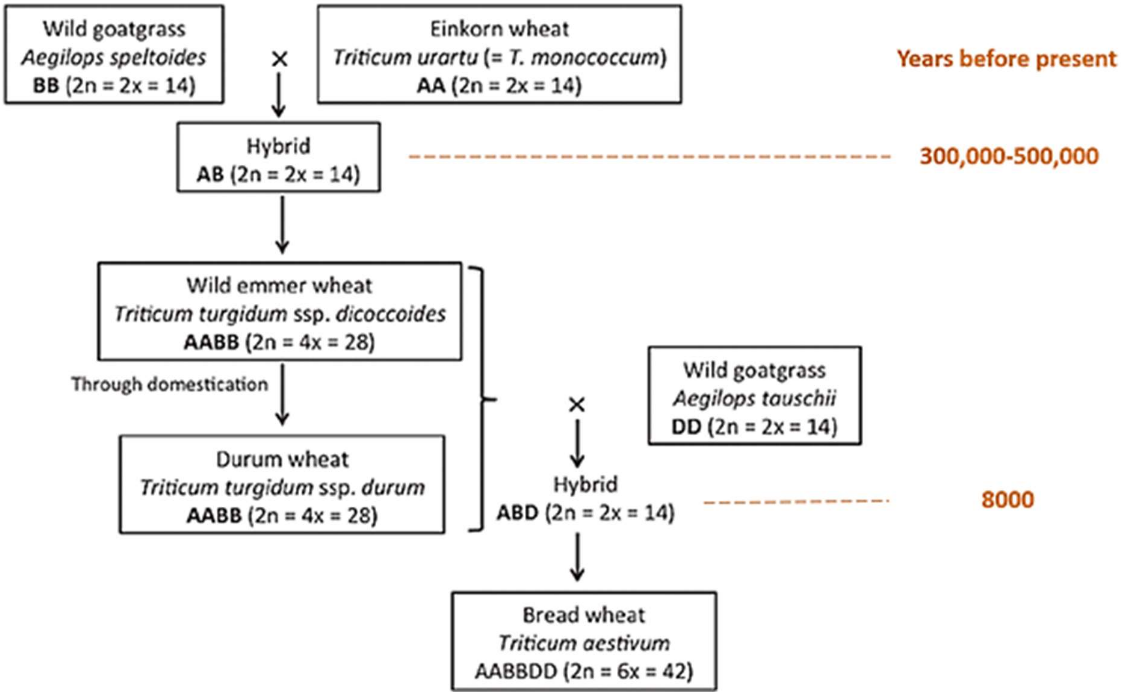
Zadoks JC, Chang TT, Konzak CF (1974) A decimal code for the growth stages of cereals. Weed Res 14(6): 415-21.

Zhang Y, Skolnick J (2005) TM-align: a protein structure alignment algorithm based on the TM-score. Nucleic Acids Res 33(7): 2302–2309.

Zhang Y (2008) I-TASSER server for protein 3D structure prediction. BMC bioinformatics 9(1): 40.

Zuccaro A, Lahrmann U, Güldener U, Langen G, Pfiffi S, Biedenkopf D, Wong P, Samans B, Grimm C, Basiewicz M, Murat C (2011) Endophytic life strategies decoded by genome and transcriptome analyses of the mutualistic root symbiont *Piriformospora indica*. PLoS Pathog 7: e1002290.

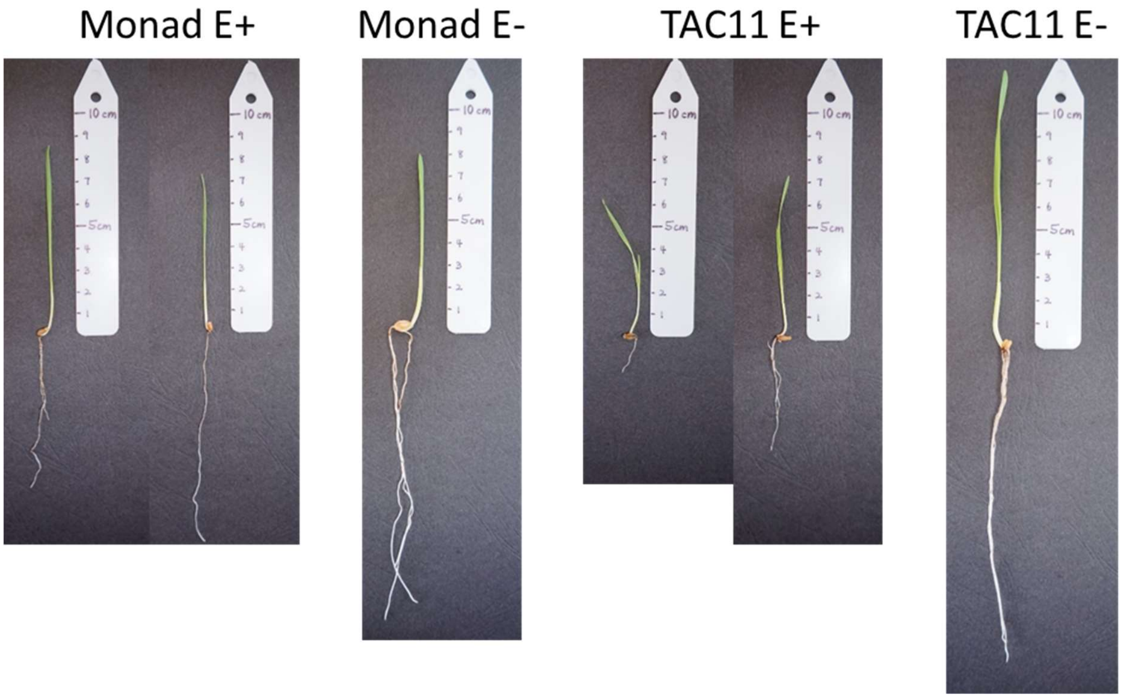
7. Appendices



Appendix 1. Schematic representation of the evolution of hexaploid bread wheat.

(based on Harris et al. 2014)

Appendix 2. Dendrogram of genetic similarity of endophytes isolated from 73 Triticeae accessions (Card et al 2014). AR3060 and AR3018 are marked with red rectangular. Unweighted pair group method of arithmetic averages was used.



Appendix 3. Examples of progeny seedlings of *E. bromicola*-symbiotic "Monad" and "TACBOW0011" wheat. Aerial parts of the seedlings at Zadoks decimal growth stage Z1.1 (first leaf emerged) were sampled for transcriptome analysis of progenies of *E. bromicola* strain AR3060 infected wheat (Chapter 4). Endophyte-symbiotic wheat seedlings showed weaker root growth. E+, endophyte-symbiotic; E-, endophyte-free; TAC11, "TACBOW0011" wheat.

Appendix 4. Summary of SOLiD-SAGE read alignments to reference genomes.

Plant[a]	Endohpyte[b]	Total reads	Not mapped	%	Unique	%	Endophyte[c]
Elymus	**AR3018**	28,528,588	20,774,513	73%	5,591,904	20%	2.55%
Elymus	**AR3018**	29,123,356	21,022,979	72%	5,610,854	19%	1.15%
Elymus	**AR3018**	31,882,138	23,412,797	73%	6,126,473	19%	1.92%
Elymus	**AR3018**	26,996,660	20,634,706	76%	4,636,279	17%	1.87%
Elymus	**AR3018**	26,596,374	19,400,191	73%	5,197,079	20%	2.64%
Elymus	**AR3018**	27,724,395	20,266,988	73%	5,184,907	19%	1.12%
Elymus	**AR3018**	35,390,951	26,519,779	75%	6,401,324	18%	1.83%
Elymus	**AR3018**	26,111,204	20,079,794	77%	4,386,059	17%	1.83%
Monad	**AR3018**	29,579,763	17,290,705	58%	9,228,625	31%	2.79%
Monad	**AR3018**	30,594,833	18,239,352	60%	9,350,262	31%	2.98%
Monad	**AR3018**	24,936,682	14,808,037	59%	7,553,627	30%	2.49%
Monad	**AR3018**	27,276,468	15,922,876	58%	8,557,023	31%	3.02%
Monad	**AR3018**	25,508,494	15,214,802	60%	7,900,305	31%	3.21%
Monad	**AR3018**	22,463,143	13,310,988	59%	6,875,949	31%	2.76%
Elymus	**AR3060**	24,647,942	17,941,939	73%	4,776,745	19%	1.18%
Elymus	**AR3060**	28,204,864	20,603,883	73%	5,284,979	19%	1.46%
Elymus	**AR3060**	30,601,779	21,836,242	71%	5,890,075	19%	3.03%
Elymus	**AR3060**	30,653,359	21,981,517	72%	5,796,375	19%	2.98%
Elymus	**AR3060**	20,404,207	14,840,127	73%	3,973,513	19%	1.19%
Elymus	**AR3060**	27,141,319	19,812,163	73%	5,096,585	19%	1.48%
Elymus	**AR3060**	25,238,747	18,007,738	71%	4,876,465	19%	3.19%
Elymus	**AR3060**	19,412,283	13,876,583	71%	3,727,506	19%	3.14%
Monad	**AR3060**	29,814,253	17,719,224	59%	9,094,965	31%	3.14%
Monad	**AR3060**	29,959,178	18,071,300	60%	8,920,493	30%	3.46%
Monad	**AR3060**	24,573,906	14,334,872	58%	7,872,564	32%	2.96%
Monad	**AR3060**	21,967,501	12,888,268	59%	6,804,113	31%	2.75%
Monad	**AR3060**	32,202,010	19,105,121	59%	9,885,358	31%	3.39%
Monad	**AR3060**	25,501,276	15,317,324	60%	7,665,610	30%	3.72%
Monad	**AR3060**	21,318,961	12,449,345	58%	6,843,133	32%	3.10%
Monad	**AR3060**	21,511,847	12,543,290	58%	6,769,261	31%	3.03%
Elymus	**E-**	26,596,788	19,092,563	72%	5,031,480	19%	2.52%
Elymus	**E-**	26,604,604	19,018,949	71%	5,073,897	19%	2.67%
Elymus	**E-**	22,653,457	16,262,414	72%	4,312,930	19%	2.65%
Elymus	**E-**	23,265,007	16,619,880	71%	4,460,641	19%	2.87%
Monad	**E-**	27,602,140	16,156,883	59%	8,631,144	31%	2.28%
Monad	**E-**	24,340,746	14,182,378	58%	7,693,522	32%	2.46%

a Elymus, native host of AR3060 (*E. excelsus*) and AR3018 (*E. dahuricus*); Monad, a commercial spring wheat.

b AR3060 and AR3018, *Epichloë bromicola* strains. E-, Endophyte-free plant.

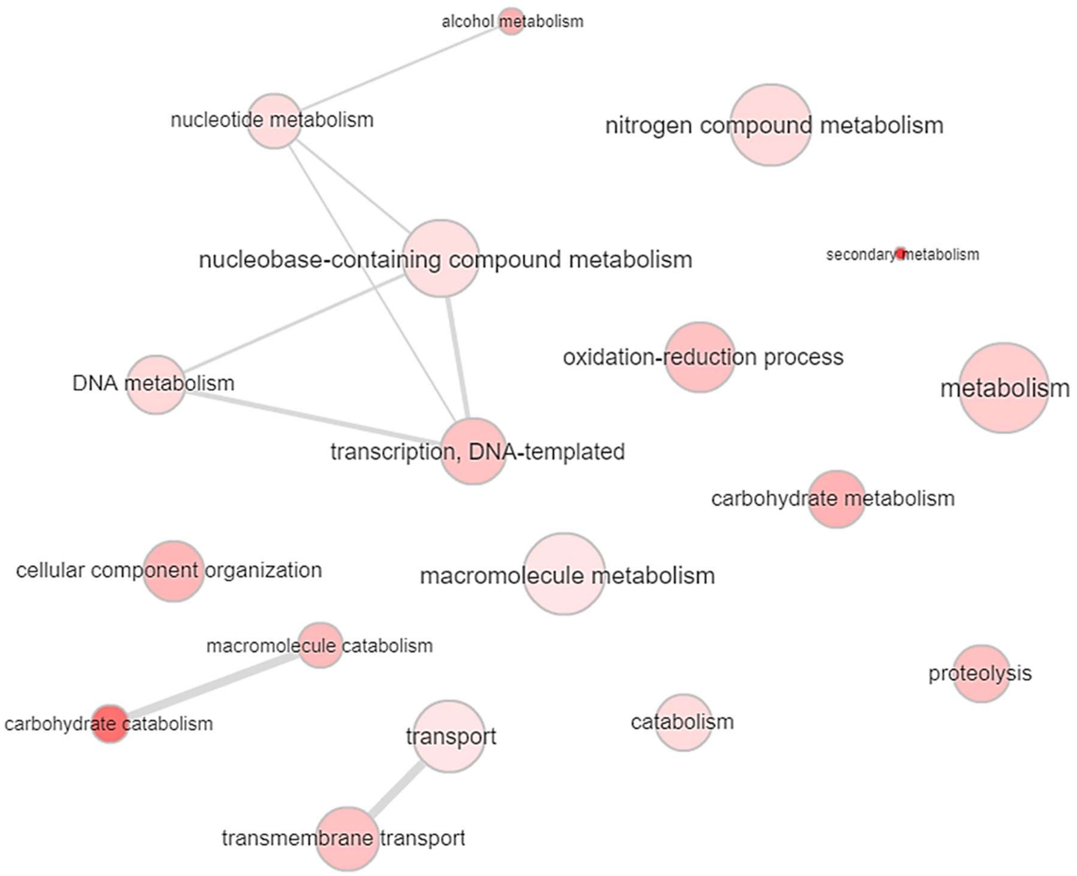
c Percentage of endophyte reads of total SOLiD reads.

Appendix 5. Top 10 significantly enriched GO terms of DEGs in pTAC3060 vs pEl3060.

GO term [a]	Description	Number[b]	p-value	FDR
GO:0019748	secondary metabolic process	5	0.00073	0.00073
GO:0016052	carbohydrate catabolic process	6	0.0057	0.0057
GO:0006066	alcohol metabolic process	5	0.099	0.099
GO:0017171	serine hydrolase activity	7	0.0017	0.0017
GO:0008236	serine-type peptidase activity	7	0.0017	0.0017
GO:0004252	serine-type endopeptidase activity	5	0.0052	0.0052
GO:0016798	hydrolase activity, acting on glycosyl bonds	9	0.0072	0.0072
GO:0004553	hydrolase activity, hydrolyzing O-glycosyl compounds	8	0.015	0.015
GO:0004175	endopeptidase activity	7	0.04	0.04
GO:0070011	peptidase activity, acting on L-amino acid peptides	9	0.065	0.065

a Gene ontology groups of Biological Process and Molecular Function.

b Number of genes of this GO term found in the list of differentially expressed genes (DEGs) in TAC11 vs *Elymus*.



Appendix 6. Fungal gene ontology groups of Biological Process differentially expressed *in planta*. REVIGO interactive map of GO terms under Biological Process category enriched in differentially expressed genes between AR3060-symbiotic "TAC11" wheat and *Elymus*. The sizes of the red circles indicate the relative frequencies of GO terms. Darker shading indicates lower P values. Highly similar GO terms are linked by lines with arbitrary length.

Appendix 7. Top 20 highly expressed fungal genes in at least one sample of RNA-seq I

		Normalised Counts [a]			Monad vs Elymus				
Gene	EbCEs	Culture	Monad	Elymus	Fold change	PValue	DEGs	Putative Function [b]	SP [c]
AR3060G0113041	N	1516	193	227	-1.17441	0.755686	N	Heat shock 70 kDa protein	N
AR3060G0104637	N	1469	3	10	-3.15268	0.165232	N	30 kDa heat shock protein	N
AR3060G0101861	N	1205	18	206	-11.4114	2.82E-05	Y	Cell wall mannoprotein 1	N
AR3060G0109514	N	925	14	20	-1.41367	0.598756	N	30 kDa heat shock protein	N
AR3060G0107109	N	883	14	22	-1.56583	0.47512	N	Heat shock protein hsp98	N
AR3060G0104197	N	883	1	5	-4.5162	0.281681	N		Y
AR3060G0112220	N	793	117	145	-1.24104	0.683874	N	Heat shock protein 90 homolog	N
AR3060G0103434	N	718	19	177	-9.29748	0.000123	Y	Clock-controlled protein 6	Y
AR3060G0100128	N	674	1	1	1.744852	1	N		N
AR3060G0105077	N	492	139	223	-1.6034	0.356157	N	Elongation factor 1-alpha	N
AR3060G0100103	N	456	0	1	-13.287	0.528324	N		N
AR3060G0100077	N	441	128	53	2.410408	0.095545	N		N
AR3060G0101448	N	390	48	69	-1.43938	0.502737	N		Y
AR3060G0106900	N	389	107	191	-1.78229	0.265965	N		Y
AR3060G0108909	N	386	47	11	4.152063	0.016339	Y		N
AR3060G0100250	N	381	0	0	0	1	N		N
AR3060G0112584	N	357	40	49	-1.21865	0.733419	N		N
AR3060G0104401	N	346	111	47	2.351323	0.106917	N		N
AR3060G0102970	N	309	142	120	1.188138	0.740824	N	Glyceraldehyde-3-phosphate dehydrogenase	N
AR3060G0115277	N	296	7	11	-1.60016	0.617378	N	Growth hormone-inducible transmembrane protein	N
AR3060G0114149	Y	17	7596	25865	-3.40502	0.016753	Y		Y
AR3060G0114255	Y	0	5042	5661	-1.12277	0.816333	N		Y
AR3060G0116318	Y	0	4924	2308	2.133378	0.132584	N		Y
AR3060G0105064	Y	0	4658	905	5.145997	0.001742	Y	LysM domain-containing protein ARB_05157	Y
AR3060G0117041	N	9	2342	2699	-1.15252	0.776238	N	Serine/threonine-protein kinase RAD53	N

Gene	EbCEs	Culture	Monad	Elymus	Fold change	PValue	DEGs	Putative Function	SP
AR3060G0117031	N	24	1374	1370	1.002742	0.996426	N		N
AR3060G0114485	Y	11	1201	298	4.022776	0.007278	Y	Endochitinase 1	Y
AR3060G0104034	Y	0	1083	104	10.35619	2.12E-05	Y		Y
AR3060G0110800	Y	9	1043	1922	-1.84288	0.224132	N	Glucan endo-1,6-beta-glucosidase B	Y
AR3060G0105601	Y	2	961	217	4.426564	0.004391	Y		Y
AR3060G0107873	Y	0	865	898	-1.03768	0.942346	N		Y
AR3060G0103728	Y	1	840	195	4.316746	0.005055	Y		Y
AR3060G0102050	Y	0	612	417	1.465948	0.447246	N		Y
AR3060G0112476	Y	5	454	44	10.18929	2.95E-05	Y		Y
AR3060G0106426	N	0	413	48	8.549642	9.35E-05	Y		Y
AR3060G0100801	N	0	383	3254	-8.494	7.46E-05	Y		N
AR3060G0105086	Y	10	332	138	2.409781	0.087383	N		N
AR3060G0108293	N	1	330	349	-1.05713	0.914505	N		N
AR3060G0106159A	Y	31	316	410	-1.29722	0.608315	N		Y
AR3060G0102234	N	245	315	186	1.688814	0.301543	N	Cell wall protein SED1 OS=Saccharomyces cerevisiae	Y
AR3060G0114870	Y	0	78	2143	-27.4445	1.73E-08	Y		Y
AR3060G0106335	Y	5	86	853	-9.90473	3.17E-05	Y	Cuticle-degrading protease	Y
AR3060G0114148	N	0	173	564	-3.2575	0.022554	Y		N
AR3060G0113595	Y	0	48	559	-11.6249	1.27E-05	Y		Y
AR3060G0108080	N	36	91	498	-5.46639	0.001473	Y	Glucose-repressible gene protein	N
AR3060G0106159	Y	0	41	355	-8.64928	0.00011	Y		Y
AR3060G0100356	N	0	108	316	-2.92097	0.039091	Y	Agroclavine dehydrogenase	N

a Normalised read counts of candidate genes in RNA-seq Experiment II. Culture: *Epichloë bromicola* AR3060 grown in axenic culture. *Elymus*: AR3060 symbiotic with native host *Elymus excelsus*. Monad: AR3060 with "Monad" wheat.

b Features are generated based on BLAST hits from databases (Swiss-prot, Uniprot, Refseq, Pfam) and HHpred search results.

c SP, signal peptide predicted by SignalP.

Appendix 8. Top 20 highly expressed fungal genes in at least one sample of RNA-seq II

		Normalised Counts [a]			TAC11 vs Elymus			
Gene	**EbCEs**	**Culture**	**Elymus**	**Tac11**	**Fold Change**	**adj.P.Val**	**Hit Description [b]**	**SP [c]**
AR3060G0101861	0	9602	1130	152	-9.15	0.30	Cell wall mannoprotein 1	N
AR3060G0113041	0	7089	3845	4845	1.04	1.00	Heat shock 70 kDa protein	N
AR3060G0104703	0	6382	383	122	-5.46	0.89		Y
AR3060G0112220	0	4916	2302	2580	-1.06	1.00	Heat shock protein 90 homolog	N
AR3060G0104253	0	4706	65	6	-17.28	0.89	Linear gramicidin synthase subunit D	N
AR3060G0105077	0	4517	2467	4836	1.62	0.89	Elongation factor 1-alpha	N
AR3060G0105408	0	4326	25	3	-1.90	0.98	Uncharacterized protein F14F9.5	Y
AR3060G0114395	0	3920	515	777	1.23	0.95	Probable glucan endo-1,3-beta-glucosidase eglC	Y
AR3060G0115760	0	3288	160	356	1.74	0.89	Ammonium transporter 1	N
AR3060G0104798	0	3250	851	1368	1.34	0.92		Y
AR3060G0107478	0	2964	351	123	-2.65	0.92		N
AR3060G0106158	0	2947	346	683	1.64	0.96	Mitochondrial phosphate carrier protein	N
AR3060G0103434	0	2904	1430	1256	-1.37	0.94	Clock-controlled protein 6	Y
AR3060G0102729	0	2877	48	0	-40.54	0.33		Y
AR3060G0104570	0	2826	703	1243	1.47	0.96	ADP,ATP carrier protein	N
AR3060G0101448	0	2742	1218	1393	-1.07	0.99		Y
AR3060G0116849	0	2645	1061	1832	1.43	0.89	Histone H3	N
AR3060G0111882	0	2546	2152	2214	-1.17	0.96	Glutamine synthetase	N
AR3060G0108179	0	2392	1497	1444	-1.26	0.96	Polyubiquitin	N
AR3060G0108080	0	2257	1823	750	-2.99	0.89	Glucose-repressible gene protein	N
AR3060G0114149	1	61	153970	75639	-2.57	0.89		Y
AR3060G0100801	0	0	48830	4589	-14.16	0.29		N
AR3060G0117041	0	0	44796	15598	-3.52	0.29	Serine/threonine-protein kinase RAD53	N

Gene	EbCEs	Culture	Elymus	Tac11	Fold Change	adj.P.Val	Hit Description	SP [c]
AR3060G0114255	1	27	44144	26468	-2.09	0.89		Y
AR3060G0110800	1	390	25905	15145	-2.07	0.89	Glucan endo-1,6-beta-glucosidase B	Y
AR3060G0116318	1	0	18748	2451	-8.96	0.29		Y
AR3060G0107873	1	12	12609	7571	-2.14	0.89		Y
AR3060G0105064	1	5	11648	33173	2.36	0.89	LysM domain-containing protein ARB_05157	Y
AR3060G0117032	0	36	8882	4542	-2.39	0.60		N
AR3060G0108293	0	78	6888	4471	-1.86	0.89		N
AR3060G0114870	1	0	6273	504	-16.71	0.59		Y
AR3060G0114148	0	4	6069	3287	-2.22	0.89		N
AR3060G0108288	0	25	5427	4061	-1.63	0.89		N
AR3060G0117033	0	39	4441	2292	-2.34	0.89		N
AR3060G0101652	1	160	4388	5142	-1.03	1.00	Alkali-sensitive linkage protein 1	Y
AR3060G0102050	1	0	4327	605	-9.03	0.29		Y
AR3060G0106161	0	566	4305	7992	1.54	0.89		N
AR3060G0104129	0	525	4081	3254	-1.55	0.90	Purine-cytosine permease fcyB	N
AR3060G0103267	0	577	3620	3475	-1.26	0.90	Protein disulfide-isomerase	Y
AR3060G0114485	1	1193	2875	14065	3.99	0.46	Endochitinase 1	Y
AR3060G0103728	1	42	2407	9684	3.30	0.56		Y
AR3060G0112476	1	1215	875	8848	8.26	0.29		Y
AR3060G0102234	0	1813	2490	7858	2.62	0.37	Cell wall protein SED1	Y
AR3060G0101890	1	2	2835	7566	2.30	0.77		Y
AR3060G0116257	0	575	2637	4269	1.35	0.90	78 kDa glucose-regulated protein homolog	Y
AR3060G0103195	0	439	2407	4250	1.48	0.89		Y

a Normalised read counts of candidate genes in RNA-seq Experiment II. Culture: *Epichloë bromicola* AR3060 grown in axenic culture. *Elymus*: AR3060 symbiotic with native host *Elymus excelsus*. TAC11: AR3060 with "TACBOW0011" wheat.

b Features are generated based on BLAST hits from databases (Swiss-prot, Uniprot, Refseq, Pfam) and HHpred search results.

c SP, signal peptide predicted by SignalP.