Genomic features and evolution of the conditionally dispensable chromosome in the tangerine pathotype of *Alternaria alternata*

Mingshuang Wang[1,2], Huilan Fu[1], Xing-Xing Shen[2], Ruoxin Ruan[3], Nicholas Pun[4], Jianping Xu[4], Hongye Li[1*] and Antonis Rokas[2*]

[1]Institute of Biotechnology, Zhejiang University, Hangzhou 310058, China

[2]Department of Biological Sciences, Vanderbilt University, Nashville, Tennessee 37235, USA

[3]Hangzhou Academy of Agricultural Sciences, Hangzhou 310024, China

[4]Department of Biology, McMaster University, Hamilton, Ontario L8S 4K1, Canada

*Corresponding authors: Hongye Li and Antonis Rokas

E-mail address: hyli@zju.edu.cn (H. Li); antonis.rokas@vanderbilt.edu (A. Rokas)

Abstract

The tangerine pathotype of the ascomycete fungus *Alternaria alternata* is the causal agent of citrus brown spot, which can result in significant losses of both yield and marketability for tangerines and tangerine hybrids worldwide. A conditionally dispensable chromosome (CDC), which harbors the host-selective ACT toxin gene cluster, is required for tangerine pathogenicity of *A. alternata*. To understand the genetic makeup and evolution of the tangerine pathotype CDC, we analyzed the function and evolution of the CDC genes present in the *A. alternata* Z7 strain. The 1.84Mb long CDC contains 512 predicted protein-coding genes, which are enriched in functional categories associated with 'metabolic process' (132 genes, p-value = 0.00192) including 'oxidation-reduction process' (48 genes, p-value = 0.00021) and 'lipid metabolic process' (11 genes, p-value = 0.04591). Relatively few of the CDC genes can be classified as CAZymes (13), kinases (3) and transporters (20). Differential transcriptome analysis of H_2O_2 treatment and control conditions revealed that 29 CDC genes were significantly up-regulated and 14 were significantly down-regulated, suggesting that CDC genes may play a role in coping with oxidative stress. Evolutionary analysis of the 512 CDC proteins showed that their evolutionary conservation tends to be restricted within the genus *Alternaria* and that the CDC genes evolve faster than genes in the essential chromosomes. Interestingly, phylogenetic analysis suggested that the genes of 13 enzymes and one sugar transporter residing in the CDC were likely horizontally transferred from distantly related species. Among these genes, 5 were likely transferred as a physically linked cluster of genes from *Cryptococcus* (Basidiomycota) or *Penicillium* (Eurotiomycetes) and another 4 genes might have been transferred from *Colletotrichum* (Sordariomycetes). One carboxylesterase gene was transferred from bacteria but functionally knocking out this gene did not affect the pathogenicity of the Z7 strain.

These results provide new insights into the function and evolution of CDC genes in *Alternaria*.

Keywords: plant pathogen, Dothideomycetes, accessory chromosome, evolutionary origin, horizontal gene transfer

Introduction

Alternaria alternata fungi can be ubiquitously found in soil, various plants, and decaying plant debris (Thomma, 2003). Some *A. alternata* strains, which are known as pathotypes, can cause plant diseases and result in severe crop losses worldwide. Specifically, at least seven pathogenic *A. alternata* pathotypes, each producing a unique host-selective toxin (HST) essential to pathogenicity, have been recognized to cause diseases in Japanese pear, strawberry, tangerine, apple, tomato, rough lemon and tobacco (Tsuge et al., 2013). Generally, genes required for HST biosynthesis are clustered on relatively small chromosomes of 1.0 ~ 2.0 Mb in size in *A. alternata* (Tsuge et al., 2013). These chromosomes, also known as conditionally dispensable chromosomes (CDCs), are highly variable among species and are required for pathogenicity in *A. alternata*. However, CDCs are generally not required for fungal growth and reproduction on artificial media (Johnson et al., 2001; Hatta et al., 2002).

The importance of CDCs conferring pathogenicity to fruits has been demonstrated by the construction of *A. alternata* hybrids in laboratory conditions. More specifically, two distinct laboratory hybrids were constructed from tomato and strawberry pathotypes and separately for apple and tomato pathotypes. The resulting hybrids harbored two CDCs from their parents allowing for the production of HSTs that caused diseases in both parental plant species (Akamatsu et al., 2001; Akagi et al., 2009b). Furthermore, these studies support the hypothesis that CDCs can be transmitted between different stains thereby facilitating the spread and evolutionary diversification of fungal phytopathogens.

Besides *A. alternata*, many other fungal phytopathogens have CDCs in their genomes including *Fusarium oxysporum*, *Nectria haematococca*, *Zymoseptoria tritici*

(previously known as *Mycosphaerella graminicola* and *Septoria tritici*), and *Colletotrichum gloeosporioides* (Coleman et al., 2009; Ma et al., 2010; Stukenbrock et al., 2010). Because CDCs are typically found in some, but not all, strains, they have been proposed to have different evolutionary origins than the essential chromosomes (Covert, 1998; Tsuge et al., 2013), e.g., through horizontal transfer from distantly related species (Covert, 1998; Hatta et al., 2002). The possibility of horizontal transfer of CDCs between species has been demonstrated in several studies. For example, a 2-Mb chromosome was transferred between two different biotypes of *C. gloeosporioides* during vegetative co-cultivation in the laboratory (He et al., 1998). Importantly, CDC acquisition can facilitate the transition from the non-pathogenic to pathogenic phenotype in the recipient organism. For example, co-incubation of non-pathogenic and pathogenic genotypes of *F. oxysporum* revealed that the transfer of a CDC from the pathogenic (donor) to non-pathogenic genotype (recipient) enabled the recipient to become pathogenic to tomatoes (Ma et al., 2010).

Previously, the ~1.0 Mb CDC of the tomato pathotype of *A. alternata* has been identified and characterized (Hu et al., 2012). Genes in that CDC are abundant in the categories of “metabolic process” and “biosynthetic process”, and include 36 polyketide and non-ribosomal peptide synthetase domain-containing genes; furthermore, the GC content of third codon positions, codon usage bias, and repeat region load in the CDC are different from that in the essential chromosomes (Hu et al., 2012). The authors also provided evidence supporting that the *A. arborescens* CDC was acquired through horizontal transfer from an unrelated fungus (Hu et al., 2012). Finally, a recent study claimed the availability of almost complete sequences of the CD chromosomes from the strawberry, apple and tomato pathotypes of *A.alternata* and the presence of large syntenic regions among the three CD

chromosomes, but did not provide any evidence (Tsuge et al., 2016).

The tangerine pathotype of *A. alternata*, which can cause citrus brown spot on tangerines and tangerine hybrids, was demonstrated to harbour an additional chromosome of about 1.9~2.0Mb by pulse-field gel electrophoresis studies (Miyamoto et al., 2008; Miyamoto et al., 2009). Seven genes, *ACTTR*, *ACTT2*, *ACTT3*, *ACTT5*, *ACTT6*, *ACTTS2* and *ACTTS3*, were located onto this chromosome and found to be required for the biosynthesis of the ACT-toxin, a unique HST produced by the tangerine pathotype (Tsuge et al., 2013). However, relatively little is known about the content, potential biological functions, and evolution of the other genes in the CDC of the tangerine pathotype of *A. alternata.* To address this question, we examined the functional annotation, transcriptional activity, and evolution of all 512 genes in the CDC of the tangerine pathotype strain Z7 of *A. alternata*. We found that these genes are enriched in functions associated with 'metabolic process'. Furthermore, 43 CDC genes were differentially expressed in response to oxidative stress. Finally, we found that conservation for the majority of the 512 Z7 CDC genes was restricted to the genus *Alternaria* and that 14 CDC genes likely originated via horizontal gene transfer (HGT) events from other fungi or bacteria. These results suggest that CDC genes in *Alternaria* are involved in processes associated with metabolism, that they are rapidly evolving, and that they sometimes originate via HGT.

Material and Methods

Genomic and transcriptomic data retrieval

The assembled *A. alternata* Z7 genome and proteome were downloaded from

GenBank under the accession number LPVP00000000 (Wang et al., 2016). Genome data from other *Alternaria* species were downloaded from the *Alternaria* genomes database (Dang et al., 2015) and from other species of Dothideomycetes from the GenBank database (Table S1) (last access on April 15, 2017). The transcriptome data of *A. alternata* after H_2O_2 treatment were downloaded from the NCBI's Sequence Read Archive (SRA) database with accession number SRP071688 (last access on May 8, 2017).

Identification of the contigs comprising the CDC

To identify all the contigs that are part of the *A. alternata* Z7 CDC, we used a previously described method with slight modifications (Hu et al., 2012). Briefly, all *A. alternata* Z7 contigs with sequence lengths greater than 5 kb were aligned to contigs from *Alternaria brassicicola*, a species that belongs to the same genus as *A. alternata* but which does not carry CDCs, using MUMmer 3.0 with an identity cut-off at 80% (Delcher et al., 2003). All contigs whose sequence coverage when aligned to *A. brassicicola* was less than 20% were considered to be parts of the *A. alternata* Z7 CDC.

Functional annotation of the genes in the CDC

To functionally annotate the 512 genes (and their protein products) in the CDC, we performed gene ontology analysis using the topGO version 2.28.0 (Alexa and Rahnenfuhrer, 2016) and classified the 512 proteins into protein families using the Pfam, version 31.0 databases (last access on June 1, 2016) (Finn et al., 2014). We predicted CDC genes that were parts of fungal secondary metabolite pathways using the web-based analytical tool SMURF (http://www.jcvi.org/smurf/index.php, last access on Jun 20, 2016) (Khaldi et al., 2010). To identify CDC proteins that are

carbohydrate-active enzymes, we searched with the CAZymes Analysis Toolkit (http://www.cazy.org/, last access on Jul 10, 2016) based on sequence and Pfam annotation (Lombard et al., 2014). To identify secreted proteins, we used SignalP, version 4.1 to predict transmembrane domains (Petersen et al., 2011) and we excluded non-extracellular and GPI-anchored proteins by using ProtComp, version 5 (Klee and Ellis, 2005) and fragAnchor (http://navet.ics.hawaii.edu/~fraganchor/NNHMM/NNHMM.html, last access on July 10, 2016) (Poisson et al., 2007), respectively. All resulting secreted proteins that were shorter than 200 amino acids (aa) in length and contained at least 4 cysteine residues were considered as small secreted cysteine-rich proteins. Kinases were searched and classified by an automated pipeline (Kosti et al., 2010). Cytochrome P450s were classified based on Pfam and the P450 database, version 1.2 (last access on June 18, 2016) (Moktali et al., 2012). Transporters were identified by performing BLASTp against the Transporter Classification Database (last access on June 7, 2016) (Saier et al., 2016).

Species phylogeny inference and evaluation of sequence conservation of CDC genes across Dothideomycetes

To understand the origin and evolution of genes in the *A. alternata* CDC, we first constructed a species phylogeny using genomic data from all available *Alternaria* species as well as for representative species from other Dothideomycetes. We constructed the species phylogeny using 1,754 conserved, fungal BUSCO genes as described previously (Simao et al., 2015; Shen et al., 2016). Briefly, the gene structure of each genome was predicted by AUGUSTUS 3.1 (Stanke and Waack, 2003), then the sequences of these predicted genes were aligned to the HMM alignment profile of each BUSCO gene in the OrthoDB v9 database (Simao et al.,

2015), and the ones with alignment bit-scores higher than 90% of the lowest bit-score among the reference genomes were kept for tree construction. The isolates *Alternaria porri BMP0178* and *Alternaria destruens BMP0317* were excluded from downstream analyses due to their poor (<60%) coverage of BUSCO genes (Table S2). We also compared the sequence similarity of each of the 512 CDC proteins to proteins in the genomes of other Dothideomycetes by calculating the value of BLAST identity * query coverage.

Examination of relative evolutionary rate between genes in the CDC and genes in the essential chromosomes

To estimate the relative evolutionary rate between Z7 CDC and essential chromosomes (ECs), ortholog groups within the *Alternaria* Clade I containing Z7 proteins from both CDC and ECs were extracted using OrthorMCL v2.0.9 and reciprocal BLASTp with identity>50% and query coverage >50% as cut-offs (Chen et al., 2006). For each ortholog group containing more than 7 (50% of total number) species, only one sequence per isolate/species (the one that was the best hit to the Z7 protein) was kept for further analysis. The coding sequences of those ortholog proteins were then aligned with MAFFT v7.023b using the E-INS-I strategy (Yamada et al., 2016) and trimmed with trimAl v1.4.rev11 using its automated1 strategy (Capella-Gutierrez et al., 2009). The maximum-likelihood (ML) phylogenetic trees were inferred using IQ-TREE 1.5.4, with the best model selected by ModelFinder (Nguyen et al., 2015; Kalyaanamoorthy et al., 2017) and with 1,000 bootstrap replicates. The significance of the difference in the average total branch length of the ML phylogenetic trees of CDC genes against that of the EC genes was determined by Wilcoxon test.

Identification of CDC genes that underwent HGT

To detect gene candidates that experienced HGT in *A.alternata* Z7 CDC, we first performed a BLASTp search of the local NCBI's nonredundant protein database (nr, last access on May 21, 2017) using Z7 CDC proteins as queries. We next selected proteins with the following characteristics as HGT candidates for further phylogenetic analyses: a) an Alien Index (AI) score larger than 0 (Gladyshev et al., 2008; Wisecaver et al., 2016), b) at least 80% of the top 200 BLASTp hits of the query protein are from organisms other than Dothideomycetes, and c) the sequence identity of the query protein across its entire length to its best BLASTp hit is equal or greater than 50%.

All genes that fit these three criteria were used as query sequences in BLASTp searches against the nr database and phylogenetic trees of their most closely related sequences across the tree of life were constructed. To reduce the number of sequences used to build each phylogenetic tree, we kept only one sequence per species (the one with the best BLASTp hit to the HGT candidate), then we selected the top 200 hits from the Blast results. The resulting sequences were used as input for multiple sequence alignment, trimming and phylogenetic inference, which were performed as described above.

The phylogenetic tree of each HGT candidate was manually inspected and only those trees that were evidently incongruent with the species phylogeny and strongly supported (bootstrap value > 95%) were retained as HGT candidates. For those HGT candidates, we used the Consel software, version V0.1i (Shimodaira and Hasegawa, 2001; Shimodaira, 2002) to perform the approximately unbiased (AU) comparative topology test between the unconstrained ML tree and the constrained ML tree in

which the *Alternaria* gene sequence was forced to be monophyletic with the rest of the sequences from Dothideomycetes. All phylogenetic trees were visualized using ITOL version 3.0 (Letunic and Bork, 2016).

Deletion of a horizontally transferred gene

The AALTg12037 gene was knocked out using a fungal protoplast transformation protocol, as described previously (Chen et al., 2017). Briefly, the two flanking 900bp fragments and a bacterial phosphotransferase B gene (HYG) were fused together, the resulting fragment was then introduced into fungal protoplasts using polyethylene glycol and $CaCl_2$. The transformants growing on a medium supplemented with 150 μg/ml hygromycin were selected and examined by PCR with specific primer pairs. All the primers used in this study are listed in Table S3. Fungal virulence was assessed on *Citrus poonensis* and *Citrus* × *clementina* leaves inoculated by placing a 5mm plug taken from the media for 2 days. Each strain was tested on at least 5 leaves and experiments were repeated two times.

Data availability

All data generated in this study, including CDC contigs, CDC gene annotation, multiple sequence alignments and phylogenetic trees, have been deposited on the figshare repository at DOI: 10.6084/m9.figshare.5549077 (the data will be made publicly available upon acceptance of the manuscript).

Results

General features of CDC

To identify the genome content of CDC of the tangerine pathotype of *A. alternata* strain Z7, we compared the genome sequence of the Z7 strain to that of *A. brassicicola*, which is known to not have a CDC (see methods). This strategy identified 43CDC contigs with a combined total length of 1.84Mb, which is close to the CDC size estimated by the pulse-field electrophoresis experiment (1.9~2.0Mb) (Miyamoto et al., 2008; Miyamoto et al., 2009). The overall G+C content of the CDC was 47.7%, while that of the ECs was 51.2%. The percentage of repetitive sequences on the Z7 CDC was 1.23%, over 2-fold of that of ECs (0.51%). The average gene length and gene density of CDC were significantly smaller than those of ECs (Table 1).

Functional annotation of the genes in the CDC

The CDC was predicted to comprise of 512 protein-coding genes. To predict their functions, gene ontology analysis was performed and 233 genes were assigned to 154 gene ontology terms related to biological process. 132 genes are assigned to the GO term metabolic process (p-value 0.00192) (Figure 1). Within the metabolic process, the GO terms oxidation-reduction process (48, p-value 0.00021) and lipid metabolic process (11, p-value 0.04591) were also significantly enriched (Figure 1). The reduction-oxidation (redox)-associated genes include monoxygenase, dehydrogenase, and reductase, suggesting that the CDC may be involved in intracellular redox homeostasis.

To functionally annotate the 512 genes in the CDC, we classified their protein products into protein families using several different approaches. Based on Pfam domain characterization, 307 / 512 genes belonged to 195 protein families (Table S4). We identified 13 CAZyme genes in the CDC, accounting for 3.48% (13 / 373) of the

total CAZymes of Z7. These include 3 Glycosyl Transferases (GTs), 6 Glycoside Hydrolases (GHs) and 5 Auxiliary Activities (AAs) (Table S4). A total of 29 secreted proteins were predicted in the CDC, including 3 plant cell wall-degrading enzymes, 10 small-secreted cysteine-rich proteins (SSCPs), 3 peptidases, 1 lipase, and 1 LysM domain-containing protein (Table S4). Only 3 kinases from 3 different kinase families were found in the CDC (Table S4). We identified 20 transcription factors in the CDC, which can be divided into 5 subfamilies: Zinc finger Zn_2-Cys_6 (8), Zinc finger C_2H_2 (5), Myb-like DNA-binding (1), helix-turn-helix, Psq (4) and high mobility group box (2) (Table S4). Twenty-six transporter-encoding genes were found in the CDC (Table S4). Among the 48 redox related genes in the CDC, 13 were predicted to be Cytochrome P450 monooxygenases. Based on comparisons with proteins in the PHI-database, 35 CDC proteins are related to pathogenicity (Table S4). Finally, the CDC contained the ACT-toxin biosynthetic gene cluster present only in the tangerine pathotype of *A. alternata*, which has been described in detail previously (Wang et al., 2016).

Expression of CDC genes under H_2O_2 stress

The production of the host-selective ACT toxin is crucial for the pathogenicity of the tangerine pathotype of *A. alternata* (Miyamoto et al., 2008; Miyamoto et al., 2009; Tsuge et al., 2013). Additionally, recent studies have shown that the ability to eliminate ROS by the tangerine pathotype of *A. alternata* is also of vital importance for pathogenesis to citrus (Lin et al., 2009; Chen et al., 2013; Yang et al., 2016). To discover which genes in the *A. alternata* Z7 CDC are potentially involved in coping with oxidative stress, we performed transcriptome analysis using the previously published transcriptome data of *A. alternata* Z7 after H_2O_2 treatment using no H_2O_2 treatment as a control (Wang et al., 2016).

Of the 512 CDC genes, 43 were significantly differentially expressed during the H_2O_2 stress condition (Table 2, Table S5). These included 29 CDC genes that were significantly upregulated and 14 that were significantly downregulated. The set of upregulated genes includes proteins such as oxidoreductases, hydrolases, transcription factors and phosphatases. Interestingly, the polyketide synthase (AALT_g11750, log2FC 3.3), which was hypothesized to be crucial for the biosynthesis of the ACT toxin (Miyamoto et al., 2010), was strongly induced after H_2O_2 treatment. A cluster of four genes (AALT_g11772, AALT_g11773, AALT_g11774 and AALT_g11775) located on CDC was also highly upregulated (log2FC from 3.4 to 4.4) during H_2O_2 stress. After comparing these four proteins to the Pfam database, proteins AALT_g11772 and AALT_g11774 were found to contain an oxidoreductase family domain and a NmrA-like family domain, respectively. However, no protein domain was predicted for AALT_g11773 and AALT_g11775 (Table 2, Table S5).

The Evolutionary origin of the Z7 CDC

To explore the evolutionary origin of the Z7 CDC, we compared the sequence similarity (calculated by BLAST identity score * query coverage) of each of the 512 CDC proteins to the protein sequences from the genomes of other species in the Dothideomycetes. In the species phylogeny, *Alternaria* species are grouped into three clades (I through III), which coincides with a previously constructed phylogeny based on 200 conserved single-copy orthologs (Wang et al., 2016). Proteins in the Z7 CDC showed a wide range of sequence similarity values to proteins in other Dothideomycetes (Figure 2). As expected, the highest degree of similarity was with proteins of other *Alternaria* species, and in particular with those from clade I (Figure 2, Figure S1). For example, we found that 442 / 512 (86.3%) of the Z7 CDC proteins

(442, 86.3%) showed >50% similarity to proteins in *A. turkisafria* (Figure 2, Figure S1), which can also cause citrus brown spot. However, *A. citriarbusti* and *A. tangelonis* can also cause citrus brown spot but these exhibited much lower numbers (321 / 512, 62.7% and 310 / 512, 60.5%, respectively) of proteins with >50% similarity to proteins in Z7 CDC (Figure 2). These results suggest that the gene content and sequence similarity on CDC can be highly variable among strains of the same pathotype.

To examine if any of the Z7 CDC proteins were more similar to proteins outside those found in genomes from the genus *Alternaria*, a BLASTp search of those 512 protein sequences against the NCBI non-redundant database was performed and the result was filtered using the E-value<1e-10 and sequence identity>30% criteria. Although the best matches for 330 / 512 proteins were proteins from other *Alternaria* species, 81 proteins had their best matches to be proteins found in non-*Alternaria* members from the family Pleosporaceae, 33 from the order Pleosporales (other than Pleosporaceae), 9 from the class Dothideomycetes (other than Pleosporales), 44 from the domain Fungi (other than Dothideomycetes), and 1 from the domain Bacteria (Table S6).

To further dissect the evolutionary history of Z7 CDC genes, we reconstructed the phylogenetic tree for each Z7 CDC gene with their orthologs from other species in the Dothideomycetes. There were too few orthologs to construct phylogenetic trees for three CDC genes; among the remaining 509 gene trees, 402 showed monophyly within the genus *Alternaria*. Taken together, our results suggest that most of the *A.alternata* Z7 CDC proteins were likely present in the *Alternaria* ancestor and that some of them were likely independently lost in some of the species during the

diversification of the genus.

Relative evolutionary rate of Z7 CDC

To figure out if genes in the CDC and genes in the ECs in Z7 evolve differently, we built a phylogenetic tree for each Z7 gene which has orthologs in more than 7 (50% of total number) species within Alternaria Clade I. A total of 191 CDC and 10,060 EC genes were used to calculate the average branch length for each tree. The average branch length for most genes in both the CDC and the ECs are very low (Figure 3), which indicates that most genes are highly conserved within the genus *Alternaria* Clade I. However, as a whole, we found that EC genes had lower average branch lengths than CDC genes (P= 2.2e-16, Figure 3), suggesting that the Z7 CDC genes are evolving faster than the Z7 EC genes.

HGT of genes in the CDC

To examine whether any of the *A. alternata* Z7 CDC genes originated via HGT, we calculated the Alien Index (Gladyshev et al., 2008; Wisecaver et al., 2016) of all 512 genes. A total of 43 genes show AI > 0 and at least 80% of their top 200 BLASTp hits with a taxonomic classification other than Dothideomycetes. The validity of these 43 HGT candidates was further examined phylogenetically. The phylogenetic trees for most of these 43 HGT candidates were weakly supported, but the evolutionary origin of 14 of these genes was strongly supported to be outside Dothideomycetes (Table 3, Fig S2-15). The AU test for each of the 14 genes significantly rejected the hypothesis that they formed a monophyletic group with the rest of the sequences from Dothideomycetes (Table 3). As the genes inferred to have undergone HGT are also found in other *Alternaria* species, we infer that the HGT events occurred before the divergence of the Z7 strain from the other *Alternaria* genomes examined and not

after. Specifically, 4 of the horizontally transferred genes are found in the genomes of species in *Alternaria* Clades I and II, 6 genes are also found in the genomes of species in *Alternaria* Clade I, and 4 genes are found only in the Japanese pear, strawberry and tangerine pathotypes (Table 3).

All HGT genes encode enzymes except for AALT_g11770, which is a hexose transporter (Table 3). According to their predicted functions, these HGT candidates are involved in two biological processes: oxidation reduction and carbohydrate metabolism (Table 3). Interestingly, five of the transferred genes are physically clustered in the Z7 strain CDC and almost always appear on the gene phylogeny as sisters to sequences from either *Cryptococcus* (Basidiomycota) or *Penicillium* (Eurotiomycetes) fungi (Figure 4A, Fig S2-6). Specifically, both *Penicillium flavigenum* and *Cryptococcus gattii* have 4 clustered genes that are highly similar and group together on the gene phylogeny with 4 of these 5 genes (Figure 4A, B); *Penicillium flavigenum* lacks the neuraminidase encoding gene homolog (AALT_g11771) while *Cryptococcus gattii* lacks the oxidoreductase encoding gene homolog (AALT_g11772) (Fig 4B). However, the gene order and orientation is quite different among the clusters of these three genomes (Figure 4B). From these results, the most likely scenario is that the genes were horizontally transferred from *Cryptococcus* or *Penicillium* species and were subsequently rearranged.

The ACT toxin is essential for the pathogenicity of *A. alternata* Z7 to citrus leaves, the synthesis of which is predicted to be controlled by a cluster composed of about 25 genes (Wang et al., 2016). Surprisingly, 4 / 25 of the ACT cluster genes contained in the Z7 CDC strain are always grouped together with sequences from the unrelated *Colletotrichum* (Sordariomycetes) in their gene phylogenies (Figure 5A, Fig

S7-10). Interestingly, the orthologous genes in *Colletotrichum tofieldiae* are physically linked with each other and are part of a secondary metabolite biosynthetic gene cluster predicted by antiSMASH 4.0 (Blin et al., 2017) (Figure 5A, B), although the gene order and orientation of the two clusters is different (Figure 5B). Besides *A.alternata* Z7 (the tangerine pathotype), this cluster is also present in the Japanese pear pathotype and the strawberry pathotype (Wang et al., 2016). Previously, deletion of the AKT3 gene in the Japanese pear pathotype, which is the ortholog of the HMG-CoA hydrolase gene ACTT3 (AALT_g11755), produced toxin-deficient and non-pathogenic mutants (Tanaka and Tsuge, 2000). Taken together, these results raise the hypothesis that the HGT of four genes from a lineage related to *Colletotrichum* may have contributed to the composition of the HST gene clusters found in the CDCs of three pathotypes of *Alternaria*.

There is only one gene (AALT_g12037) that was likely transferred from bacteria (Figure 6A, Fig S11). The phylogenetic tree of this gene contains a large proportion of bacteria (Figure 6A). There are only 15 fungal species predicted to contain this gene, including 13 *Alternaria* species, *Fusarium oxysporum* and *Pyrenochaeta sp. DS3sAY3a* (Figure 6B). AALT_g12037 is 1,428 bp in length and contains no introns. It encodes one of the carboxylesterases that are responsible for the hydrolysis of carboxylic acid esters into their corresponding acid and alcohol (Potter and Wadkins, 2006). This protein in Z7 CDC shows high sequence identity (75%, 73% and 74%) and query coverage (99%, 100% and 98%) to its orthologs in *Fusarium oxysporum*, *Pyrenochaeta sp. DS3sAY3a* and *Bacillus subtilis*, one of the potential donors. To test the functional significance of this gene, the coding region of AALT_g12037 was knocked out in *A.alternata Z7* (Figure 7A, B). However, the virulence of both the wild-type and the mutant did not differ (Figure 7C), indicating that this horizontally

434 transferred carboxylesterase is not associated with pathogenicity in *A.alternata Z7*.

435

436 **Discussion**

437 Conditionally dispensable chromosomes (CDCs) are commonly found in fungal
438 phytopathogens and play key roles in pathogenicity (Johnson et al., 2001; Hatta et al.,
439 2002). However, despite their importance, our knowledge about the gene content and
440 evolutionary origin of CDCs is very limited. In this study, we identified and
441 characterized the 1.84 Mb CDC of the tangerine pathotype strain Z7 of *A. alternata*
442 and examined the function and evolutionary history of its genes. Our results suggest
443 that CDC genes in *Alternaria* are involved in processes associated with metabolism,
444 that they are conserved within the genus *Alternaria* and are rapidly evolving, and that
445 they sometimes originate via HGT. Below, we discuss our findings in the context of
446 the genome content of *Alternaria* CDCs and the evolutionary origin of *Alternaria*
447 CDC genes.

448

449 In this study, we functionally annotated the CDC of the tangerine pathotype strain
450 Z7 of *A. alternata*. Although the Z7 CDC is much larger than the *A. arborescens*
451 CDC, several Z7 CDC gene families are comparable to those present in the *A.*
452 *arborescens* CDC, including CAZymes, SSCPs, kinases, transcription factors and
453 transporters (Table S4) (Hu et al., 2012). This result suggests that there are some
454 shared properties between CDCs from these two divergent species. Our analyses also
455 identified differences between CDCs in *A. arborescens* and in *A. alternata* strain Z7.
456 For example, the enrichment of “biosynthetic process” genes was discovered in the *A.*
457 *arborescens* CDC but not in the *A. alternata* CDC (Hu et al., 2012). In addition, the *A.*
458 *arborescens* CDC contains 10 SM clusters and harbors 36 polyketide and non-
459 ribosomal peptide synthetase genes that might be responsible for the biosynthesis of

460 the backbone structures of several groups of secondary metabolites (Hu et al., 2012).
461 However, in Z7 CDC, except for the host selective ACT-toxin gene cluster, no other
462 gene cluster involved in the biosynthesis of secondary metabolites was identified.
463 About 39.2% (201) of the 512 genes found in Z7 CDC were described as hypothetical
464 proteins according to their BLAST hits in the NCBI nr database, indicating that the
465 functions of a great number of Z7 CDC genes are not known (Table S4).
466
467 Evolutionary analysis showed that most of the Z7 CDC proteins are highly
468 conserved within the genus *Alternaria* and likely originated in the *Alternaria* last
469 common ancestor (Figure 2). These results contrast with those of a previous study
470 suggesting that the *A. arborescens* CDC originated from an unknown species through
471 HGT (Hu et al., 2012). The discrepancy is most likely due to the small amount of
472 available data used in the previous study; specifically, *A. arborescens* protein-coding
473 genes were compared to only two small databases respectively composed of either
474 only *A. brassicicola* proteins or of three fungal species from other filamentous fungal
475 genera (*Leptosphaeria maculans*, *Pyrenophora tritici-repentis*, and *Aspergillus*
476 *oryzae*) (Hu et al., 2012). In contrast, the much larger number of *Alternaria* genomes
477 examined in our study shows that the evolutionary history of most CDC genes is
478 consistent with the species phylogeny and with vertical, not horizontal, transmission.
479
480 Comparison of the Z7 CDC proteins to those in the proteomes of other
481 *Alternaria* species showed that CDCs in different species or even in the same
482 pathotype can be very variable (Figure 2). This result is consistent with the fact that
483 the length of CDCs in distinct pathotypes of *A.alternata* is very variable. For example,
484 both the strawberry and the tomato pathotypes contain a ~1.0 Mb CDC (Ito et al.,
485 2004; Hu et al., 2012); both the Japanese pear and the tangerine pathotypes contain a

1.9-2.0 Mb CDC (Miyamoto et al., 2008; Miyamoto et al., 2009); the apple pathotype contains a 1.1-1.8 Mb CDC (Harimoto et al., 2007); while the rough lemon pathotype contains a 1.2-1.5 Mb CDC (Masunaka et al., 2005). Although the mechanism underlying the formation of these diverse CDCs in *Alternaria* species is largely unknown, a recent population genomics study in *Zymoseptoria tritici* identified the precise breakpoint locations of insertions that give rise to the highly differentiated gene contents in *Z. tritici* CDCs (Croll et al., 2013). That same study also reported the occurrence of CDC losses in progeny because of nondisjunction during meiosis as well as the emergence of a new CDC through a fusion between sister chromatids (Croll et al., 2013). Thus, the CDCs of *Z. tritici* were proposed to originate mainly from ancient core chromosomes through a degeneration process involving breakage-fusion-bridge cycles and large insertions (Croll et al., 2013). Although determining whether the breakage-fusion-bridge cycles model holds for *Alternaria* CDCs will require further sequencing of CDCs and analyses, we note that our finding of evolutionary conservation of CDC genes within *Alternaria* is consistent with the *Z. tritici* model.

Horizontal transfer of CDCs between different strains of the same or closely related fungal species has been documented for *Fusarium oxysporum*, *Colletotrichum gloeosporioides*, as well as for some *Alternaria* species (He et al., 1998; Akamatsu et al., 2001; Akagi et al., 2009a; Ma et al., 2010). Horizontal transfer of entire CDCs may contribute not only to the introduction of CDCs in fungal populations but also to their subsequent spread and to their acquisition of virulence factors that are essential for pathogenicity. Since fungal CDCs are known to be able to transfer between closely related species, it may be very difficult to distinguish between vertical and horizontal transmission of CDCs within *Alternaria*. On the other hand, horizontal

transfer of CDCs between distantly related species has never been demonstrated, making it a less likely explanation for the evolutionary origin of *Alternaria* CDCs.

Although horizontal transfer is an unlikely explanation for the formation of CDCs in *Alternaria*, our study found evidence that HGT is a mechanism for the origin of some of the genes residing in its CDCs. Specifically, we found 14 genes in the Z7 CDC that were likely horizontally transferred from distantly related species, including 9 genes that formed 2 gene clusters (Table 3). Previously, the 23-gene secondary metabolic gene cluster involved in the biosynthesis of the mycotoxin sterigmatocystin was shown to have been horizontally transferred from *Aspergillus* to *Podospora* (Slot and Rokas, 2011). HGT of intact gene clusters would not only contribute to fungal metabolic diversity but also potentially provide its recipient with a competitive advantage offered by the ability to synthesize a novel secondary metabolite. Although the horizontally transferred gene cluster in this study contains fewer genes, the HMG-CoA hydrolase coding gene AKT3 found in one of the two Z7 CDC transferred clusters was shown to be absolutely required for the HST production and virulence to host plant (Tanaka and Tsuge, 2000). This is consistent with the view that HGT events, including ones involving the transfer of entire clusters, have played important roles over the course of the evolution of filamentous fungi (Fitzpatrick, 2012; Soanes and Richards, 2014; Wisecaver and Rokas, 2015).

One of the Z7 CDC genes, a carboxylesterase, was likely acquired from Bacteria (Figure 6A). Carboxylesterases are ubiquitous enzymes that exist in almost all living organisms and whose function is to hydrolyze carboxylesters into the corresponding carboxylic acid and alcohol (Satoh and Hosokawa, 1998). By degrading exogenous xenobiotics that contain esters, carboxylesterases are thought to be associated with

538 detoxification (Hatfield et al., 2016), although, to date, no endogenous
539 carboxylesterase substrates have been identified. Functional analysis of the
540 horizontally transferred carboxylesterase showed that this gene is not associated with
541 the pathogenicity of *A.alternata Z7* (Figure 7C). Thus, the acquisition of this gene
542 may not have been related to the transition to a pathogenic lifestyle but with nutrient
543 utilization and survival in nutritionally adverse environments.
544

545 **References**

546 Akagi, Y., Akamatsu, H., Otani, H., and Kodama, M. (2009a). Horizontal chromosome transfer, a
547 mechanism for the evolution and differentiation of a plant-pathogenic fungus. *Eukaryot Cell*
548 8(11)**,** 1732-1738.
549 Akagi, Y., Taga, M., Yamamoto, M., Tsuge, T., Fukumasa-Nakai, Y., Otani, H., et al. (2009b).
550 Chromosome constitution of hybrid strains constructed by protoplast fusion between the
551 tomato and strawberry pathotypes of Alternaria alternata. *Journal of General Plant Pathology*
552 75(2)**,** 101-109.
553 Akamatsu, H., Fukumasa-Nakai, Y., Otani, H., and Kodama, M. (2001). Construction and genetic
554 analysis of hybrid strains between apple and tomato pathotypes of Alternaria alternata by
555 protoplast fusion. *Journal of General Plant Pathology* 67(2)**,** 97-105.
556 Alexa, A., and Rahnenfuhrer, J. 2016. topGO: Enrichment Analysis for Gene Ontology. (R package
557 version 2.28.0).
558 Blin, K., Wolf, T., Chevrette, M.G., Lu, X., Schwalen, C.J., Kautsar, S.A., et al. (2017). antiSMASH
559 4.0-improvements in chemistry prediction and gene cluster boundary identification. *Nucleic*
560 *Acids Res* 28(10).
561 Capella-Gutierrez, S., Silla-Martinez, J.M., and Gabaldon, T. (2009). trimAl: a tool for automated
562 alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25(15)**,** 1972-1973.
563 Chen, F., Mackey, A.J., Stoeckert, C.J., Jr., and Roos, D.S. (2006). OrthoMCL-DB: querying a
564 comprehensive multi-species collection of ortholog groups. *Nucleic Acids Res* 34(Database
565 issue)**,** D363-368.
566 Chen, L.H., Lin, C.H., and Chung, K.R. (2013). A nonribosomal peptide synthetase mediates
567 siderophore production and virulence in the citrus fungal pathogen Alternaria alternata. *Mol*

Plant Pathol 14(5)**,** 497-505.

Chen, L.H., Tsai, H.C., Yu, P.L., and Chung, K.R. (2017). A Major Facilitator Superfamily Transporter-Mediated Resistance to Oxidative Stress and Fungicides Requires Yap1, Skn7, and MAP Kinases in the Citrus Fungal Pathogen Alternaria alternata. *PLoS One* 12(1).

Coleman, J.J., Rounsley, S.D., Rodriguez-Carres, M., Kuo, A., Wasmann, C.C., Grimwood, J., et al. (2009). The genome of Nectria haematococca: contribution of supernumerary chromosomes to gene expansion. *PLoS Genet* 5(8)**,** 28.

Covert, S.F. (1998). Supernumerary chromosomes in filamentous fungi. *Curr Genet* 33(5)**,** 311-319.

Croll, D., Zala, M., and McDonald, B.A. (2013). Breakage-fusion-bridge cycles and large insertions contribute to the rapid evolution of accessory chromosomes in a fungal pathogen. *PLoS Genet* 9(6)**,** 13.

Dang, H.X., Pryor, B., Peever, T., and Lawrence, C.B. (2015). The Alternaria genomes database: a comprehensive resource for a fungal genus comprised of saprophytes, plant pathogens, and allergenic species. *BMC Genomics* 16(239)**,** 015-1430.

Delcher, A.L., Salzberg, S.L., and Phillippy, A.M. (2003). Using MUMmer to identify similar regions in large sequence sets. *Curr Protoc Bioinformatics* 10(10).

Finn, R.D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R.Y., Eddy, S.R., et al. (2014). Pfam: the protein families database. *Nucleic Acids Res* 42(Database issue)**,** 27.

Fitzpatrick, D.A. (2012). Horizontal gene transfer in fungi. *FEMS Microbiol Lett* 329(1)**,** 1-8.

Gladyshev, E.A., Meselson, M., and Arkhipova, I.R. (2008). Massive horizontal gene transfer in bdelloid rotifers. *Science* 320(5880)**,** 1210-1213.

Harimoto, Y., Hatta, R., Kodama, M., Yamamoto, M., Otani, H., and Tsuge, T. (2007). Expression profiles of genes encoded by the supernumerary chromosome controlling AM-toxin biosynthesis and pathogenicity in the apple pathotype of Alternaria alternata. *Mol Plant Microbe Interact* 20(12)**,** 1463-1476.

Hatfield, M.J., Umans, R.A., Hyatt, J.L., Edwards, C.C., Wierdl, M., Tsurkan, L., et al. (2016). Carboxylesterases: General detoxifying enzymes. *Chem Biol Interact* 259(Pt B)**,** 327-331.

Hatta, R., Ito, K., Hosaki, Y., Tanaka, T., Tanaka, A., Yamamoto, M., et al. (2002). A conditionally dispensable chromosome controls host-specific pathogenicity in the fungal plant pathogen Alternaria alternata. *Genetics* 161(1)**,** 59-70.

He, C., Rusu, A.G., Poplawski, A.M., Irwin, J.A., and Manners, J.M. (1998). Transfer of a supernumerary chromosome between vegetatively incompatible biotypes of the fungus

Colletotrichum gloeosporioides. *Genetics* 150(4), 1459-1466.

Hu, J., Chen, C., Peever, T., Dang, H., Lawrence, C., and Mitchell, T. (2012). Genomic characterization of the conditionally dispensable chromosome in Alternaria arborescens provides evidence for horizontal gene transfer. *BMC Genomics* 13(171), 1471-2164.

Ito, K., Tanaka, T., Hatta, R., Yamamoto, M., Akimitsu, K., and Tsuge, T. (2004). Dissection of the host range of the fungal plant pathogen Alternaria alternata by modification of secondary metabolism. *Molecular Microbiology* 52(2), 399-411.

Johnson, L.J., Johnson, R.D., Akamatsu, H., Salamiah, A., Otani, H., Kohmoto, K., et al. (2001). Spontaneous loss of a conditionally dispensable chromosome from the Alternaria alternata apple pathotype leads to loss of toxin production and pathogenicity. *Curr Genet* 40(1), 65-72.

Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., von Haeseler, A., and Jermiin, L.S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods* 14(6), 587-589.

Khaldi, N., Seifuddin, F.T., Turner, G., Haft, D., Nierman, W.C., Wolfe, K.H., et al. (2010). SMURF: Genomic mapping of fungal secondary metabolite clusters. *Fungal Genet Biol* 47(9), 736-741.

Klee, E.W., and Ellis, L.B. (2005). Evaluating eukaryotic secreted protein prediction. *BMC Bioinformatics* 6, 256.

Kosti, I., Mandel-Gutfreund, Y., Glaser, F., and Horwitz, B.A. (2010). Comparative analysis of fungal protein kinases and associated domains. *BMC Genomics* 11(133), 1471-2164.

Letunic, I., and Bork, P. (2016). Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res* 44(W1), 19.

Lin, C.H., Yang, S.L., and Chung, K.R. (2009). The YAP1 homolog-mediated oxidative stress tolerance is crucial for pathogenicity of the necrotrophic fungus Alternaria alternata in citrus. *Mol Plant Microbe Interact* 22(8), 942-952.

Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P.M., and Henrissat, B. (2014). The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res* 42(Database issue), 21.

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

Masunaka, A., Ohtani, K., Peever, T.L., Timmer, L.W., Tsuge, T., Yamamoto, M., et al. (2005). An Isolate of Alternaria alternata That Is Pathogenic to Both Tangerines and Rough Lemon and

Produces Two Host-Selective Toxins, ACT- and ACR-Toxins. *Phytopathology* 95(3)**,** 241-247.

Miyamoto, Y., Ishii, Y., Honda, A., Masunaka, A., Tsuge, T., Yamamoto, M., et al. (2009). Function of genes encoding acyl-CoA synthetase and enoyl-CoA hydratase for host-selective ACT-toxin biosynthesis in the tangerine pathotype of Alternaria alternata. *Phytopathology* 99(4)**,** 369-377.

Miyamoto, Y., Masunaka, A., Tsuge, T., Yamamoto, M., Ohtani, K., Fukumoto, T., et al. (2008). Functional analysis of a multicopy host-selective ACT-toxin biosynthesis gene in the tangerine pathotype of Alternaria alternata using RNA silencing. *Mol Plant Microbe Interact* 21(12)**,** 1591-1599.

Miyamoto, Y., Masunaka, A., Tsuge, T., Yamamoto, M., Ohtani, K., Fukumoto, T., et al. (2010). ACTTS3 encoding a polyketide synthase is essential for the biosynthesis of ACT-toxin and pathogenicity in the tangerine pathotype of Alternaria alternata. *Mol Plant Microbe Interact* 23(4)**,** 406-414.

Moktali, V., Park, J., Fedorova-Abrams, N.D., Park, B., Choi, J., Lee, Y.-H., et al. (2012). Systematic and searchable classification of cytochrome P450 proteins encoded by fungal and oomycete genomes. *BMC Genomics* 13(525)**,** 1471-2164.

Nguyen, L.-T., Schmidt, H.A., von Haeseler, A., and Minh, B.Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* 32(1)**,** 268-274.

Petersen, T.N., Brunak, S., von Heijne, G., and Nielsen, H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods* 8(10)**,** 785-786. doi: doi: 10.1038/nmeth.1701.

Poisson, G., Chauve, C., Chen, X., and Bergeron, A. (2007). FragAnchor: a large-scale predictor of glycosylphosphatidylinositol anchors in eukaryote protein sequences by qualitative scoring. *Genomics Proteomics Bioinformatics* 5(2)**,** 121-130.

Potter, P.M., and Wadkins, R.M. (2006). Carboxylesterases--detoxifying enzymes and targets for drug therapy. *Curr Med Chem* 13(9)**,** 1045-1054.

Saier, M.H., Jr., Reddy, V.S., Tsu, B.V., Ahmed, M.S., Li, C., and Moreno-Hagelsieb, G. (2016). The Transporter Classification Database (TCDB): recent advances. *Nucleic Acids Res* 44(D1)**,** 5.

Satoh, T., and Hosokawa, M. (1998). The mammalian carboxylesterases: from molecules to functions. *Annu Rev Pharmacol Toxicol* 38**,** 257-288.

664 Shen, X.-X., Zhou, X., Kominek, J., Kurtzman, C.P., Hittinger, C.T., and Rokas, A. (2016).
665 Reconstructing the Backbone of the Saccharomycotina Yeast Phylogeny Using Genome-Scale
666 Data. *G3* 6(12)**,** 3927-3939.
667 Shimodaira, H. (2002). An approximately unbiased test of phylogenetic tree selection. *Syst Biol* 51(3)**,**
668 492-508.
669 Shimodaira, H., and Hasegawa, M. (2001). CONSEL: for assessing the confidence of phylogenetic tree
670 selection. *Bioinformatics* 17(12)**,** 1246-1247.
671 Simao, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., and Zdobnov, E.M. (2015). BUSCO:
672 assessing genome assembly and annotation completeness with single-copy orthologs.
673 *Bioinformatics* 31(19)**,** 3210-3212.
674 Slot, J.C., and Rokas, A. (2011). Horizontal transfer of a large and highly toxic secondary metabolic
675 gene cluster between fungi. *Curr Biol* 21(2)**,** 134-139.
676 Soanes, D., and Richards, T.A. (2014). Horizontal gene transfer in eukaryotic plant pathogens. *Annu*
677 *Rev Phytopathol* 52**,** 583-614.
678 Stanke, M., and Waack, S. (2003). Gene prediction with a hidden Markov model and a new intron
679 submodel. *Bioinformatics* 19(2)**,** ii215-225.
680 Stukenbrock, E.H., Jorgensen, F.G., Zala, M., Hansen, T.T., McDonald, B.A., and Schierup, M.H.
681 (2010). Whole-genome and chromosome evolution associated with host adaptation and
682 speciation of the wheat pathogen Mycosphaerella graminicola. *PLoS Genet* 6(12)**,** e1001189.
683 Tanaka, A., and Tsuge, T. (2000). Structural and functional complexity of the genomic region
684 controlling AK-toxin biosynthesis and pathogenicity in the Japanese pear pathotype of
685 Alternaria alternata. *Molecular Plant-Microbe Interactions* 13(9)**,** 975-986.
686 Thomma, B.P. (2003). Alternaria spp.: from general saprophyte to specific parasite. *Mol Plant Pathol*
687 4(4)**,** 225-236.
688 Tsuge, T., Harimoto, Y., Akimitsu, K., Ohtani, K., Kodama, M., Akagi, Y., et al. (2013). Host-selective
689 toxins produced by the plant pathogenic fungus Alternaria alternata. *FEMS Microbiology*
690 *Reviews* 37(1)**,** 44-66.
691 Tsuge, T., Harimoto, Y., Hanada, K., Akagi, Y., Kodama, M., Akimitsu, K., et al. (2016). Evolution of
692 pathogenicity controlled by small, dispensable chromosomes in Alternaria alternata
693 pathogens. *Physiological and Molecular Plant Pathology* 95**,** 27-31.
694 Wang, M., Sun, X., Yu, D., Xu, J., Chung, K., and Li, H. (2016). Genomic and transcriptomic analyses
695 of the tangerine pathotype of Alternaria alternata in response to oxidative stress. *Sci Rep*

6(32437).

Wisecaver, J.H., Alexander, W.G., King, S.B., Hittinger, C.T., and Rokas, A. (2016). Dynamic Evolution of Nitric Oxide Detoxifying Flavohemoglobins, a Family of Single-Protein Metabolic Modules in Bacteria and Eukaryotes. *Mol Biol Evol* 33(8)**,** 1979-1987.

Wisecaver, J.H., and Rokas, A. (2015). Fungal metabolic gene clusters-caravans traveling across genomes and environments. *Front Microbiol* 6(161).

Yamada, K.D., Tomii, K., and Katoh, K. (2016). Application of the MAFFT sequence alignment program to large data-reexamination of the usefulness of chained guide trees. *Bioinformatics* 32(21)**,** 3246-3251.

Yang, S.L., Yu, P.L., and Chung, K.R. (2016). The glutathione peroxidase-mediated reactive oxygen species resistance, fungicide sensitivity and cell wall construction in the citrus fungal pathogen Alternaria alternata. *Environ Microbiol* 18(3)**,** 923-935.

Acknowledgements

We thank members of the Rokas lab for helpful discussions. We thank Abigail Leavitt Labella and Jacob Steenwyk for their critical comments on this paper. This work was conducted in part using the resources of the Advanced Computing Center for Research and Education (ACCRE) at Vanderbilt University.

Author contributions

HL and AR supervised the work; MW, XS, NP, and JX conducted the analyses; HF and RR performed the experiment; MW drafted the initial version; and all authors contributed to the writing of the manuscript.

Funding

This work was supported by the National Foundation of Natural Science of China (31571948), the earmarked fund for China Agriculture Research System (CARS-27), the China Postdoctoral Science Foundation (2016M601945), and the US National Science Foundation (DEB-1442113 to AR).

Competing financial interests

The authors declare no competing financial interests.

Legends for Figures

FIGURE 1 GO enrichment analysis of the Z7 CDC genes in the category "Biological Process". Significantly enriched GO terms ($p<0.05$) are illustrated by rectangles. Rectangle color corresponds to degree of statistical significance and ranges from bright yellow (least significant) to dark red (most significant). For each node, the GO identifier, the GO term name, and p-value of each functional category is shown. The final line inside each node shows the number of genes that belong to the functional category in the CDC and in the whole genome of *A. alternata* Z7, respectively.

FIGURE 2 Sequence conservation of Z7 CDC genes across the phylogeny of *Alternaria* and representative species of Dothideomycetes. The phylogeny on the left depicts the evolutionary relationships of species of *Alternaria* and representative Dothideomycetes. The maximum likelihood phylogeny was inferred from the concatenation-based analysis of an amino acid data matrix comprised from 1,754 single-copy BUSCO genes under the LG+R10 substitution model. Branches with bootstrap support values of 100% are not shown; branches with bootstrap values <100% are shown near each branch. The heat map on the right was constructed using the sequence conservation value (BLAST identity * query coverage) of each Z7 CDC protein to its best counterpart in each of the species included in this analysis. Cells with red color correspond to proteins (in specific species) that exhibit high sequence conservation to a given Z7 CDC protein. The numbers next to each species' name correspond to the number of proteins that exhibit sequence conservation values greater or equal to 0.5 when compared to Z7 CDC proteins.

FIGURE 3 Distribution of the average branch length from all individual Z7 CDC (red line) or EC (blue line) gene trees constructed from groups of orthologous genes

within *Alternaria* Clade I. The dashed lines denote the mean values of the two distributions.

FIGURE 4 Horizontal transfer of a cluster of 5 genes in the tangerine pathotype of *A.alternata* Z7 CDC. A) Phylogenetic evidence of the HGT of these genes. For each gene, the maximum likelihood phylogeny was inferred under the best substitution model automatically selected by ModelFinder, as implemented in IQ-TREE 1.5.4. Branch colors indicate the taxonomic lineages to which the different taxa included in each phylogeny belong. The red asterisk indicates the *Alternaria* clade on each phylogeny. The full phylogenetic trees of the individual genes can be found in Figs. S2 - S6. B) Conservation of synteny among the cluster of 5 genes in *A. alternata* and the evolutionary related clusters present in *Penicillium flavigenum* and *Cryptococcus gatti*. Orthologs among different species are marked with same color and homologous regions are shown by the gray boxes. Arrows indicate gene direction.

FIGURE 5 Horizontal transfer of 4 genes in the ACT toxin gene cluster in the tangerine pathotype of *A.alternata* Z7 CDC. A) Phylogenetic evidence of the HGT of these genes. For each gene, the maximum likelihood phylogeny was inferred under the best substitution model automatically selected by ModelFinder, as implemented in IQ-TREE 1.5.4. Branch colors indicate the taxonomic lineages to which the different taxa included in each phylogeny belong. The red asterisk indicates the *Alternaria* clade on each phylogeny. The full phylogenetic trees of the individual genes can be found in Figs. S7 - S10. B) Conservation of synteny among the ACT toxin gene cluster in *A.alternata* Z7 and a predicted secondary metabolic gene cluster in *Colletotrichum tofieldiae* 0861. Orthologs among different species are marked with same color and homologous regions are shown by the gray boxes. Arrows indicate

gene direction.

FIGURE 6 The gene AALT_g12037 was horizontally transferred from Bacteria to the *A. alternata* Z7 CDC. A) ML phylogeny of AALT_g12037 homologs across the tree of life. The gene's maximum likelihood phylogeny was inferred under the best substitution model automatically selected by ModelFinder, as implemented in IQ-TREE 1.5.4. Branch colors indicate the taxonomic lineages to which the different taxa included in each phylogeny belong. The red asterisk indicates the *Alternaria* clade on each phylogeny. The full tree can be found in Fig. S11. B) The presence of the HGT-acquired gene family in other *Alternaria* genomes. Light grey indicates the lack of this HGT gene in the corresponding species.

FIGURE 7 Functional analysis of the role of the horizontally transferred CDC gene AALT_g12037 in pathogenicity. A) Schematic depiction of gene disruption within AALT_g12037 via a homologous integration. Numbers denote primers listed in supplementary table S3. B) PCR verification of the deletion of AALT_g12037 gene. Long bands can be only amplified in the mutants (KO) using the outside PCR primer pairs 7/8 and 9/10 while the short band can be only amplified in the wildtype (WT) using the inner PCR primer pair 11/12. C) Virulence of the wild-type Z7 and the AALT_g12037 gene mutant to citrus leaves.

Legends for supplementary Figures

Fig. S1. Bar plot of the degree of sequence conservation of Z7 CDC proteins to the proteomes of other species in the Dothideomycetes. The sequence conservation value of each Z7 CDC protein to its best counterpart in each of the species was calculated by BLAST identity * query coverage.

Figs. S2 – S15. Full ML phylogenetic trees of HGT gene homologs across the tree of life. Each gene's maximum likelihood phylogeny was inferred under the best substitution model automatically selected by ModelFinder, as implemented in IQ-TREE 1.5.4. Branch colors indicate the taxonomic lineages to which the different taxa included in each phylogeny belong. The *A. alternata* Z7 CDC sequence that was used as a query in the BLAST search is shown in red font. Bootstrap values are shown near each branch. Fig. S2: AALT_g11769; Fig. S3: AALT_g11770; Fig. S4: AALT_g11771; Fig. S5: AALT_g11772; Fig. S6: AALT_g11773; Fig. S7: AALT_g12032; Fig. S8: AALT_g11755; Fig. S9: AALT_g11757; Fig. S10: AALT_g11758; Fig. S11: AALT_g12037; Fig. S12: AALT_g11567; Fig. S13: AALT_g11777; Fig. S14: AALT_g11781; Fig. S15: AALT_g11892.

Table 1. General features of the essential chromosomes (EC) and conditionally dispensable chromosome (CDC) in *A.alternata* strain Z7

Features	*A. alternata* EC	*A. alternata* CDC
Genome size (Mb)	32.53	1.84
Number of contigs	115	41
GC content (%)	51.2	47.7
Protein-coding genes	11536	512
Gene density (number of genes per Mb)	355	279
Mean gene length (bp)	1736	1516
Mean number of exons per gene	2.8	2.5
Mean length of exons	559	531
Mean number of introns per gene	1.8	1.5
Mean length of introns	92	121
Percentage of genes without intron (%)	24	32
Repeat rate (%)	0.51	1.23
tRNA genes	115	0

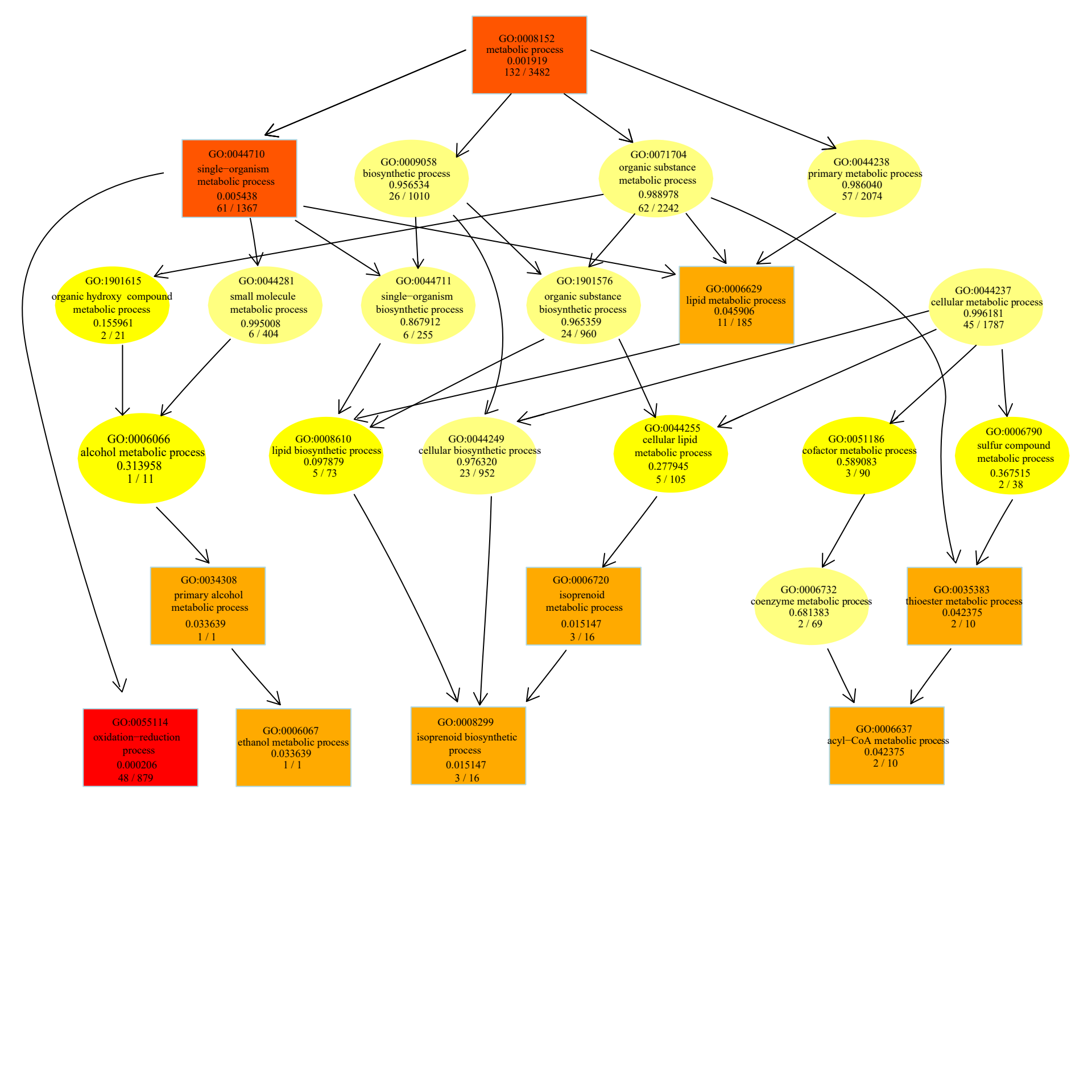
Table 2 Annotation of the differentially expressed genes located in the conditionally dispensable chromosome with absolute log2FC >2 during H_2O_2 stress

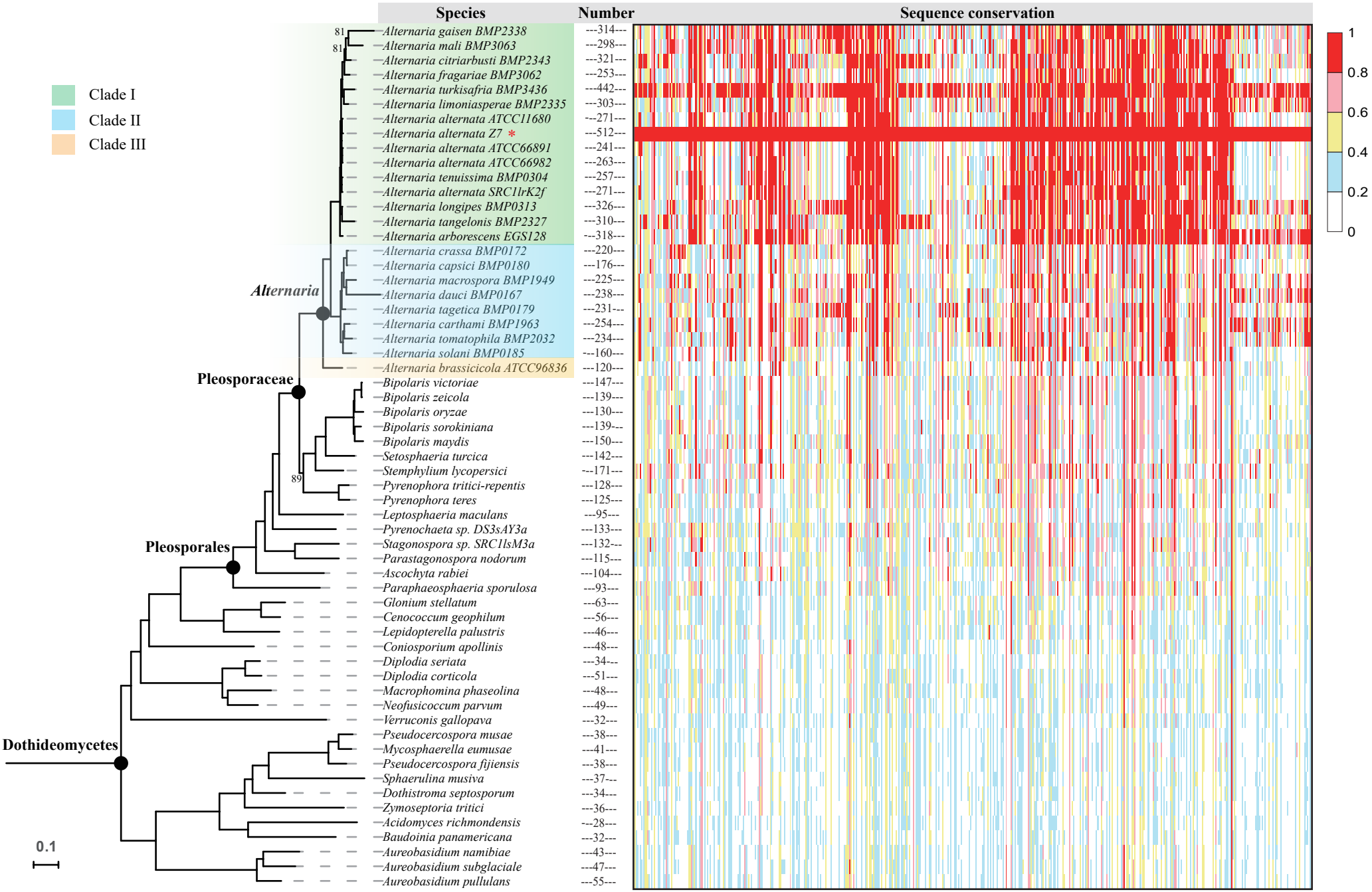
gene id	log2FC	padj	Pfam	domain	Description
AALT_g11773	4.3712	1.09E-07			oxidoreductase
AALT_g11774	4.2924	4.21E-08	PF05368	NmrA-like family	nucleoside-diphosphate-sugar epimerase family
AALT_g11717	3.9966	9.78E-05			hypothetical protein
AALT_g11883	3.5338	7.90E-04			hypothetical protein
AALT_g11772	3.4436	3.14E-05	PF01408	Oxidoreductase family, NAD-binding Rossmann fold	myo-inositol 2-dehydrogenase
AALT_g11775	3.4189	4.31E-09			hypothetical protein
AALT_g11750	3.3427	4.95E-10	PF14765, PF00698, PF08659	Polyketide synthase dehydratase, Acyl transferase domaindomain, Phosphopantetheine, attachment site, Methyltransferase domain,	polyketide synthase
AALT_g11216	3.2515	9.34E-09			hypothetical protein
AALT_g11903	3.2159	1.51E-33	PF12697	Abhydrolase_6	alpha beta hydrolase
AALT_g11194	3.2043	1.67E-26	PF01425	Amidase	glutamyl-tRNA(Gln) amidotransferase subunit A
AALT_g11797	3.1694	3.79E-35	PF05199, PF00732	GMC oxidoreductase	glucose oxidase
AALT_g11943	2.6733	1.63E-04			hypothetical protein
AALT_g11172	2.5340	1.50E-11	PF00797	N-acetyltransferase	Arylamine N-acetyltransferase
AALT_g11565	2.4221	7.76E-11	PF06331	Transcription factor TFIIH complex subunit Tfb5	RNA polymerase II transcription factor B subunit 5
AALT_g11916	2.3833	1.03E-07	PF12311	Protein of unknown function	hypothetical protein
AALT_g11577	2.2458	4.84E-11	PF02586	SOS response associated peptidase	putative duf159 domain protein
AALT_g11978	2.1658	4.45E-04	PF00651	BTB/POZ domain	hypothetical protein
AALT_g11054	-2.0725	3.00E-03	PF03184, PF03221, PF05225	DDE superfamily endonuclease, Tc5 transposase DNA-binding domain, helix-turn-helix, Psq domain	putative transposase
AALT_g11894	-2.0867	5.54E-05	PF00175, PF00667, PF00258, PF00067	Oxidoreductase NAD-binding domain, FAD binding domain, Flavodoxin, Cytochrome P450	bifunctional P-450/NADPH-P450 reductase
AALT_g11037	-2.2114	1.68E-04			hypothetical protein
AALT_g11895	-2.8403	2.37E-05	PF00487	Fatty acid desaturase	omega-6 fatty acid desaturase (delta-12 desaturase)

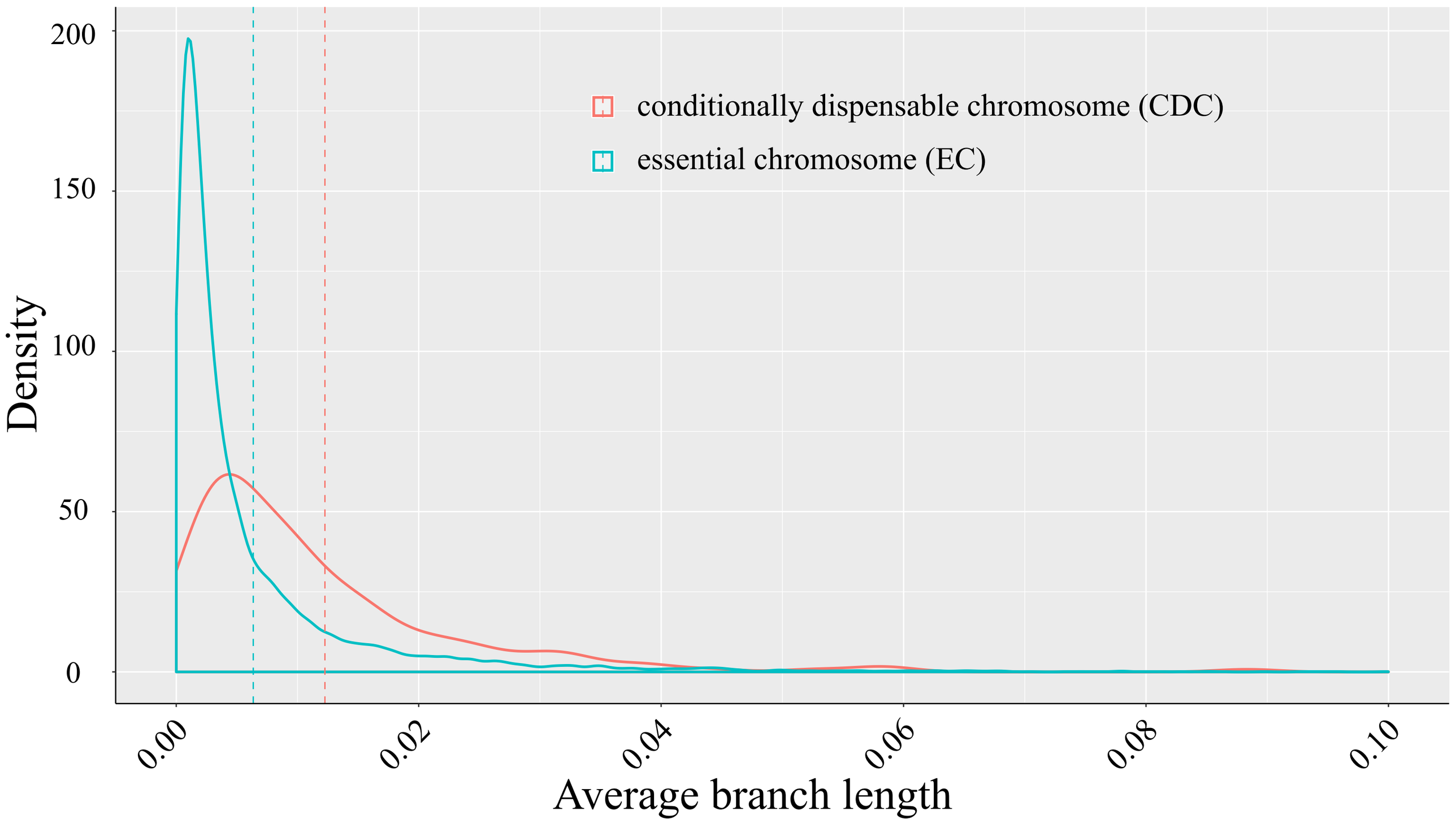
Table 3 Summary list of all horizontal gene transfers (HGTs) in *A.alternata* Z7 CDC.

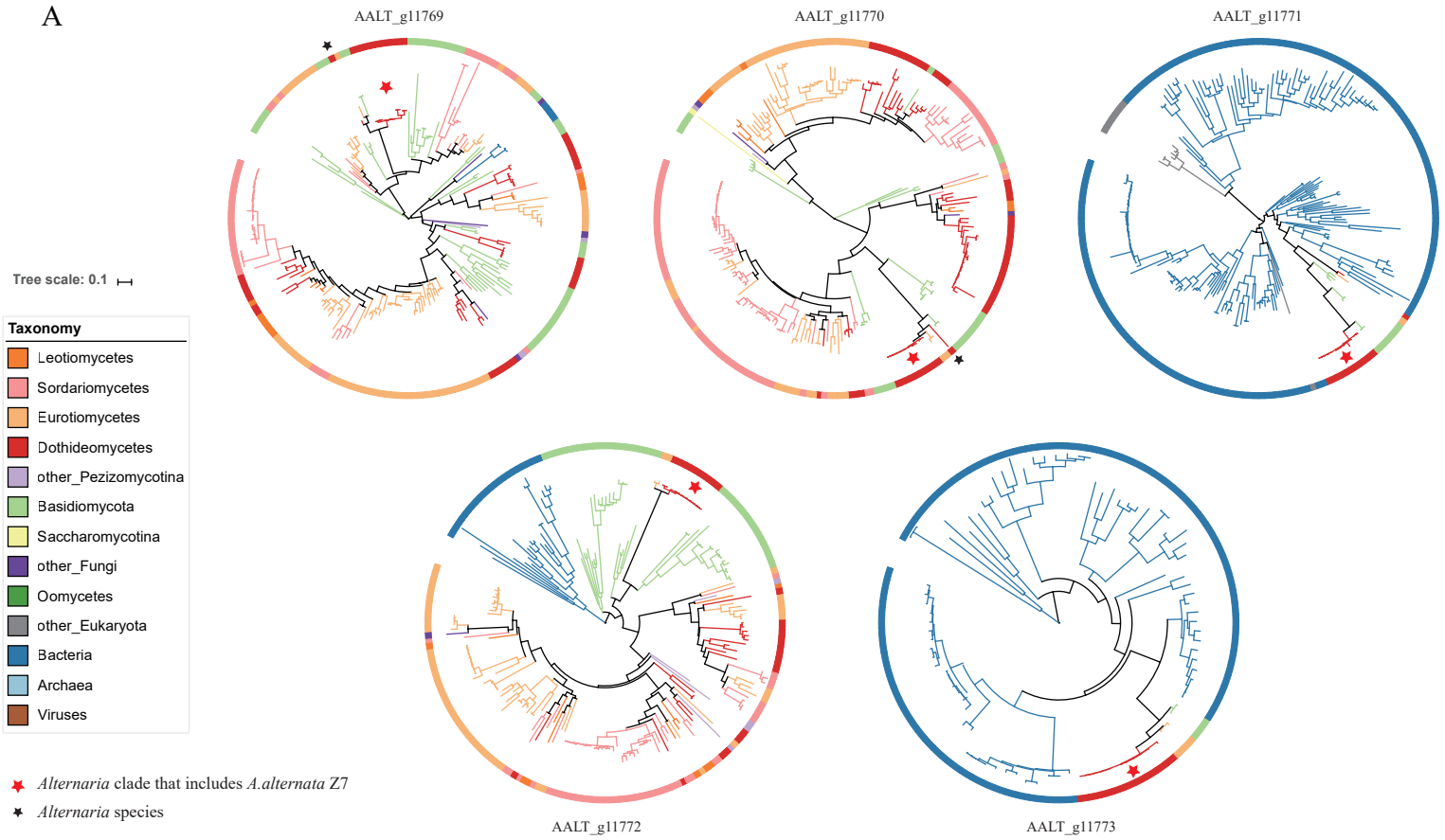
Gene ID	Protein length	Closest sequence	Closest species	Proteome Identity	Gene Identity	Description	Distribution in *Alternaria*	AU test *P* value
AALT_g11567	667	EGU73680.1	*Fusarium oxysporum*	50	90	alcohol oxidase	Clade I, II	2E-08
AALT_g11755	296	KZL71655.1	*Colletotrichum tofieldiae*	51	64	HMG-CoA hydrolase	3 pathotypes	2E-50
AALT_g11757	2349	OLN84341.1	*Colletotrichum chlorophyti*	50	59	polyketide synthase	3 pathotypes	3E-52
AALT_g11758	439	OLN84339.1	*Colletotrichum chlorophyti*	50	74	Cytochrome P450, Aft11-1	3 pathotypes	2E-56
AALT_g11769	320	OQE18800.1	*Penicillium flavigenum*	51	71	short chain dehydrogenase	Clade I	2E-40
AALT_g11770	554	OQE18785.1	*Penicillium flavigenum*	51	89	hexose transporter	Clade I	2E-44
AALT_g11771	396	ADV23792	*Cryptococcus gattii*	41	83	neuraminidase	Clade I	2E-06
AALT_g11772	377	OQE18924.1	*Penicillium flavigenum*	51	89	NAD(P)-binding oxidoreductase	Clade I	3E-76
AALT_g11773	332	OQE18776.1	*Penicillium flavigenum*	51	87	gfo/Idh/MocA family oxidoreductase	Clade I	/
AALT_g11777	646	KFY76151.1	*Pseudogymnoascus sp. VKM F-103*	52	78	flavin-containing monooxygenase	Clade I, II	1E-06
AALT_g11781	362	KFZ02874.1	*Pseudogymnoascus sp. VKM F-4520*	49	80	NAD-binding Rossmann fold oxidoreductase	Clade I, II	3E-04
AALT_g11892	377	GAM40406.1	*Talaromyces cellulolyticus*	52	54	isopropanol dehydrogenase	Clade I	1E-06
AALT_g12032	349	KZL71680.1	*Colletotrichum tofieldiae*	51	72	Acyl-CoA dehydrogenase, Aft10-1	3 pathotypes	3E-73
AALT_g12037	430	EWZ28488.1	*Fusarium oxysporum*	50	73	carboxylesterase	Clade I, II	/

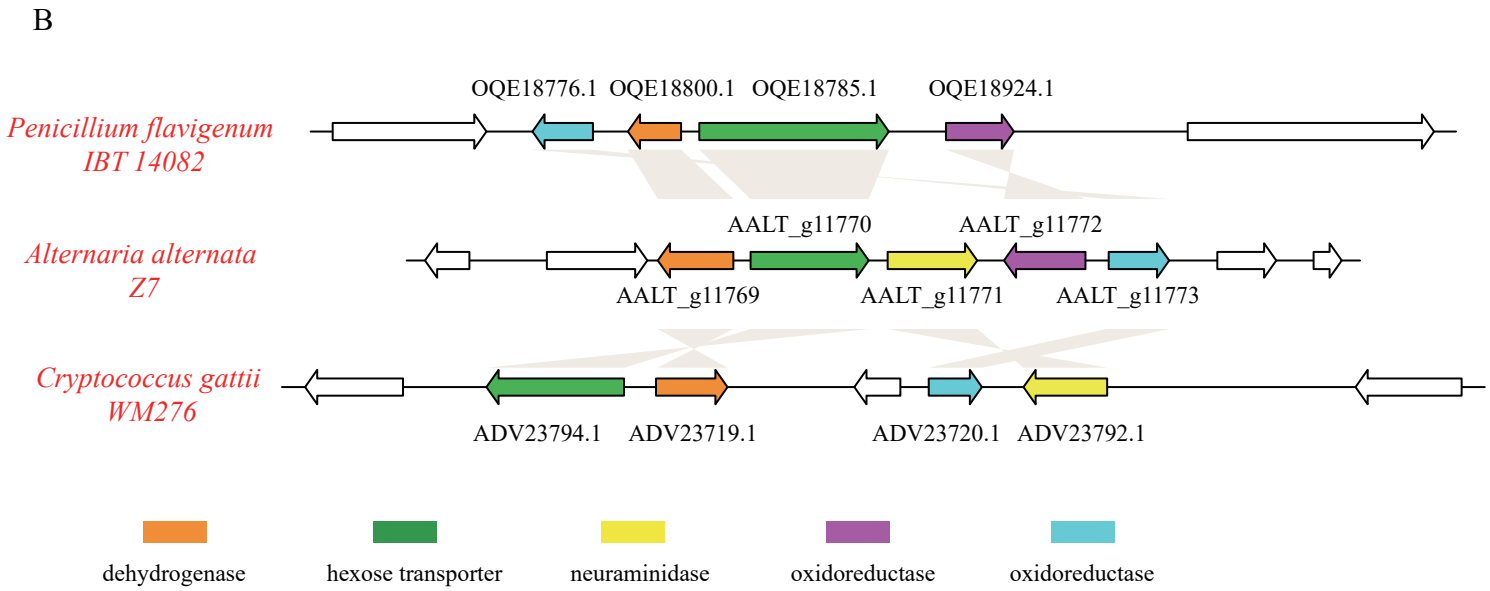
3 pathotypes: the Japanese pear, strawberry and tangerine pathotypes.


GO:0008152
metabolic process
0.001919
132 / 3482
GO:0044710
single−organism
metabolic process
0.005438
61 / 1367
GO:0009058
biosynthetic process
0.956534
26 / 1010
GO:0071704
organic substance
metabolic process
0.988978
62 / 2242
GO:0044238
primary metabolic process
0.986040
57 / 2074
GO:1901615
organic hydroxy compound
metabolic process
0.155961
2 / 21
GO:0044281
small molecule
metabolic process
0.995008
6 / 404
GO:0044711
single−organism
biosynthetic process
0.867912
6 / 255
GO:1901576
organic substance
biosynthetic process
0.965359
24 / 960
GO:0006629
lipid metabolic process
0.045906
11 / 185
GO:0044237
cellular metabolic process
0.996181
45 / 1787
GO:0006066
alcohol metabolic process
0.313958
1 / 11
GO:0008610
lipid biosynthetic process
0.097879
5 / 73
GO:0044249
cellular biosynthetic process
0.976320
23 / 952
GO:0044255
cellular lipid
metabolic process
0.277945
5 / 105
GO:0051186
cofactor metabolic process
0.589083
3 / 90
GO:0006790
sulfur compound
metabolic process
0.367515
2 / 38
GO:0034308
primary alcohol
metabolic process
0.033639
1 / 1
GO:0006720
isoprenoid
metabolic process
0.015147
3 / 16
GO:0006732
coenzyme metabolic process
0.681383
2 / 69
GO:0035383
thioester metabolic process
0.042375
2 / 10
GO:0055114
oxidation−reduction
process
0.000206
48 / 879
GO:0006067
ethanol metabolic process
0.033639
1 / 1
GO:0008299
isoprenoid biosynthetic
process
0.015147
3 / 16
GO:0006637
acyl−CoA metabolic process
0.042375
2 / 10

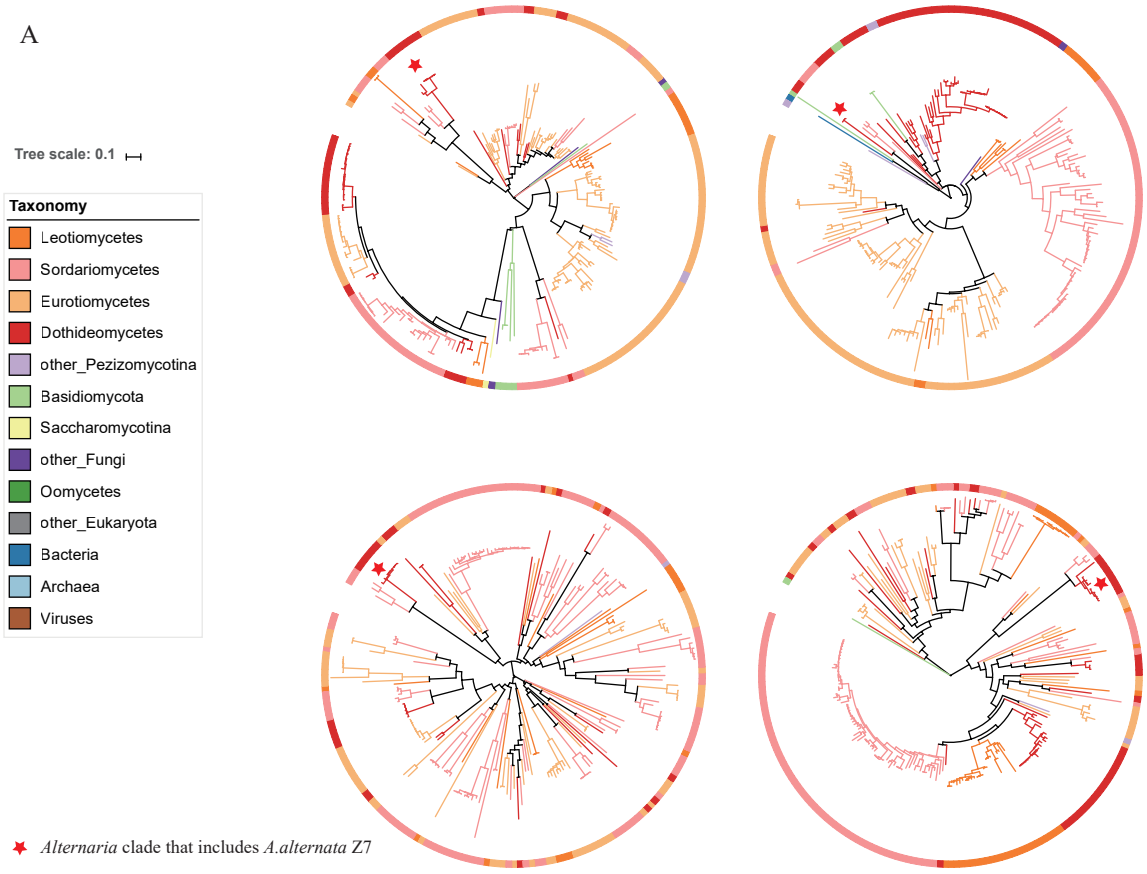

Species
Number
Sequence conservation
Clade I
Clade II
Clade III
Alternaria gaisen BMP2338 ---314---
Alternaria mali BMP3063 ---298---
Alternaria citriarbusti BMP2343 ---321---
Alternaria fragariae BMP3062 ---253---
Alternaria turkisafria BMP3436 ---442---
Alternaria limoniasperae BMP2335 ---303---
Alternaria alternata ATCC11680 ---271---
Alternaria alternata Z7 * ---512---
Alternaria alternata ATCC66891 ---241---
Alternaria alternata ATCC66982 ---263---
Alternaria tenuissima BMP0304 ---257---
Alternaria alternata SRC1lrK2f ---271---
Alternaria longipes BMP0313 ---326---
Alternaria tangelonis BMP2327 ---310---
Alternaria arborescens EGS128 ---318---
Alternaria crassa BMP0172 ---220---
Alternaria capsici BMP0180 ---176---
Alternaria macrospora BMP1949 ---225---
Alternaria dauci BMP0167 ---238---
Alternaria tagetica BMP0179 ---231---
Alternaria carthami BMP1963 ---254---
Alternaria tomatophila BMP2032 ---234---
Alternaria solani BMP0185 ---160---
Alternaria brassicicola ATCC96836 ---120---
Bipolaris victoriae ---147---
Bipolaris zeicola ---139---
Bipolaris oryzae ---130---
Bipolaris sorokiniana ---139---
Bipolaris maydis ---150---
Setosphaeria turcica ---142---
Stemphylium lycopersici ---171---
Pyrenophora tritici-repentis ---128---
Pyrenophora teres ---125---
Leptosphaeria maculans ---95---
Pyrenochaeta sp. DS3sAY3a ---133---
Stagonospora sp. SRC1lsM3a ---132---
Parastagonospora nodorum ---115---
Ascochyta rabiei ---104---
Paraphaeosphaeria sporulosa ---93---
Glonium stellatum ---63---
Cenococcum geophilum ---56---
Lepidopterella palustris ---46---
Coniosporium apollinis ---48---
Diplodia seriata ---34---
Diplodia corticola ---51---
Macrophomina phaseolina ---48---
Neofusicoccum parvum ---49---
Verruconis gallopava ---32---
Pseudocercospora musae ---38---
Mycosphaerella eumusae ---41---
Pseudocercospora fijiensis ---38---
Sphaerulina musiva ---37---
Dothistroma septosporum ---34---
Zymoseptoria tritici ---36---
Acidomyces richmondensis ---28---
Baudoinia panamericana ---32---
Aureobasidium namibiae ---43---
Aureobasidium subglaciale ---47---
Aureobasidium pullulans ---55---
Alternaria
Pleosporaceae
Pleosporales
Dothideomycetes
81
81
89
0.1
1
0.8
0.6
0.4
0.2
0


conditionally dispensable chromosome (CDC)
essential chromosome (EC)
Density
200
150
100
50
0
0.00
0.02
0.04
0.06
0.08
0.10
Average branch length

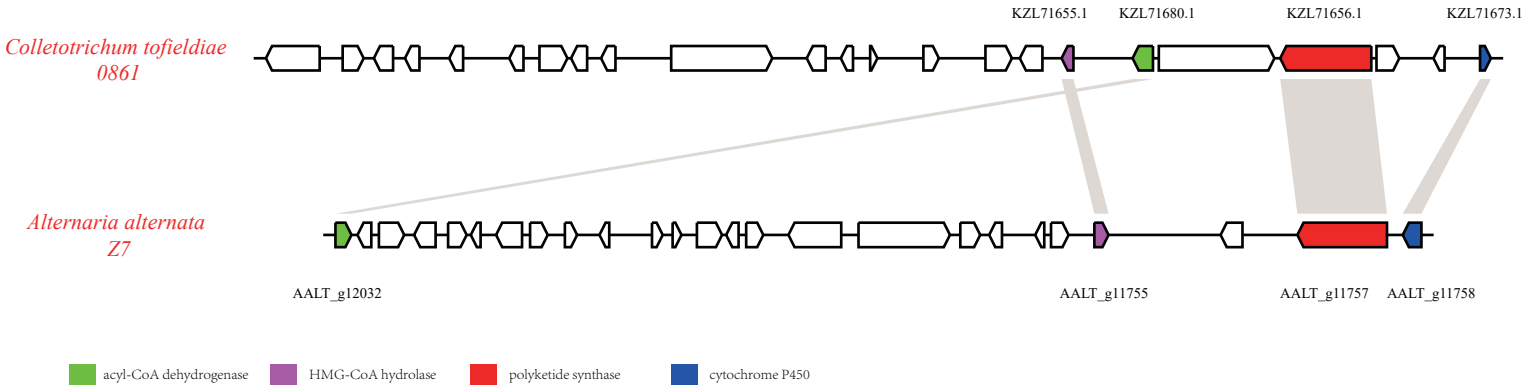




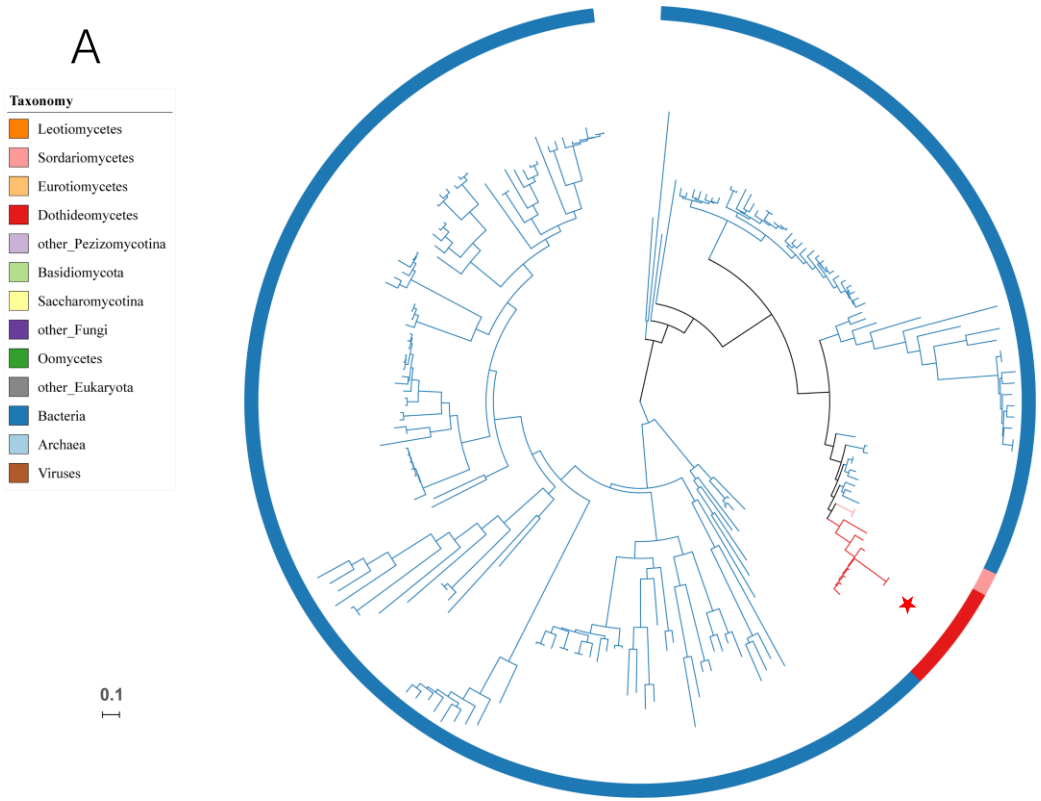
A



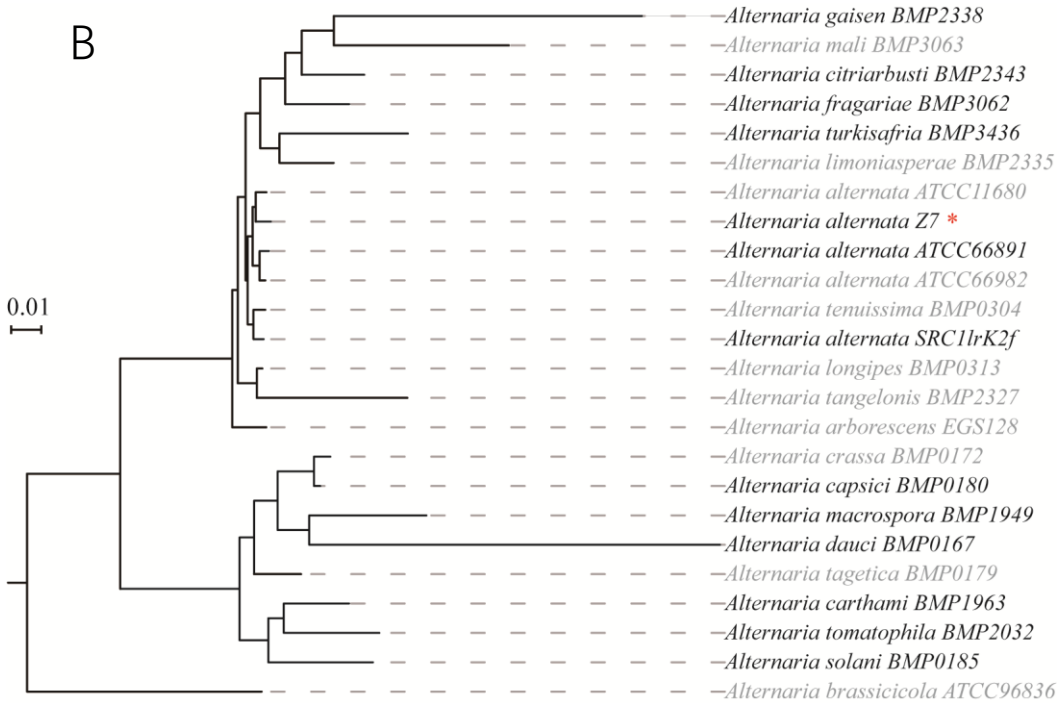
B



A

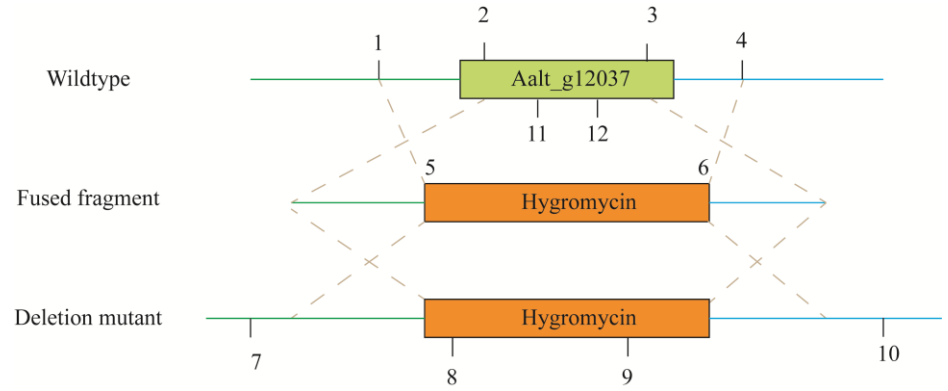


B

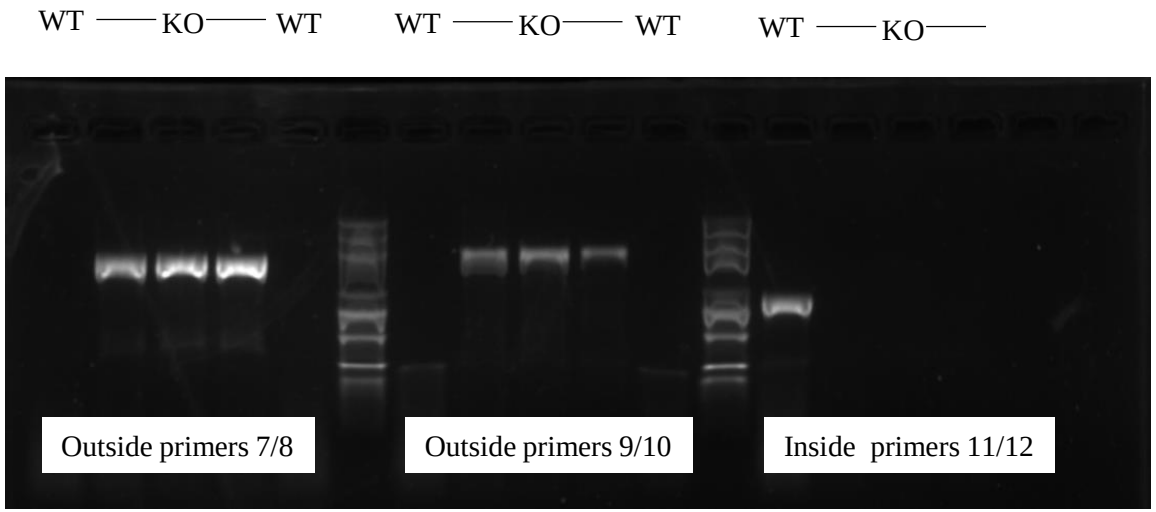


* *Alternaria* clade that includes *A.alternata* Z7

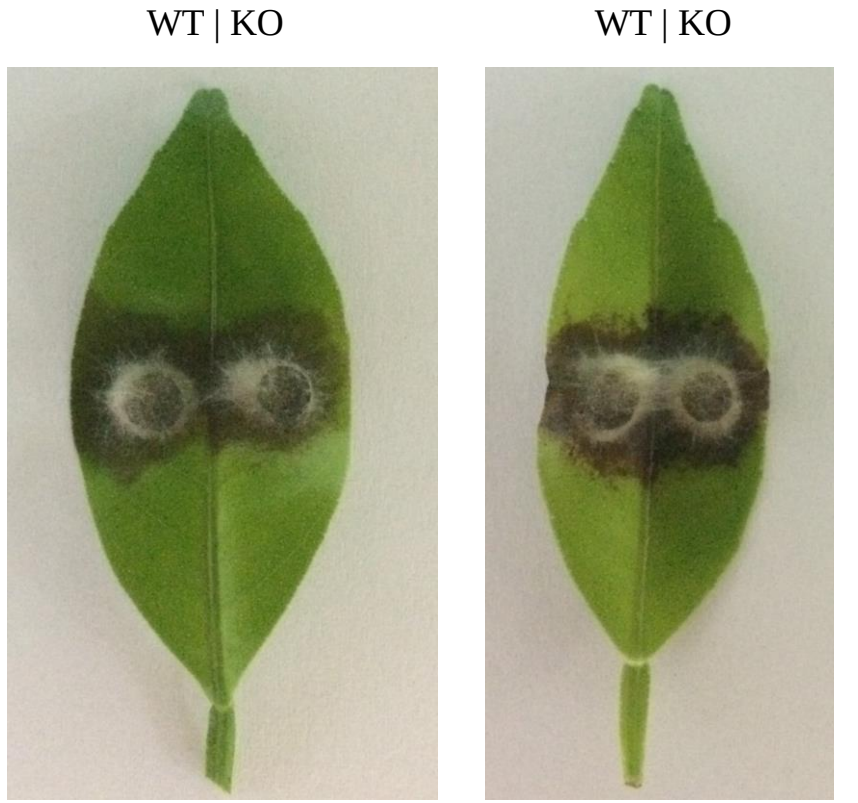
A


Wildtype
Aalt_g12037
Fused fragment
Hygromycin
Deletion mutant
Hygromycin
1
2
3
4
5
6
7
8
9
10
11
12

B


WT KO WT
WT KO WT
WT KO
Outside primers 7/8
Outside primers 9/10
Inside primers 11/12

C


WT | KO
WT | KO
Citrus × clementina
Citrus poonensis