Genomic analysis for Biosynthesis and Metabolic Pathway of Elsinochrome Toxin produced by *Elsinoë arachidis*

Wenli Jiao
Shenyang Agricultural University
Mengxue Xu
Shenyang Agricultural University
Rujun Zhou (zhourujun@syau.edu.cn)
Shenyang Agricultural University
Lu Liu
Shenyang Agricultural University
Zibo Li
Shenyang Agricultural University
Caiyun Xue
Shenyang Agricultural University

Research article

Keywords: Elsinoë arachidis, Genomic analysis, Elsinochrome

Posted Date: April 10th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-21061/v1

Abstract

Background: *Elsinoë arachidis*, an important peanut pathogenic fungus that distributes widely and leads to large-scale losses in peanut producing regions in China, produce elsinochromes (ESCs) as the vital toxin through pathogenic process. However, the biosynthesis of elsinochromes have not been investigated and the transcriptional response of the light on synthesis of elsinochrome in *Elsinoë* is poorly understood.

Results: In this paper, high-quality genome of *E. arachidis* by PacBio RS II sequencing method was reported. The 33.18Mb genome encodes 9056 predicted genes, of which the proportion of genes encoding secreted proteins 8.0% (734 secreted proteins), 124 transporter-related genes, 949 signal peptides, 1,829 transmembrane protein-coding genes, 127 non-coding RNAs and 13 pseudogenes. Mapped the *E. arachidis* assembly to 16 scaffold that contain 86 secondary metabolites gene clusters, including six polyketide synthase gene clusters encoding two melanin, one elsinochromes and three T-toxin were also identified in the genome. Additionally, ESC biosynthesis-related gene cluster predicted to contain *ESCB1* which was high expression under light condition.

Conclusion: Analysis of genomic information of *E. arachidis* lays a solid molecular data foundation for further exploration of pathogenic mechanisms and toxin biosynthesis pathways. Taken together, we provide a valuable foundation in the biosynthesis of elsinochrome and essential information for further comprehend its virulence mechanism.

Background

Peanut (*Arachis hypogaea* Linn.), the primary plant lipid resources, plays a vital role in China agricultural production. The peanut scab caused by *Sphaceloma arachidis* Bitaucourt et Jenkins have ever made tremendous yield loss in peanut growing areas in China. The reduction in the general field is 10% ~ 30%, and the serious disease field loss can reach more than 50%.[1–2] There have been many projects dedicated to discover patterns of disease occurrence, however, the characteristics of the molecular mechanism of the host-pathogen interaction is still blank, which cannot provide the theoretical basis for the innovation of prevention and control strategies[3–6].

The pathogen, *E. arachidis*, causes scab on leaves, petioles and stems, and distortions of peanut. Species of the phytopathogenic genus *Elsinoë* (conidial stage: *Sphaceloma*) cause scab and spot anthracnose on field crops such as cassava, ornamentals and economic crops including avocado, mango, grape and citrus, several of them produce ESCs [7, 8]. ESCs produced by *E. arachidis* is necessary to fungal virulence, which share the same 4,9-dihydroxy-3,10-perylenequinone chromophore with cercosporin (*Cercospora* spp.) and altertoxin I (*Alternaria alternata*) whose structures fit the classification of a group of perylenequinone toxins [9–11]. Most of the identified perylenequinone toxins are the products of the polyketide synthase pathway of the Ascomycota fungi [12]. However, compared to the fumonisin and cercosporin, the biosynthetic pathways and regulation is still in infancy. To explore the role and mode of

plant pathogenic mycotoxins and biosynthetic pathways is of great significance for revealing the pathogenic infection ways, pathogenic mechanisms and interaction mechanisms.

Recent years, with the development of gene sequencing technology, the whole genome sequencing technology has been rapidly popularized and applied in the research field of phytopathogenic fungi. So far, thousands of fungal genomes have been sequenced and analyzed, such as *Pyreriaia grisea* [13], *Fusarium graminearum* [14] and *Exserohilum turcicum* [15]. The discovery of the genome provided reliable molecular evidence for scientific questions such as genetic evolution, pathogenic differentiation, secondary metabolism, biosynthetic pathways and pathogenic mechanisms of plant pathogenic fungi. *Magnaporthe oryzae* is one of the main diseases affecting rice production. Sequencing of its genome, help scholars understand the pathogenic mechanism more comprehensively, such as MoEnd3, one of the secret proteins in the genome, have been demonstrated to regulates the appressorium formation of *Magnaporthe oryzae* [16]. Meanwhile, Genomic analysis of the closely related *Sclerotinia sclerotiorum* and *Botrytis cinerea* which offer insight into the different features of their pathogenic process [17].

In this study, we first report the high-quality genome sequence of *E. arachidis*, which could be used for comparative genomics to explore evolutionary relationships between species. In addition, key gene *ESCB1* involved in ESC biosynthetic gene cluster were predicted, which would provide new directions and ideas for further in-depth research of toxin biosynthesis. Analysis of pathogen-host interaction genes allow researchers to more understanding *E. arachidis* during the investigation of pathogenic mechanism.

Results

Genome features

The genome of *E. arachidis* LNFT-H01 was sequenced (100 × coverage), high-quality sequencing data, a total of 6.28 Gb, was obtained and de novo assembled into 16 scaffolds (N50, 3,376,838 bp) with a total size of around 33.18 Mb by CANU, large than *E. australis* (23.34 Mb) (GenBank: NHZQ00000000) and *Sphaceloma murrayae* (20.72 Mb) (GenBank: NKHZ00000000). In which, sixteen scaffolds above 1Kb, the length is 33,184,353 bp and the longest scaffold is 4,426,246 bp.

The completeness of *E. arachidis* LNFT-H01 genome was evaluated to be > 99%, totally encoded 9174 protein genes similar to *E. australis* (9223) and more than *S. murrayae* (8256). The proportion of genes encoding secreted proteins in *E. arachidis* LNFT-H01 was 8.0% (734 proteins), the proportions of secreted proteins in *E. arachidis* was close to 7–10%. Sixteen scaffolds were displayed by circus-plot (Fig. 1), the gene density is 285 genes per 1 Mb and have 127 Non-coding RNA and 13 pseudogenes are predicted in the genome.

Phylogenetic Analysis And Collinear Analysis

Phylogenetic analysis shows that *E. arachidis* is close to *Sphaceloma murrayae* and *E. australis*(Fig. 2A). In addition, synteny analysis of *E. arachidis* genome with *E. australis*, reveals that *E. arachidis* highest synteny with *E. australis*. For example, scaffolds 1, 5, 6 and 17 of *E. australis* correspond well with the scaffold 1 of *E. arachidis*, scaffold 32 and 37show well synteny to the scaffold 10 of *E. arachidis*(Fig. 2B).

Repetitive DNA Sequences And Methylation Sites

In eukaryotic genomes, repetitive DNA sequences have a critical role in genes function and genome structure, meanwhile, the different types and the proportion of repetitive sequences in a genome are different between species.[48, 49] Among 16 scaffolds of *E. arachidis* LNFT-H01 genome, 7,033,311 bp repeat sequences were totally identified, such as DNA transposon and LTR retrotransposon (Table S2), which were accounts for 21.4% of the genome, in which, LTRs accounting for 78.46%.

DNA methylation is significant in epigenetic processes and cell processes [50, 51].The fungus genome contains a variety of DNA modifications, the most common of which are adenine methylation and cytosine methylation. 1,033,888 4 m-C (4-methyl-cytosine) and 28,762 6 m-A (6-methyl-adenosine) were identified in *E. arachidis* LNFT-H01 genome, additionally, m4C are the majority of methylation (97.3%), whereas m6A only 2.7%. As for methylase-specific motif based on DNA methylation, we identified by DNA polymerase kinetic information and detect the fungus-specific motif(Table S3).

Functional annotation of E. arachidis

Functional annotation analysis of *E. arachidis*, a total of 8644 of the 9174 encoded protein sequences annotated (94.22%). Among them, GO, KEGG and KOG analysis respectively annotated 3237, 3055 and 4958 genes, accounting for 35.28%, 33.30% and 54.04%, respectively.

In order to clarify the secondary metabolic pathway of *E. arachidis*, KEGG used to identify the biological pathway in *E. arachidis*. The results of KEGG annotation show that the substance metabolism in this pathogen is active, including not only the formation of nutrients such as amino acids and sugars, but also the synthesis of some secondary metabolites. 4958 genes of *E. arachidis* were assigned to 24 functional regions of KOG annotation, and the number of genes distributed in different KOG categories was significantly different (Fig S1). The functional regions of the gene that account for a high percentage of the annotated results are posttranslational modification, protein turnover, signal transduction mechanisms, carbohydrate transport and metabolism, amino acid transport and metabolism, and secondary metabolites biosynthesis. The functional genes involved in transport and catabolism are abundant, and there are many genes involved, including lipid transport and metabolism, transport and metabolism related genes such as ions and coenzymes.

In the GO analysis, 3237 genes were further divided into 42 GO functional classify in biological process, cellular component and molecular function (Fig S2). The proportion of catalytic activity and metabolic

process is high, and detoxification and antioxidant activity related to pathogen self-detoxification are also noted. The annotation of these gene functions for further study of the secondary metabolic biosynthesis and transport process of toxins provides a rich data base of *E. arachidis*.

Gene Family Analysis

Analyses of the E. arachidis **genome for pathogenicity proteins**

In order to explore the potential pathogenic genes of *E. arachidis*, using the pathogen-host interaction database for Blastp alignment, 2,752 genes were screened from the *E. arachidis* genome, including secondary metabolic synthesis of key genes, cytochrome P450, ATP-binding cassette superfamily (ABC) transporter and Major Facilitator Superfamily (MFS) and other related genes, can be speculated on the complexity of the disease(Table S4).

Gene Associated With Detoxification

ESC, which produced by *E. arachidis*, can produce a large amount of active oxygen under light conditions. Active oxygen can act on cell membranes and destroy its structure. *E. arachidis* can also grow and develop in the case, indicating that it has a certain detoxification effect.

The MFS transporter and ABC transporter are the two largest families of fungal transporters[52–53]. A total of 57 ABC superfamily transporter genes and 190 MFS superfamily transporter genes were obtained from the genome of *E. arachidis*. In addition, the cytochrome P450 enzyme system is a multifunctional oxidoreductase. [54]. In the genome of *E. arachidis*, 78 cytochrome P450 enzymes were predicted, these may be involved in the synthesis and the detoxification of toxins.

The CAZyme

The carbohydrate-active enzymes secreted by pathogenic fungi are involved in the process of pathogen infection of host plant cells [57], which play an essential role in the decomposition of monosaccharides and polysaccharides, synthesis and modification of carbohydrates [58]. As one of the four major organic molecules in the organism, sugar metabolism is the center of the entire biological metabolism. It is not only an important structural component in the growth and development of plants, but also a signal molecule for communication between cells. Mapped *E. arachidis* genomes with CAZy database to detect the presence of CAZymes. 602 genes may be code carbohydrate-active enzyme (CAZymes) were defined (Table 2), including glycosyl transferases(114), carbohydrate esterases(109), glycoside hydrolases(271) and polysaccharide lyases(16).

Table 1

Gene annotation summary statistics

Genome features	
Scaffold Number	16
Scaffold Length (bp)	33,184,353
Scaffold N50 (bp)	3,376,838
Scaffold N90 (bp)	2,306,82
Scaffold Max (bp)	4,426,246
Gap total Length (bp)	0
Genome assembly (Mb)	33.18
Number of coding sequence genes	9,174
Average Exons length	641.1
Average Introns length	101.9
Total Genes length	15,965,221
CDSs Percentage of genome	43.9427%
GC Content (%)	48.24
Secreted protein	734
Transmembrane protein	1,829
Theproteins withsignal peptide	949
PHI	2,752
TCDB	124

Table 2
The carbohydrate-active enzymes

Classification	Number
Glycoside Hydrolases (GHs)	271
Polysaccharide Lyases (PLs)	16
Carbohydrate Esterases (CEs)	109
Glycosyl Transferases (GTs)	114
Auxiliary activities (AAs)	84
Carbohydrate-Binding Modules (CBMs)	66
Additional Files	
FigS1 KOG annotation of *E. arachidis*.	
FigS2 GO annotation of *E. arachidis*.	

The cuticle and cell wall on the plant surface are the first barriers to prevent the invasion of pathogen. Most pathogen invade the host defense system by producing cutinase and cell wall degrading enzymes. Through, further analysis detect pectin lyase(15), cutinase(13) and cellulase(19), the function in pathogenitic process still futher to study.

Secondary Metabolite Gene Clusters

E. arachidis can produce secondary metabolites ESC[5]. Although ESC is an important pathogenic factor in *E. arachidis*, the core gene for ESC synthesis has not be clarified in *E. arachidis*. The genome of *E. arachidis* provides the possibility to find the core genes for ESC. To identify gene clusters responsible for biosynthesis of polyketides in *E. arachidis*, using antiSMASH2 to identify all secondary metabolites clusters. Totally own 86 predicted secondary metabolites clusters, including polyketide synthase (PKS), nonribosomal peptide synthetase (NRPS), NRPS-PKS hybrid and others. The number and distribution of coding genes contained in gene clusters are different.

Identification and analysis of PKS **Genes in** E. arachidis

To further clarify the *PKS* gene cluster that regulates the biosynthesis of ESC in *E. arachidis*, a total of 19 polyketide synthase protein sequences from different species were analyzed, and phylogenetic tree was constructed (Fig. 3). *EVM0003759* is involved in ESC synthesis, namely *ESCB1* (Elsinochrome Biosynthesis gene 1).

To further analyze the differences in different polyketide synthase genes, used InterProScan to analyze the conserved domain of polyketide synthase in *E. arachidis*. Visualized the protein domain architecture,

the conserved domain of polyketide synthase was mapped using software DOG 2.0. The conserved domains of polyketide synthase in *E. arachidis* LNFT-H01 genome include AT, KS, DH, KR, MeT, ACP and ER(Fig. 4). Among them, *EVM0004732*, *EVM0003759* and *EVM0005880* all contain KS, AT, ACP and TE, and the distribution between different domains also as the same, only difference in sequence length. In addition to KS and AT, *EVM0005988*, *EVM0002563*, and *EVM0006869* contain two others domains, ER and KR, which have reducing activity. Therefore, they are further classified into non-reduced PKS (EVM0004732, EVM0003759, EVM0005880) and reduced PKS depending on the type of domain they contain. (EVM0005988, EVM0002563, EVM0006869).

RT-qPCR analysis of ESCB1

The biosynthesis of ESC in *E. arachidis* was significantly different in different light condition[47]. Under light condition, the production of ESC was 16 nmol/plug, while in the dark, no synthesis of toxins was detected. To further clarify whether *ESCB1* participate in the biosynthesis of ESC, the expression of *ESCB1* under different light conditions was examined. Results showed that the expression of *ESCB1* was the same as that of toxin production(Fig. 5).

Distribution of the ESCB1 **gene cluster of ESC**

Further analysis of the *ESCB1* cluster, 13 putative ORFs were identified including ESCB1 (Fig. 6). The EVM0001135 and EVM0007299 encode a putative polypeptide similar to O-methyltransferase and have a FAD binding domain which is involved in a number of enzymes. The EVM0006582 and EVM0006794 encode a product which similarity to major facilitator superfamily transporter. Cytochrome P450 (EVM0002495) and zinc finger transcription factor (EVM0002638) are also on this gene cluster. The specific biological functions of related functional genes in this gene cluster of ESC biosynthesis need to be further studied.

Discussion

Recent years, with supply-side structural reform, peanut planting area is increasing ranks second in the world and total output ranking first in the world. Peanut scab caused by *Elsinoë arachidis*, an important fungal disease with short incubation period and fast incidence rate, is highly eye-catching in the main peanut-producing areas of China. Recent trends in *E. arachidis* have led to a proliferation of studies about etiology, variety resistance evaluation, epidemiology and prevention measures[1, 2, 6, 52, 53]. Interestingly, previous studies demonstrated that *E. arachidis* can produced red, light-activated, nonhost-selective phytotoxin Elsinochrome (ESC), which affects the metabolism of host plants. Investigating the role and the mode of mycotoxins in the pathogenesis of pathogen, biosynthetic pathways and their regulatory mechanisms are of great significance for revealing pathogenic mechanisms and interaction mechanisms between pathogen and plant.

In this study, we first revealed whole genome sequence of *E. arachidis* which laying a foundation for further exploring the pathogenic mechanism of *E. arachidis* and make more contribution to research

pathogenicity-associated genes, additionally, to detection of ESC toxin biosynthesis pathways. In the annotation of the genome function of *E. arachidis*, a total of 8644 encoded protein sequences were annotated (94.22%), of which GO analysis, KEGG analysis and KOG analysis were annotated to 3237, 3055 and 4958 genes, respectively. KEGG analysis provides a research platform for clarifying the secondary metabolic pathways of *E. arachidis*. The higher percentages are secondary metabolic biosynthesis and carbohydrates. The substance metabolism in *E. arachidis* is relatively active, including not only the formation of amino acids and other nutrients. There are some secondary metabolites synthesized, in addition to ketones associated with the synthesis of toxins from *E. arachidis*, and transporters associated with the transport of substances. The KOG annotation and GO annotation lipid transport and metabolism, secondary metabolite biosynthesis, transport and catabolism include functional genes, and more genes involved. Functional annotation of the gene provides a data base for further study of the synthesis and metabolism of toxins in *E. arachidis*.

ESCs are perylenequinones photosensitive toxins, which can produce a large amount of reactive oxygen species (ROS) under light condition and act on the cell membrane and destroy its structure. *E. arachidis* can also grow and develop under the circumstance of high toxin production, indicating that it has certain detoxification. 57 ABC superfamily transporter genes, 190 MFS superfamily transporter genes and 78 cytochrome P450 enzymes were obtained from the genome analysis of *E. arachidis*. ABC transporter is a multi-component major active transporter with ATP-binding region, which can be transported by ATP under the conditions of active transport, including transmembrane transport of various molecules including small molecules and macromolecules, while MFS transporter is A single polypeptide-assisted vector capable of transporting small molecules in response to chemically permeable ion gradients[54–55]. Cytochrome P450 enzyme system is a multifunctional oxidoreductase[56], and it is speculated that the three may participate in the self-detoxification of *E. arachidis*, and reduce the effect of toxins on themselves by secreting toxins in vitro and redox and other physiological reactions. In addition, gene annotation includes detoxification genes and antioxidant genes related to self-detoxification, which provides ideas and references for further research on self-tolerance to ESC of *E. arachidis*.

So far, the molecular mechanism of biosynthesis of cercosporin is the most studied and characterized. The CTB gene cluster centered on the gene *CTB1* (cercosporin synthase gene 1) encoding polyketide synthase has confirmed that the biosynthesis of the toxin is synthesized by the polyketide pathway[59]. Similarly, the biosynthesis of ESC in *E. fawcettii* has been shown regulated by the polyketide synthase gene *Efpks1*, but its specific synthetic pathway remains to be further studied[60–61]. PKS includes AT, KS, DH, ER, MeT, ACP and KR domains. KR and ER are reductive enzymes, according to the domain-containing, EVM0005988, EVM0002563 and EVM0006869 are reduced PKS, EVM0004732, EVM0003759 and EVM0005880 are non-reducing PKS, containing KS, AT and ACP domains. Five PKS and PKS-NRPS were divided into melanin, elsinochrome and T-toxin, respectively. EVM0003759 is involved in elsinochrome synthesis and was named ESCB1. EVM0004732 and EVM0005880 were involved in melanin synthesis, EVM0005988, EVM0002563 and EVM0006869 were involved in T-toxin. Interestingly, by comparing the positions of the genome sequences of the six gene clusters, it was found that *ESCB1*, *EVM0004732* and *EVM0005880* are located in Contig00003. It is speculated that the positional

relationship between the three may be related to the mechanism of melanin and ESC production by *E. arachidis*. This also provides new ideas for the study of ESC secondary metabolic synthesis pathways. The gene cluster contains 13 predicted genes including *ESCB1*, including major facilitator superfamily transporter, cytochrome P450, monooxygenase and O-methyltransferase. For the specific role of these protein-coding genes in the synthesis of ESC in *E. arachidis*, the research team further verified the gene functions on the gene clusters one by one. The prediction of gene function on gene clusters helps to clarify the function of related genes and further clarify the ESC biosynthetic pathway.

Conclusions

In conclusion, this is the first report of the high-quality genome of *E. arachidis* by PacBio RS II. The basic information of the sequence, gene family and metabolic gene cluster of *E. arachidis* were clarified. Through further analysis of the key genes in different PKS gene clusters, the expression of *ESCB1* (EVM0003759) under light and dark condition was initially determined to participate in the ESC biosynthetic pathway, and the flanking sequences of this gene cluster were annotation, including major facilitator superfamily transporter, cytochrome P450, monooxygenase and O-methyltransferase. In addition to ESC toxins, genes related to mycotoxin biosynthesis such as melanin are also noted. This information provides new ideas for further exploration of the pathogenic mechanism of *E. arachidis*.

Materials And Methods

Strain and culture conditions

The strain of *Elsinoë arachidis* used in this study was LNFT-H01 which have been previously identified. LNFT-H01 was subcultured for purification by single spores and cultured on potato dextrose agar (PDA) under continuous fluorescent light (5 microeinstein (μE) $m^{-2}s^{-1}$).

Genome Sequencing And Assembly

A 20Kb library was constructed for the sequencing *Elsinoë arachidis* LNFT-H01 strain genomic DNA which performed by PacBio RS II (Biomarker Technologies)to ensure that the sequencing depth was not less than 100 ×. The genome sequence was assembled by Canu [18–20] software.

Phylogenetic And Syntenic Analysis

The protein sequences predicted by *E. arachidis* LNFT-H01 and the reference genome of *E. arachidis* were subjected to family clustering using OrthoMCL software, and then single copy gene were extracted. The phylogenetic tree was constructed using PhyML software by single-copy gene sequences of *E. arachidis* LNFT-H01 and *E. australis*. [21, 22]

The protein sequence of *E. arachidis* LNFT-H01 is BLAST aligned with the protein sequence of *E. australis*, collinearity relationship at the nucleic acid level is obtained based on the positional information of the homologous gene on the genomic sequence. Using the software MCScanX to map *E. arachidis* LNFT-H01 to *E. australis* in pairs according to the collinearity.[23]

Repetitive Sequences And DNA Methylation Analysis

Since the conservation of repetitive sequences between species is relatively low, the prediction of repetitive sequences for specific species requires the construction of a specific repetitive sequence database. Therefore, we use LTR_FINDER [24], MITE-Hunter [25], RepeatScout [26], PILER-DF [27] to construct a repetitive sequence database of sequencing data genomes based on the principles of structural and de novo prediction, using PASTEClassifier [28] classify the database, then merge it with Repbase [29] as the final repetitive sequence database, use RepeatMasker [30] software to predict the sequence of the repeated data based on the constructed repetitive sequence database.

Based on the kinetic information generated by the DNA polymerase synthesis reaction, SMRT three-generation sequencing technology can directly identify DNA methylation modification sites [31].

Gene Prediction And Annotation

The genetic structure of the sequencing data is predicted by using different strategies such as de novo prediction and homologous species prediction, and then using EVM software to integrate the prediction results. Augustus [32], GeneID [33], Genscan [34], Glimmer HMM [35], SNAP [36] for de novo prediction; GeMoMa [37] for homologous species-based prediction; finally, EVM [38] integration of the above methods. Gene function was annotated with Pfam[39], NR, KOG/COG,[40] KEGG[41] and GO[42] databases.

Protein Family Classifications And Function Characterization

In order to further analyze the protein family in *E. arachidis*, the Carbohydrate active enzymes database (http://www.cazy.org/) [43], transporter classification database (TCDB) [44] and pathogen-host interaction database [45] were used to analyze CAZymes and other important gene families for further analysis.

Secondary Metabolite Cluster Identification, Characterization And Visualization

Secondary metabolite clusters in *E. arachidis* genome sequence were analyzed by using antiSMASH2 (https://fungismash.secondarymetabolites.org). In addition, PKS, for polyketide synthase clusters were identified by screening the InterProScan (http://www.ebi.ac.uk/Interpro/).

Phylogenetic and domain analyses of fungal polyketide synthases (PKS).

Fungal polyketide synthase (PKSs) have ketosynthase (KS), acyl transferase (AT), and acyl carrier protein (ACP) domain, meanwhile, ketoreductase (KR), enoyl reductase (ER), and methyltransferase (MT) also exist[46].

ESCB1, a polyketide synthase in *E. arachidis* was identified in NCBI. Then, the conservative domains of the ESCB1 were identified using InterProScan and the visualize of the protein domain architectures was printed by software DOG 2.0. Phylogenetic tree was constructed by MEGA 10.0.5, which based on polyketide synthase protein sequences from different organisms with 1000 replicates bootstrap analysis. The GI number of the polyketide synthase is shown in table S1.

Measurements of toxin contents and Expression of ESCB1

Elsinochrome was extraction and quantitative analysis as before[47]. To estimate the expression of *ESCB1* transcripts at different light condition, the strain was cultured on PDA under dark conditions for 7 days, then take part to light condition for 48 h and isolated total RNA, respectively. ESCB1-F (ATCCGAGGTCATTGGTGATG)/ ESCB1-R(GAGGTTGACATCTGGCATTTG) were used to study the *ESCB1* expression.

Abbreviations

ESC: elsinochromes; PKS: polyketide synthase; NRPS: nonribosomal peptide synthetase; ABC: ATP-binding cassette superfamily transporter; MFS: Major Facilitator Superfamily transporter

Declarations

Competing interests

The authors declare that they have no competing interests.

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Funding

Natural Science Foundation of Liaoning Province (CN) (2019-MS-278)

Authors' contributions

Designed the experiments: Zhou Rujun, Jiao Wenli. Performed the experiments: Jiao Wenli, Xu Mengxue, Zhou Rujun, Liu Lu. Analyzed data: Jiao Wenli, Xu Mengxue, Zhou Rujun. Wrote manuscript: Jiao Wenli, Zhou Rujun.

All authors have read and approved the final version of the manuscript.

Availability of data and materials

We have submitted our data in NCBI (http://www.ncbi.nlm.nih), the accession number is:JAAPAX000000000.

Acknowledgements

We thank Natural Science Foundation of Liaoning Province (CN) (2019-MS-278) supported.

References

1. Zhou RJ, Xu Z, Fu JF, Cui JC, He JJ, Xue C. Y. Resistance evaluation of peanut varieties to peanut scab and the epidemic dynamics in Liaoning Province. Acta Phytophylacica Sinica. 2014;41:597–601.
2. Fang SM, Wang ZR, Ke YQ, Chen YS, Huang CM, Yu J. X. The Evaluation of Resistance and Resistant Mechanisms of Peanut Varieties to Scab Disease. Scientia Agricultura Sinica. 2007;40:291–7.
3. Fan XL, Barreto RW, Groenewald JZ, et al. Phylogeny and taxonomy of the scab and spot anthracnose fungus *Elsinoë* (Myriangiales, Dothideomycetes). Studies in Mycology. 2017; 87, 1–41.
4. Zhao JF, Zhou RJ, Li YJ, Lu L, Fu JF, Xue C. Y. Infectious condition of *Sphaceloma arachidis* and physiological responses to pathogen infection in peanut. Chinese Journal of Oil Crop Sciences. 2017; 1.
5. Liu L, Jiao WL, Zhou RJ, Li YJ, Xu MX, Fu JF. Extraction Technology and Activity Analysis of *Elsinoë arachidis* Toxin. Journal of Shenyang Agricultural University. 2018;49:272–8.
6. Fang SM, Wang ZR, Guo JM. Fungicides selection for peanut scab disease. Chinese Journal of Oil Crop Sciences. 2006;28:220–3.
7. Weiss U, Flon H, Burger WC. The photodynamic pigment of some species of *Elsinoë*, and *Sphaceloma*. Archives of Biochemistry & Biophysics. 1957; 69, 311–319.
8. Weiss U, Merlini L, Nasini G. Naturally occurring perylenequinones, Fortschritte der Chemie organischer Naturstoffe/Progress in the Chemistry of Organic Natural Products. Springer Vienna. 1987; 1–71.

9. Daub ME. Cercosporin, a photosensitizing toxin from *Cercospora* species. Phytopathology. 1982;72:370–4.
10. Stack ME, Mazzola EP, Page SW, et al. Mutagenic perylenequinone metabolites of alternaria alternata: altertoxins i, ii, and iii. J Nat Prod. 1986;49:866–71.
11. Arnone A, Assante G, Modugno VD, Merlini L, Nasini G. Perylenequinones from cucumber seedlings infected with *cladosporium cucumerinum*. Phytochemistry. 1988;27:1675–8.
12. Goodwin SB, Dunkle LD. Cercosporin production in Cercospora and Related Anamorphs. In: Lartey RT, Weiland JJ, Panella L, Crous PW, and Windels, C.E. editors. Cercospora Leaf Spot of Sugar Beet and Related Species. St. Paul: MN:APS Press. p.; 2010. pp. 97–108.
13. Dong Y, Li Y, Zhao M, et al. Global Genome and Transcriptome Analyses of *Magnaporthe oryzae* Epidemic Isolate 98 – 06 Uncover Novel Effectors and Pathogenicity-Related Genes, Revealing Gene Gain and Lose Dynamics in Genome Evolution. PLOS Pathogens. 2015;11:e1004801.
14. Walkowiak S, Rowland O, Rodrigue N, et al. Whole genome sequencing and comparative genomics of closely related Fusarium Head Blight fungi: *Fusarium graminearum*, *F. meridionale* and *F. asiaticum*. BMC Genomics. 2016; 17, 1014.
15. Ohm RA, Feau N, Henrissat B. Diverse Lifestyles and Strategies of Plant Pathogenesis Encoded in the Genomes of Eighteen Dothideomycetes Fungi. PLoS Pathog. 2012;8:e1003037.
16. Li X, Gao C, Li L, et al. MoEnd3 regulates appressorium formation and virulence through mediating endocytosis in rice blast fungus *Magnaporthe oryzae*. PLoS Pathog. 2017;13:e1006449.
17. Amselem J, Cuomo CA, van Kan JA, et al. Genomic Analysis of the Necrotrophic Fungal Pathogens *Sclerotinia sclerotiorum* and *Botrytis cinerea*. PLoS Genet. 2011;7:e1002230.
18. Chin CS, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. Nature methods. 2013;10:563–9.
19. Berlin K, Koren S, Chin CS, Drake JP, Landolin JM, Phillippy AM. Assembling large genomes with single-molecule sequencing and locality-sensitive hashing. Nature biotechnology. 2015;33:623–30.
20. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Research. Genome; gr.215087.116v2. 2017.
21. Li, OrthoMCL L. Identification of Ortholog Groups for Eukaryotic Genomes. Genome Res. 2003;13:2178–89.
22. Vincent L, Jean-Emmanuel L, Olivier GSMS. Smart Model Selection in PhyML. Molecular Biology and Evolution. 2017; 9.
23. Wang Y, Tang H, Debarry JD, et al. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic Acids Res. 2012;40:e49–9.
24. Xu Z, Wang H. LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. Nucleic Acids Res. 2007;35:W265–8.

25. Han Y, Wessler SR. MITE-Hunter: a program for discovering miniature inverted-repeat transposable elements from genomic sequences. Nucleic Acids Res. 2010;38:199.
26. Price AL, Jones NC, Pevzner PA. De novo identification of repeat families in large genomes. Bioinformatics. 2005;21:351–8.
27. Edgar RC, Myers EW. PILER: identification and classification of genomic repeats. Bioinformatics. 2005;21:152–8.
28. Wicker T, Sabot F, Hua-Van., et al. A unified classification system for eukaryotic transposable elements. Nat Rev Genet. 2007;8:973–82.
29. Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, Walichiewicz J. Repbase Update, a database of eukaryotic repetitive elements. Cytogenet Genome Res. 2005;110:462–7.
30. Tarailo-Graovac M, Chen N. Using RepeatMasker to Identify Repetitive Elements in Genomic Sequences. Current protocols in bioinformatics. Chapter 4, Unit 4.10. 2009.
31. Flusberg BA, Webster DR, Lee JH, et al. Direct detection of DNA methylation during single-molecule, real-time sequencing. Nat Methods. 2010;7:461–5.
32. Stanke M, Waack S. Gene prediction with a hidden Markov model and a new intron submodel. Bioinformatics. 2003;19:215–25.
33. Blanco E, Parra G, Guigó R. Using geneid to Identify Genes, Current Protocols in Bioinformatics. 2007; 18.
34. Burge C, Karlin S. Prediction of complete gene structures in human genomic DNA. Journal of molecular biology. 1997;268:78–94.
35. Majoros WH, Pertea M, Salzberg SL. TigrScan and GlimmerHMM: two open source ab initio eukaryotic gene finders. Bioinformatics. 2004;20:2878–9.
36. Korf I. Gene finding in novel genomes. BMC Bioinformatics. 2004;5:59.
37. Jens K, Frank, Hartung, Jan, and Grau. Gemoma: homology-based gene prediction utilizing intron position conservation and rna-seq data. Methods in molecular biology (Clifton, N.J.). 2019.
38. Haas BJ, Salzberg SL, Zhu W, Mihaela P. Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. Genome Biol. 2008;9:R7.
39. Finn RD, Bateman A, Clements J, et al. Pfam: the protein families database. Nucleic Acids Res. 2014;42:D222-30.
40. Tatusov RL, Galperin MY, Natale DA, Koonin EV. The COG database: a tool for genome-scale analysis of protein functions and evolution. Nucleic Acids Res. 2000;28:33–6.
41. Kanehisa M, Goto S, Kawashima S, Okuno Y, Hattori M. The KEGG resource for deciphering the genome. Nucleic Acids Res. 2004;32:D277-80.
42. Ashburner M, Ball CA, Blake JA. Gene Ontology: tool for the unification of biology. Nat Genet. 2000;25:25–9.
43. Vincent L, Hemalatha GR, Elodie D, Coutinho PM, Bernard H. The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Research. 2013.

44. Saier MH, Tran CV, Barabote RD TCDB: The Transporter Classification Database for membrane transport protein analyses and information. Nucleic Acids Research. 2006; suppl_1, D181-D186.
45. Winnenburg R, Baldwin TK, Urban M, et al. PHI-base: a new database for pathogen host interactions. Nucleic acids research. 2006;34:D459–64.
46. Kroken S, Glass NL, Taylor JW, Yoder OC, Turgeon BG. Phylogenomic analysis of type I polyketide synthase genes in pathogenic and saprobic ascomycetes. Proceedings of the National Academy of Sciences. 2003; 100, 15670–15675.
47. Jiao WL, Liu L, Zhou RJ, Xu MX, Xiao D., and Xue CY. Elsinochrome phytotoxin production and pathogenicity of Elsinoë arachidis isolates in China. PLoS ONE. 2019;14:e0218391.
48. Mehrotra S, Goyal V. Repetitive Sequences in Plant Nuclear DNA: Types, Distribution, Evolution and Function. Genomics Proteomics Bioinformatics. 2014;12:164–71.
49. Lisch D. How important are transposons for plant evolution? Nat Rev Genet. 2012;14:49–61.
50. Tamaru H, Selker EU. A histone H3 methyltransferase controls DNA methylation in *Neurospora crassa*. Nature. 2001;414:277–83.
51. Jaenisch R, Bird A. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. Nat Genet. 2003;33:245–54.
52. Xiao D, Xue CY, Zhou RJ, Fu JF, Li YJ, Liu L. Biological characteristics of *Elsinoë arachidis* causing peanut scab. Chinese Journal of oil crop sciences. 2018;40:134–9.
53. Xue CY, Fu JF, Zhou RJ, et al. A technique for isolation of *Sphaceloma arachidis* from infected peanut. Chinese Journal of Oil Crop Sciences. 2017;39:386–92.
54. Deng D, Yan N. Structural basis and transport mechanism of the major facility superfamily (MFS) transporter. Chin Sci Bull. 2015;60:720–8.
55. Hollenstein K, Frei DC, Locher KP. Structure of an ABC transporter in complex with its binding protein. Nature. 2007;446:213–6.
56. Kimura M, Tokai T, Takahashi-Ando N, Ohsato S, Fujimura M. Molecular and genetic studies of fusarium trichothecene biosynthesis: pathways, genes, and evolution. Bioscience Biotechnology Biochemistry. 2007;71:2105–23.
57. Brink JVD, Vries RP. D. Fungal enzyme sets for plant polysaccharide degradation. Appl Microbiol Biotechnol. 2011;91:1477–92.
58. Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B. The carbohydrate-active enzymes database (CAZy): an expert resource for glycogenomics. Nucleic Acids Research. 2009; 37, D233-D238.
59. Choquer M, Dekkers KL, Chen HQ, et al. The CTB1 gene encoding a fungal polyketide synthase is required for cercosporin biosynthesis and fungal virulence of *Cercospora nicotianae*. MPMI. 2005;18:468.
60. Daub ME, Herrero S, Chung KR. Photoactivated perylenequinone toxins in fungal pathogenesis of plants. FEMS Microbiol Lett. 2005;252:197–206.

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

Additional Files

FigS1 KOG annotation of *E. arachidis*.

FigS2 GO annotation of *E. arachidis*.

Table S1 PKS sequence information table.

Table S2 Repetitive DNA sequence.

Table S3. DNA methylation sequence.

Table S4 gene family analysis.

Figures

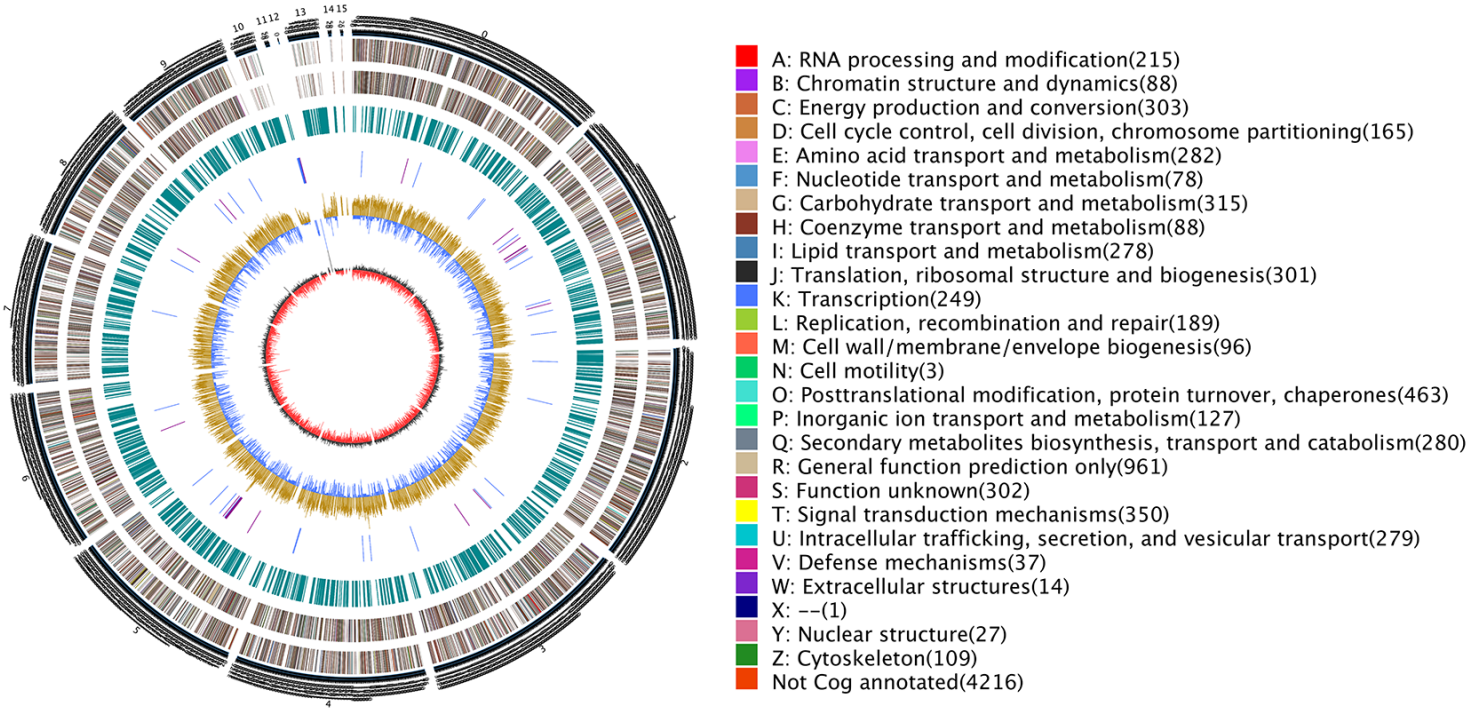

Figure 1

Circos-plot of E. arachidis. The outermost circle is the size of the genome, each scale is 5 Kb; the second and third circles are the genes on the positive and negative strands of the genome, respectively, the different colors represent different COG functional classifications; the fourth circle is repeated Sequence; the fifth circle is tRNA and rRNA, the blue is tRNA, the purple is rRNA; the sixth circle is GC content, the light yellow part indicates that the GC content of this region is higher than the average GC content of the

genome, and the higher the peak value is the average GC content. The greater the difference, the blue part indicates that the GC content of the region is lower than the average GC content of the genome; the innermost circle is GC-skew, the dark gray represents the region where the G content is greater than C, and the red represents the region where the C content is greater than G.

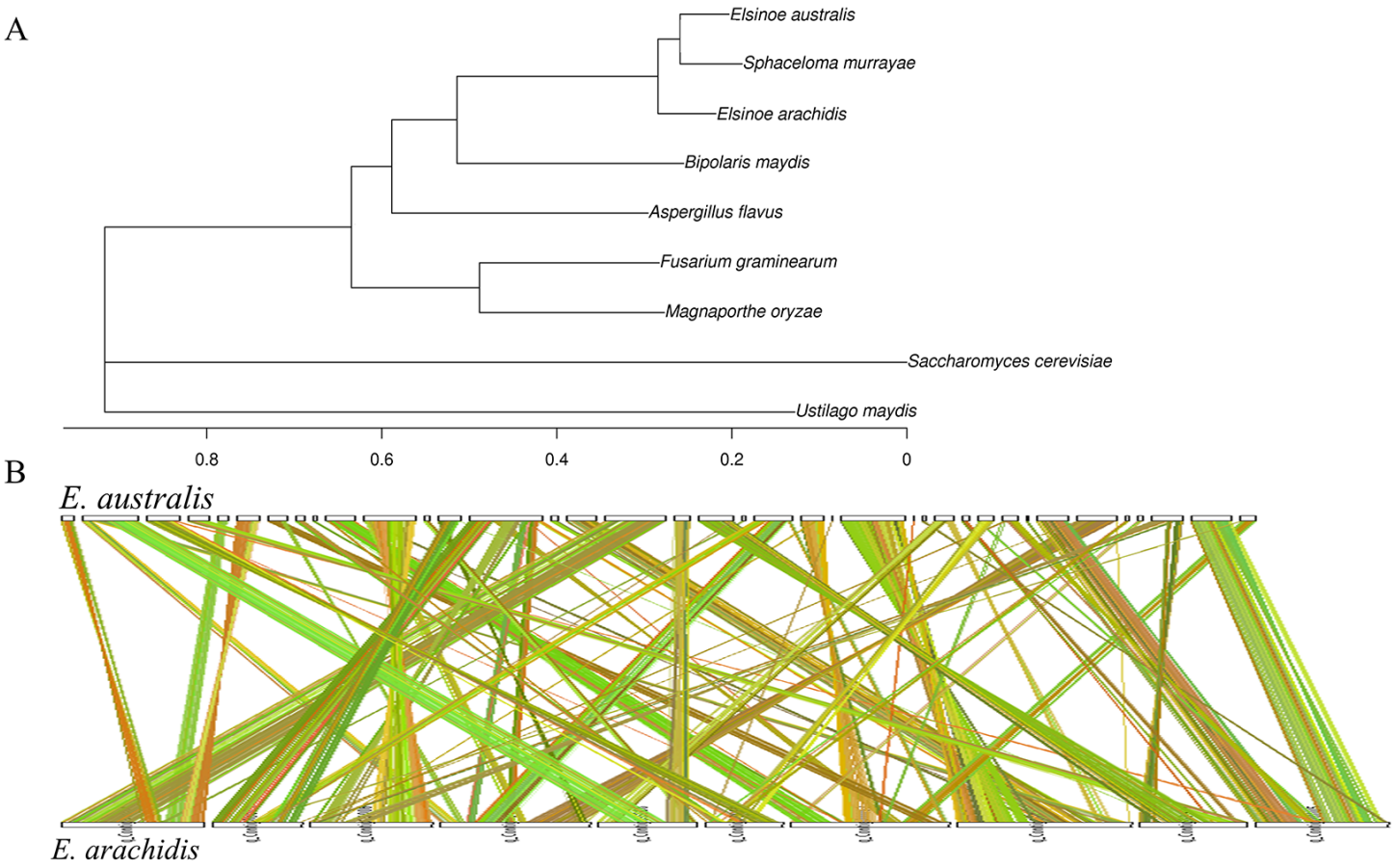

Figure 2

Collinear analysis and evolutionary analysis of E. arachidis. (A) A phylogenetic tree constructed the evolutionary relationships of E. arachidis and other fungi. (B) Collinear analysis.

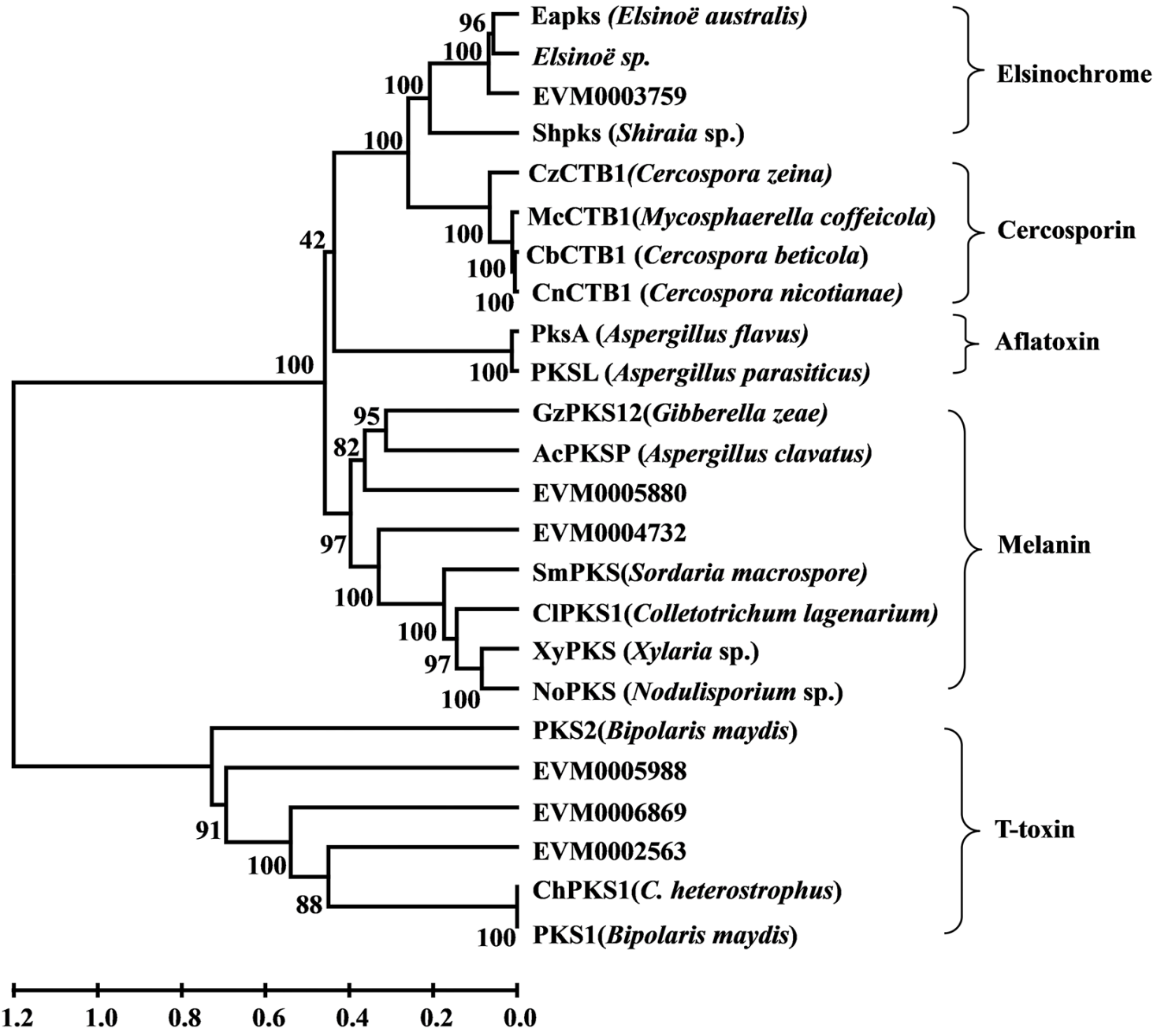

Figure 3

Phylogenetic analyses of E. arachidis and other fungal polyketide synthases (PKS). Phylogenetic tree was constructed with polyketide synthase protein sequences from different organisms which classified with the types of reducing domains are divided into five clades.

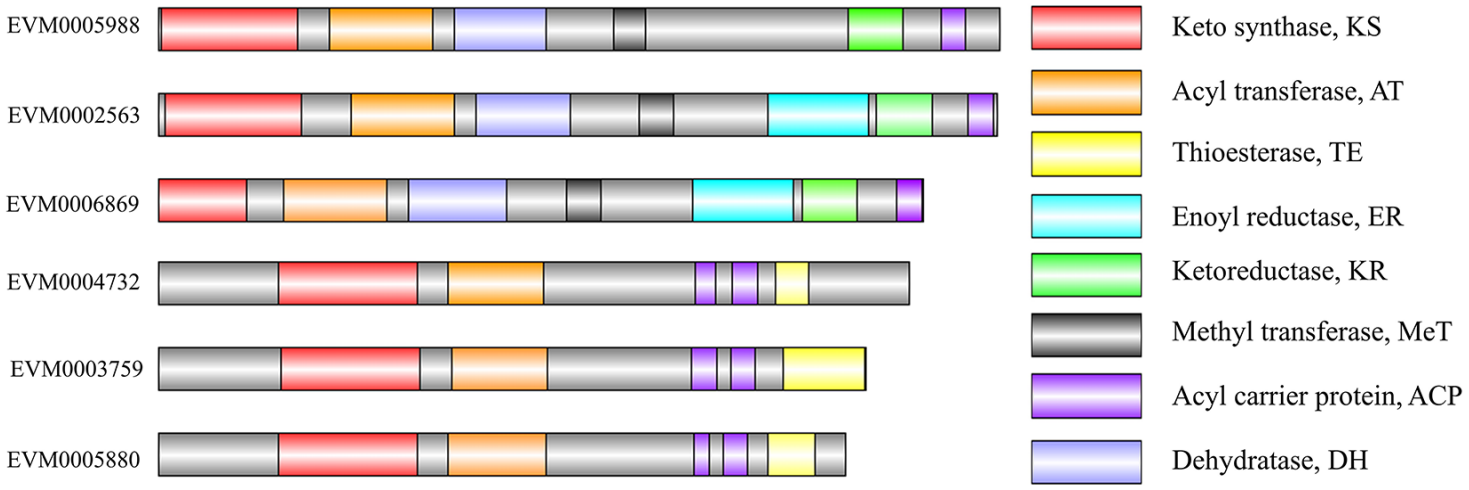

Figure 4

Structure of polyketide synthases proteins. The conservative domain of polyketide synthases was clarified by InterProScan, and visualization of the domain architectures by using software DOG 2.0.

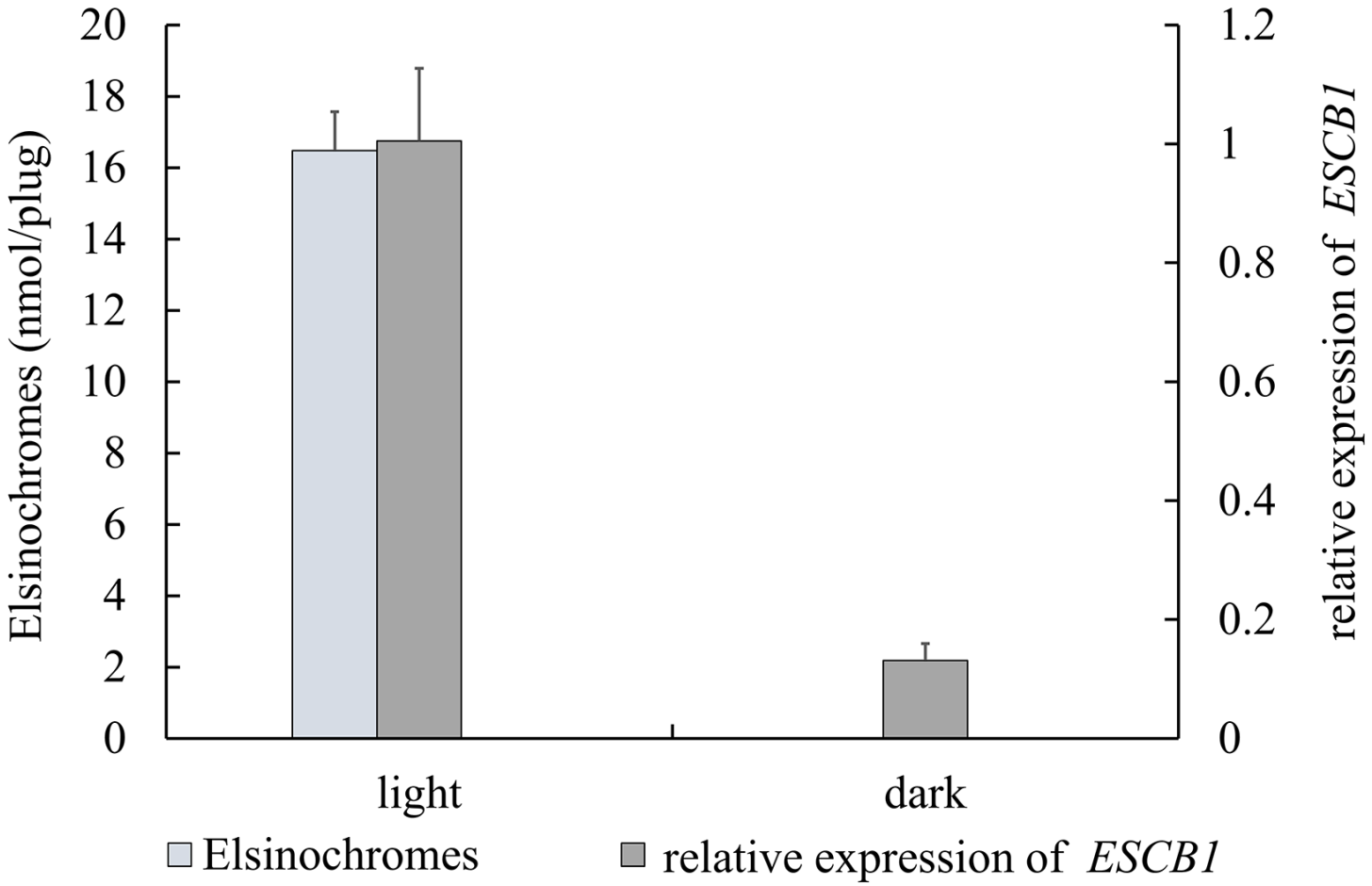

Figure 5

ESC and expression levels analysis of ESCB1. ESC and expression levels of ESCB1 was investigated in light and dark condition, respectively.

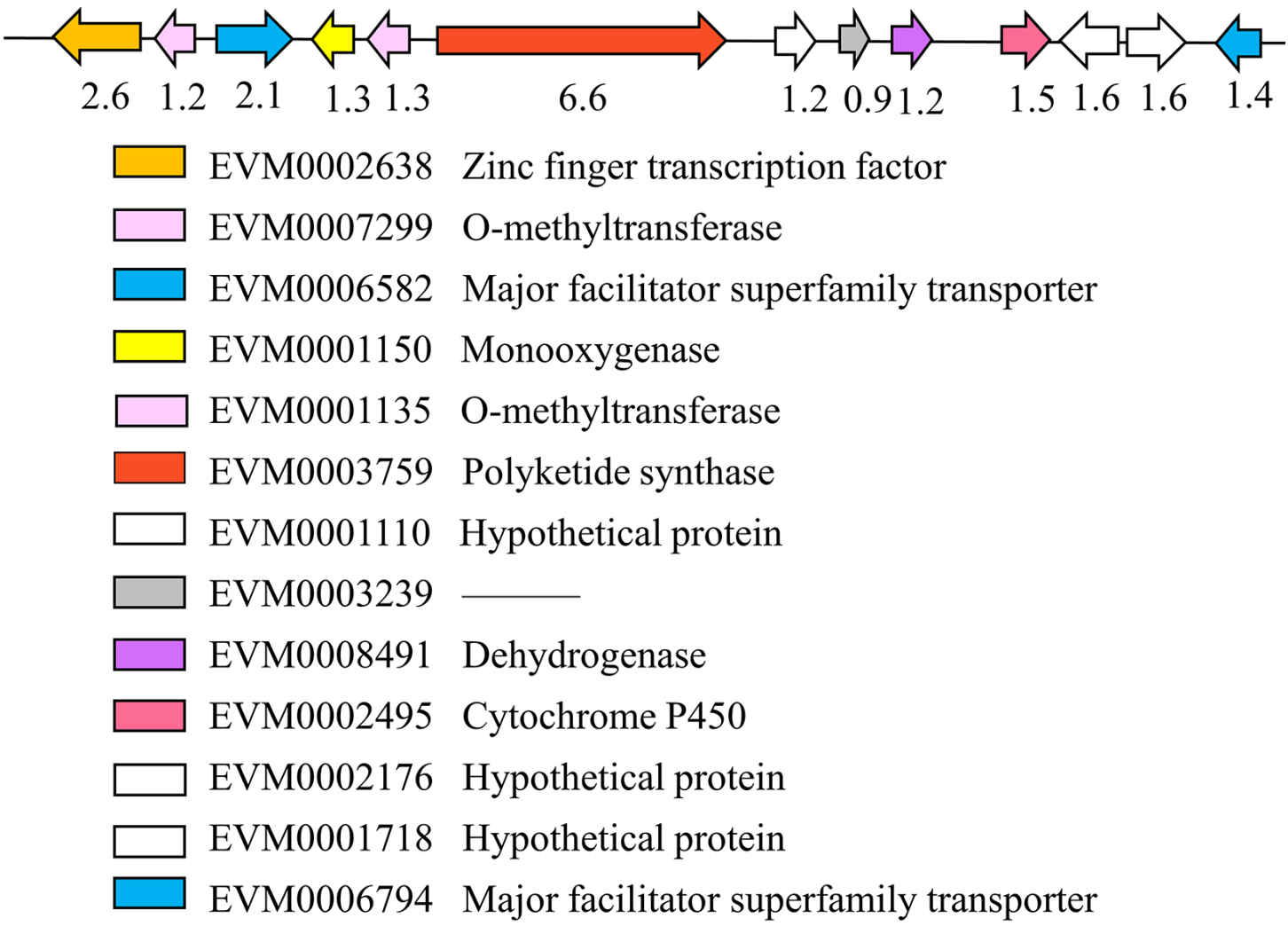

Figure 6

Distribution of the ESCB1 gene cluster. BLASTX was used to search the NCBI database to predict the function of related genes.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- FigS1KOG.tif
- TableS1PKSsequenceinformationtable.doc
- TableS3.DNAmethylation.doc
- TableS2RepetitiveDNAsequence.doc
- TableS4genefamily.xls

- FigS2GO.tif