Analysis of The Genome Sequence of Phomopsis Vexans: A Fungal Pathogen Causing Phomopsis Blight of Eggplant

Zhou Heng
Guangdong Key Laboratory for New Technology Research of Vegetables, Vegetable Research Institute, Guangdong Academy of Agricultural Sciences

Qian You
Guangdong Key Laboratory for New Technology Research of Vegetables, Vegetable Research Institute, Guangdong Academy of Agricultural Sciences

Zhiliang Li
Guangdong Key Laboratory for New Technology Research of Vegetables, Vegetable Research Institute, Guangdong Academy of Agricultural Sciences

Baojuan Sun
Guangdong Key Laboratory for New Technology Research of Vegetables, Vegetable Research Institute, Guangdong Academy of Agricultural Sciences

Xiaowan Xu
Guangdong Key Laboratory for New Technology Research of Vegetables, Vegetable Research Institute, Guangdong Academy of Agricultural Sciences

Ying Li
Guangdong Key Laboratory for New Technology Research of Vegetables, Vegetable Research Institute, Guangdong Academy of Agricultural Sciences

Zhenxing Li
Guangdong Key Laboratory for New Technology Research of Vegetables, Vegetable Research Institute, Guangdong Academy of Agricultural Sciences

Hengming Wang
Guangdong Key Laboratory for New Technology Research of Vegetables, Vegetable Research Institute, Guangdong Academy of Agricultural Sciences

Chao Gong
Guangdong Key Laboratory for New Technology Research of Vegetables, Vegetable Research Institute, Guangdong Academy of Agricultural Sciences

Li Tao (tianxing84@163.com)
Guangdong Key Laboratory for New Technology Research of Vegetables, Vegetable Research Institute, Guangdong Academy of Agricultural Sciences

Research Article

Keywords: Genome, Phomopsis vexans, Phomopsis blight of eggplant

DOI: https://doi.org/10.21203/rs.3.rs-156550/v1

Abstract

Background: *Phomopsis vexans* is a phytopathogenic fungus causing Phomopsis blight of eggplant. This disease is one of the major issues reducing eggplants production. To lay a solid foundation for the research of pathogenicity and to understand the mechanism of the disease development, the genome of an isolate PV4 was sequenced, assembled and analyzed from Guangdong, China.

Results: The assembled complete genome size of *Phomopsis vexans* was about 59.78 Mb with 51.24% G+C content and 4.93Mb contig N50. In the genome, 3,552 annotated repetitive elements were identified. A total of 15,034 genes with 1,790bp average length were predicted, of which 14,116 genes were annotated by NCBI nr database. Moreover, 1,206 genes were annotated from Carbohydrate-Active enzymes (CAZy) Database which may play a role in degrading plant cell walls. At last, 134 effector proteins were predicted.

Conclusions: The present genome is the first genome of *Phomopsis vexans*. The information obtained from this study support important resource for research on the pathogen and for sheding light on its pathogenicity mechanism.

Background

Phomopsis vexans is a phytopathogenic fungus that belongs to class Sordariomycetes, order Diaporthales, family Valsaceae[1] and causes Phomopsis blight of eggplant (*Solanum melongena*). [2]. It was first reported by Halsted in USA and has been spread widely in all eggplant growing countries[3–5]. The common symptoms of Phomopsis blight include leave spot, branch brown spots, withered and fruit rot[6]. It has caused significant yield loss of eggplants, especially at tropical and subtropical regions, where the climate is warm and humid[7]. So Phomopsis blight is one of the most economically important diseases of eggplants.

Several ways have been tried to control phomopsis blight including crop rotation, fungicide treatments, etc [2, 6, 8]. However, they are neither lack of efficiency nor caused environment and food safety problem. It is critical to understand its pathogenesis mechanism to develop efficient way for phomopsis blight management. To approach this goal, more information about *Phomopsis vexans* should be needed. Scientists have isolated, morphological and molecular characterized it[7, 9–11]. However, information about the genomic features of P. vexans which are indispensable for pathogenesis mechanism research is still lacking.

To provide information about genomic basis of *Phomopsis vexans*' pathogenicity, genome of isolate PV4 was sequenced, assembled and annotated. Potential pathogenesis mechanism was discussed. This is the first genome of *Phomopsis vexans*. It will shed light on its pathogenesis mechanism and help developing more efficient ways to manage Phomopsis blight of eggplants.

Methods

Isolation, identification, and cultivation of P.vexans isolate

A strain of *Phomopsis vexans* was isolated from infected branch of an eggplant (Solanum melongena) tree plant in Zhongluotan Town, Baiyun District, Guangzhou, Guangdong Province, China (23°23'24.5"N 113°26'19.4"E). Briefly, 1x0.5cm size of the disease-health junction fruit tissue was choosed, rinsed twice with sterile water, sterilized with sodium hypochlorite solution for 10 minutes, and then washed with sterile water three times, cultured on a PDA plate at 28°C.

Identification of Phomopsis vexans was first based on morphological characteristics and then was confirmed at molecular level. ITS region of PV4's DNA was amplified by PCR with primers ITS1, 5'-TCCGTAGGTGAACCTGCGG-3'and ITS4, 5-TCCTCCGCTTATTGATATGC-3'.

Genomic DNA extraction and sequencing

After 7 days culture, 100mg mycelia was collected from PDA plate, frozen with liquid nitrogen immediately, then grounded with a mortar and pestle. Genomic DNA was then extracted by QIAGEN Genomic-tip tool kit. Sample quality testing, library construction, library quality testing, library sequencing process were performed in accordance with the standard protocol provided by Oxford Nanopore Technologies (ONT). Library construction included the following steps: a. Nanodrop, Qubit and 0.35% agarose gel electrophoresis were used for purity, concentration, and integrity testing; b. use BluePippin automatic nucleic acid recovery system to recover large fragments of DNA; c. use SQK-LSK109 ligation kit to construct library; d. sequencing on Nanopore sequencing platform.

Data analysis and genome assembly

Clean data was obtained by trimming adapter sequences and poor quality bases for each sequence read. Canu v1.5[21] software was used to correct the errors of the filtered subreads, then wtdbg[22] software was employed to assemble the corrected subreads. Finally, Pilon[23] was used to further correct the assembled genome to get higher final accuracy data.

Repeat elements prediction

Due to the relatively low conservation of repetitive sequences between species, it is necessary to construct a specific repetitive sequence database when predicting repetitive sequences for specific species. Therefore, with the help of LTR_FINDER v1.05[24], MITE-Hunter[25], RepeatScout v1.0.5[26], PILER-DF v2.4[27], we build fungal genome repetitive sequence database. Use PASTEClassifier[28] to classify the database, and then merge with the database of Repbase[29] as the final repeated sequence database. Then RepeatMasker v4.0.6[30] was employeed to predict the repeat sequence of the fungus.

Gene prediction and annotation

Gene structure prediction mainly adopts ab initio prediction, prediction based on homologous protein and prediction based on transcriptome evidence, and then integrates the three prediction results. Genscan[31], Augustus v2.4[32], GlimmerHMM v3.0.4[33], GeneID v1.4[34], SNAP (version 2006-07-28)[35] were used for ab initio prediction. GeMoMa v1.3.1[36] was used for prediction based on homologous proteins. PASA v2.0.2 [37] was used to predict Unigene sequence based on transcriptome assembly. Finally, use EVM v1.1.1 （http://transdecoder.github.io） to integrate the prediction results obtained by the three above methods, and modify it with PASA v2.0.2. All predicted gene were annotated with NCBI non-redundant (Nr) protein, GO, KOG, KEGG, Swiss-Prot, TrEMBL Pfam and PHI（Pathogen Host Interactions） Database. Carbohydrate-Active enZYmes Database (http://www.cazy.org/) was used to predict putative proteins of the CAZy family. EffectorP v2.0[38] was used to identify potential effectors.

Results

General genome features

The genome of P. vexans isolate PV4 was sequenced by nanopore strand-sequencing. Statistics of genome sequencing and assembly are showed in **Table 1 and 2**. A total of 7,557,554,305 bp raw data was obtained, of which 7,284,861,206 bp clean data left after filtering low-quality data. The assembled complete genome was about 59.78 Mb with 51.24% G+C content and 121.85×Coverage. The contig N50 was 5,171,527 bp.

Repetitive elements identify

3,552 annotated repetitive elements were identified in the genome of PV4 (**Table 3**). The information about their annotation source, type, loci and attributes were shown in **Table S1**. They were generally classified in Class I (Retrotransposons, 738), Class II (DNA Transposons, 369), Potential host gene (455) and SSR (1,990). Major types were LTR/Copia (258) and MITE (218) in Class I and Class II separately. There were 4,931 unknown repetitive elements.

Gene prediction and annotation

As shown in **Table 4**, 15,034 genes with 26,919,018pb total length and 1,790bp average length were predicted. The total number of exons, CDS and Intro were 42,753, 27,719 and 42,689, separately. The average number of exons, CDS and Intros were 2.84, 1.84 and 2.84, respectively.

A total of 14,181 predicted genes were annotated with NCBI nr (14,116), GO (5,841), KEGG (3,729) and other four database **(Table S2)**. Moreover, 95 predicted genes did not significantly match any known genes.

Nr homologous species distribution analysis showed that PV4 had most homologous genes with Togninia minima (21.97%), Pestalotiopsis fici (7.45%) and Colletotrichum gloeosporioides (6.36%) **(Fig 1)**. Functional categorization and distribution of predicted genes by GO annotation are showed in **Fig. 2**. Distribution of annotated genes in KEGG database is shown in **Fig. 3.** Biosynthesis of amino acids (145), Carbon metabolism (124) and Ribosome (102) had the most annotated genes.

Carbohydrate-Active enzymes (CAZy) Database was employed to find genes encoding carbohydrate-activated enzymes which could be plant cell-wall degrading enzymes. 1,206 genes were annotated and separated in 6 CAZy type which was Auxiliary Activities (AAs, 253), Glycoside Hydrolases (GHs, 483), Glycosyl Transferases (GTs, 128), Polysaccharide Lyases (PLs, 43), Carbohydrate Esterases (CEs, 197) and Carbohydrate-Binding Modules (CBMs, 102) **(Table 5)**.

Pathogen-host interactions (PHI) database was used to find more information about genes related to pathogen-host interactions. Result was showed in **Table S3**. Protein subcellular location analysis predicted that there were 1,786 signal peptides, 3,223 transmembrane proteins, 1,394 secreted proteins, and 134 effector proteins (**Table S4**).

Discussion

The present study described general feature of *Phomopsis vexans'* gapless genome with third-generation sequencing technologies, Oxford Nanopore Technologies (ONT). The assembled complete genome was about 59.78 Mb with 51.24% G+C content and 4.93Mb N50 length. Compared with other reported *Phomopsis* (teleomorph *Diaporthe*) Pathogens, such as *Phomopsis longicolla* (62Mb, 1.04Mb N50length)[12], *Phomopsis phragmitis* (58.33 Mb, 3.55Mb N50length)[13], *Phomopsis obscurans* (48 Mb, 1.3kb N50length)[14] and *Diaporthe citri* (52.06 to 63.61 Mb, 13.9 kb N50length)[15], *Phomopsis vexans* has moderate size and the largest N50length. This means the present genome is well sequenced and assembled, and will provide a useful resource for the study of fungal biology, pathogen-host interaction or other aspect of *Phomopsis vexans* pathogenesis mechanism.

Repetitive elements are critical factors to determine the genome architecture and drive evolution and host adaptation of fungal genome. Here, we identified 4.84% repetitive elements, which is less than *Phomopsis longicolla* (13%)[12], but great than *Phomopsis phragmitis* (2.12%)[13]. The most important category of repetitive elements is transposable elements (TEs) which is 1.5% in *Phomopsis vexans* genome. TEs are mainly separated into two classes based on structural features and mode of transposition: Class I TEs (retrotransposons), function by reverse transcription and propagate through copy-and-paste mechanism; Class II TEs (DNA transposons), propagate through cut-and-paste mechanism[16, 17]. In this study, the majority of TEs was Class I (1.4% of whole genome), in which were long terminal repeats (LTRs), Copia

had the largest number. Class II formed a minor fraction (0.1%) of the this genome. This is consistent with previous studies[12, 18].

Passing through the plant cell wall is the first step for plant pathogen's infection[12]. Plant cell wall degrading enzymes (PCWDEs) are a subset of carbohydrate-activated enzymes (CAZy) that are produced by plant pathogens to degrade plant cell walls[19]. A total of 1,206 genes were annotated to be CAZy coding genes. In other study, the number of phomopsis pathogens were ranged from 778 to 1,702[12-15]. This result indicated that *Phomopsis vexans* has a lot of genes encoding PCWDEs.

Pathogen need effector proteins to suppress the host immune response or manipulate host cell physiology[20]. Here, we identified 134 effector proteins predicted by EffectorP. There is no effector prediction in other research on phomopsis spp. Moreover, there is no validated effector in Phomopsis spp. Since effector prediction is available now, more research was needed to character their function.

Conclusion

Phomopsis blight of eggplant is one of the most serious diseases affecting eggplant production around the world. Research about its pathogenesis mechanism is critical for its management. However, the genomic information of *Phomopsis vexans* was blank. Here, we reported the first genome sequence of *Phomopsis vexans*. Repetitive elements, carbohydrate-activated enzymes, effectors and other genes were identified and annotated. The present study provided the genome sequence information of *Phomopsis vexans*, which will be helpful to understand the pathogenicity mechanism of this disease.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

Raw data of the genome is available at: https://www.ncbi.nlm.nih.gov/bioproject/PRJNA694791

Competing interests

The authors declare that they have no competing interest

Funding

Natural Science Foundation of Guangdong Province (2020A1515011330)，Dean Fund of Guangdong Academy of Agricultural Sciences (201812)，Modern Agricultural Innovation Team Project of Guangdong Province (2020KJ110, 2020KJ106)，Guangdong Academy of Agricultural Sciences Foundation (201630TD)，Foreign cooperation projects of Guangdong Province（2018A050506052），National Modern Agricultural Industry Technology Innovation Center Creation Project in Guangzhou（2018kczx06）

Author's contributions

TL designed the experiments. ZH and QY analyzed the data and wrote the paper. ZhiliangL, ZhengingL, YL, HW managed the eggplant field and sampled fruits. XX, BS and CG isolated the *Phomopsis vexans*. All authors have read and approved the final manuscript.

Acknowledgements

We would like to thank Xiaoqing Sun and Junlin Huang for fungi sample collection and helpful discussions.

References

1. Udayanga D, Liu X, Mckenzie E H C, et al. The genus Phomopsis: biology, applications, species concepts and names of common phytopathogens [J]. Fungal Diversity, 2011, 50(1): 189–225.
2. Kalda T S, Swarup V, Choudhury B. Studies on resistance to Phomopsis blight in eggplant (Solanum melongena L.) [J]. Vegetable Science, 1976, (3): 65–70.
3. Halsted B D. A New Egg-Plant Disease [J]. Bulletin of the Torrey Botanical Club, 1891, 18(10): 302–3.
4. Harter L. Fruit rot leaf spot and stem blight of the eggplant caused by *Phomopsis vexans* [J]. journal of Agricultural Research, 1914, 2(5): 331–8.
5. Jameel Akhtar and H.S. Chaube. Variability in phomopsis blight pathogen [*Phomopsis vexans* (Sacc. & Syd.) Harter] [J]. Indian Phytopathology, 2006, 59(4): 439–44.
6. Mahadevakumar S, Janardhana GR. *Phomopsis vexans* Current Research and Future perspectives1914-2015 [J]. Journal of Botanical Sciences, 2016, S): 1–9.
7. Rohini, HG. Gowtham, P. Hariprasad and SR. Niranjana. Cultural, morphological and physiological characterization of *Phomopsis vexans* isolates from different brinjal (solanum melongena L.) growing regions of karnataka, india [J]. International journal of Pharma and bio Science, 2016, 7(3): 70–7.
8. Kalda T S. Resistance to Phomopsisblight in eggplant [J]. Vegetable Science, 1977, 4(2): 90–101.
9. Islam S J, Sitansu P. Variabilities among isolates of Phomopsis vexans[J]. Environment and ecology, 1990, 8(1B): 315–319.
10. Udayashankar A C, Chandra Nayaka S, Archana B, et al. Specific PCR-based detection of *Phomopsis vexans* the cause of leaf blight and fruit rot pathogen of Solanum melongena L [J]. Letters in Applied Microbiology, 2019, 69(5): 358–65.
11. Bhat M, Anwar A, Mughal M N, et al. Morpho-cultural and pathological variability in *Phomopsis vexans* causing leaf blight and fruit rot of brinjal in Kashmir [J]. Indian Phytopathology, 2019, 72(2): 225–33.
12. Shuxian L, Omar D, Nadim W A, et al. Analysis of the genome sequence of Phomopsis longicolla: a fungal pathogen causing Phomopsis seed decay in soybean [J]. BMC genomics, 2017, 18(1):
13. Xiaojie W, Hongjie D, Jianbin L, et al. High-Quality Genome Resource of the Pathogen of Diaporthe (Phomopsis) phragmitis Causing Kiwifruit Soft Rot [J]. 2020,
14. Udayanga D, Castlebury L. Molecular systematics and comparative draft genome analysis of strawberry leaf blight fungus [J]. 2017,
15. Gai Y, Xiong T, Xiao X, et al. The Genome Sequence of the Citrus Melanose Pathogen Diaporthe citri and Two Citrus related Diaporthe species [J]. Phytopathology, 2020, ja):
16. David J W, Austen R D, Carolyn A Y, et al. Repeat elements organise 3D genome structure and mediate transcription in the filamentous fungus Epichloe festucae [J]. PLoS genetics, 2018, 14(10): e1007467.
17. Johannes W, Anne K. Repetitive DNA elements in fungi (Mycota): impact on genomic architecture and evolution [J]. Current genetics, 2002, 41(4): 189–98.

18. Ling Y, Yunhe A, Junjie Q, et al. Genome sequence of Plasmopara viticola and insight into the pathogenic mechanism [J]. Scientific reports, 2017, 7(1):
19. Aylward J, Steenkamp E T, Dreyer L L, et al. A plant pathology perspective of fungal genome sequencing [J]. IMA Fungus, 2017, 8(1): 1–15.
20. Mareike M, Eva H S. Evolution and genome architecture in fungal plant pathogens [J]. Nature Reviews Microbiology, 2017, 15(12): 756–71.
21. Koren S, Walenz B P, Berlin K, et al. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation [J]. Genome research, 2017, 27(5): 722–36.
22. Ruan J, Li H. Fast and accurate long-read assembly with wtdbg2. bioRxiv [M]. jan. 2019.
23. Walker B J, Abeel T, Shea T, et al. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement [J]. PloS one, 2014, 9(11): e112963.
24. Xu Z, Wang H. LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons [J]. Nucleic acids research, 2007, 35(suppl_2): W265-W8.
25. Han Y, Wessler S R. MITE-Hunter: a program for discovering miniature inverted-repeat transposable elements from genomic sequences [J]. Nucleic acids research, 2010, 38(22): e199-e.
26. Price A L, Jones N C, Pevzner P A. De novo identification of repeat families in large genomes [J]. Bioinformatics, 2005, 21(suppl_1): i351-i8.
27. Edgar R C, Myers E W. PILER: identification and classification of genomic repeats [J]. Bioinformatics, 2005, 21(suppl_1): i152-i8.
28. Wicker T, Sabot F, Hua-Van A, et al. A unified classification system for eukaryotic transposable elements [J]. Nature Reviews Genetics, 2007, 8(12): 973–82.
29. Jurka J, Kapitonov V V, Pavlicek A, et al. Repbase Update, a database of eukaryotic repetitive elements [J]. Cytogenetic and genome research, 2005, 110(1–4): 462–7.
30. Tarailo-Graovac M, Chen N. Using RepeatMasker to identify repetitive elements in genomic sequences [J]. Current protocols in bioinformatics, 2009, 25(1): 4.10. 1–4.. 4.
31. Burge C, Karlin S. Prediction of complete gene structures in human genomic DNA [J]. Journal of molecular biology, 1997, 268(1): 78–94.
32. Stanke M, Waack S. Gene prediction with a hidden Markov model and a new intron submodel [J]. Bioinformatics, 2003, 19(suppl_2): ii215-ii25.
33. Majoros W H, Pertea M, Salzberg S L. TigrScan and GlimmerHMM: two open source ab initio eukaryotic gene-finders [J]. Bioinformatics, 2004, 20(16): 2878–9.
34. Blanco E, Parra G, Guig R. Using geneid to identify genes [J]. Current protocols in bioinformatics, 2007, 18(1): 4.3. 1-4.3. 28.
35. Korf I. Gene finding in novel genomes [J]. BMC bioinformatics, 2004, 5(1): 59.
36. Keilwagen J, Wenk M, Erickson J L, et al. Using intron position conservation for homology-based gene prediction [J]. Nucleic acids research, 2016, 44(9): e89-e.
37. Pertea M, Kim D, Pertea G M, et al. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown [J]. Nature protocols, 2016, 11(9): 1650.
38. Sperschneider J, Gardiner D M, Dodds P N, et al. EffectorP: predicting fungal effector proteins from secretomes using machine learning [J]. New Phytologist, 2016, 210(2): 743–61.

Tables

Table 1. Statistics of genome sequencing of Phomopsis vexans isolate PV4

Data type	Sequence Number	Total Bases (bp)	N50Length (bp)	N90Length (bp)	MeanLength (bp)	MaxLength (bp)	MeanQualification
Raw data	651,567	7,557,554,305	16,595	7,096	11,599	123,684	9.07
Clean data	554,287	7,284,861,206	16,713	7,576	13,142	123,684	9.2

Table 2. Statistics of genome assembly of Phomopsis vexans isolate PV4

Contig Length (bp)	Contig Number	Contig N50 (bp)	Contig N90 (bp)	Contig Max (bp)	GC content (%)	Gap length (bp)
59,782,554	26	5,171,527	1,392,057	8,080,950	51.24	0

Table 3. Classification of repetitive elements of Phomopsis vexans isolate PV4

Type	Number	Length (bp)	Percentage (%)
Class I	738	837,592	1.4
Class I/DIRS	1	52	0
Class I/LINE	60	84,477	0.14
Class I/LTR/Copia	258	441,034	0.74
Class I/LTR/Gypsy	97	154,276	0.26
Class I/PLE|LARD	69	84,005	0.14
Class I/SINE	218	52,857	0.09
Class I/Unknown	35	40,214	0.07
Class II	369	61,375	0.1
Class II/Helitron	7	634	0
Class II/MITE	172	47,259	0.08
Class II/TIR	125	9,091	0.02
Class II/Unknown	65	5,515	0.01
PotentialHostGene	455	90,226	0.15
SSR	1,990	692,217	1.16
Total	3,552	2,892,791	4.84

Table 4. Statistics of gene prediction of Phomopsis vexans isolate PV4

Type	Number	Length（bp）	Averange number	Averange length
Gene	15,034	26,919,018	/	1,790.54
Exon	42,753	22,983,590	2.84	537.59
CDS	27,719	3,935,428	1.84	141.98
Intro	42,689	22,604,577	2.84	529.52

Table 5. Statistics of CAZy annotation

CAZy type	Number	Percentage (%)
AA	253	20.97
CBM	102	8.45
CE	197	16.33
GH	483	40.04
GT	128	10.61
PL	43	3.56

Figures

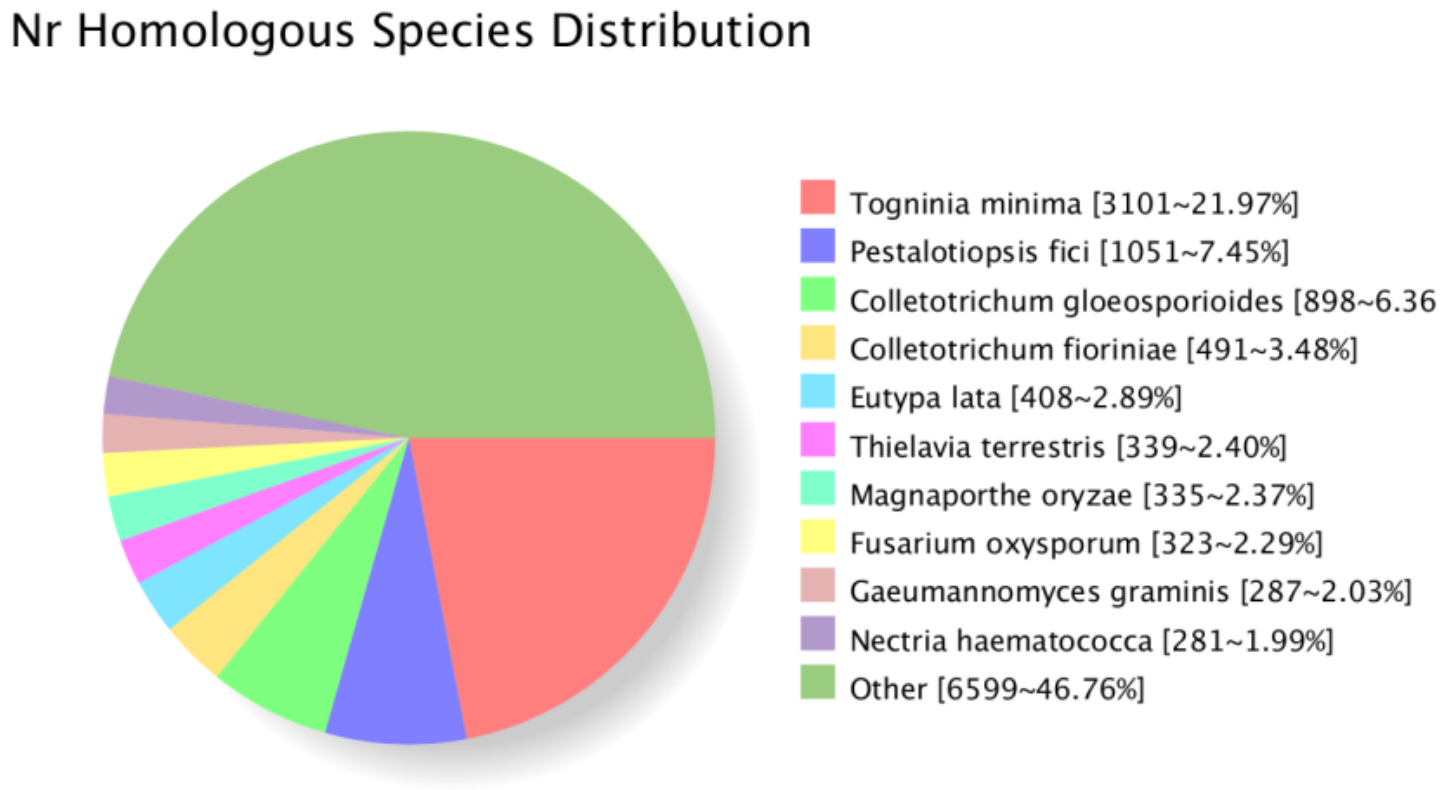

Figure 1

The statistical results of species distribution of sequences annotated from Nr database

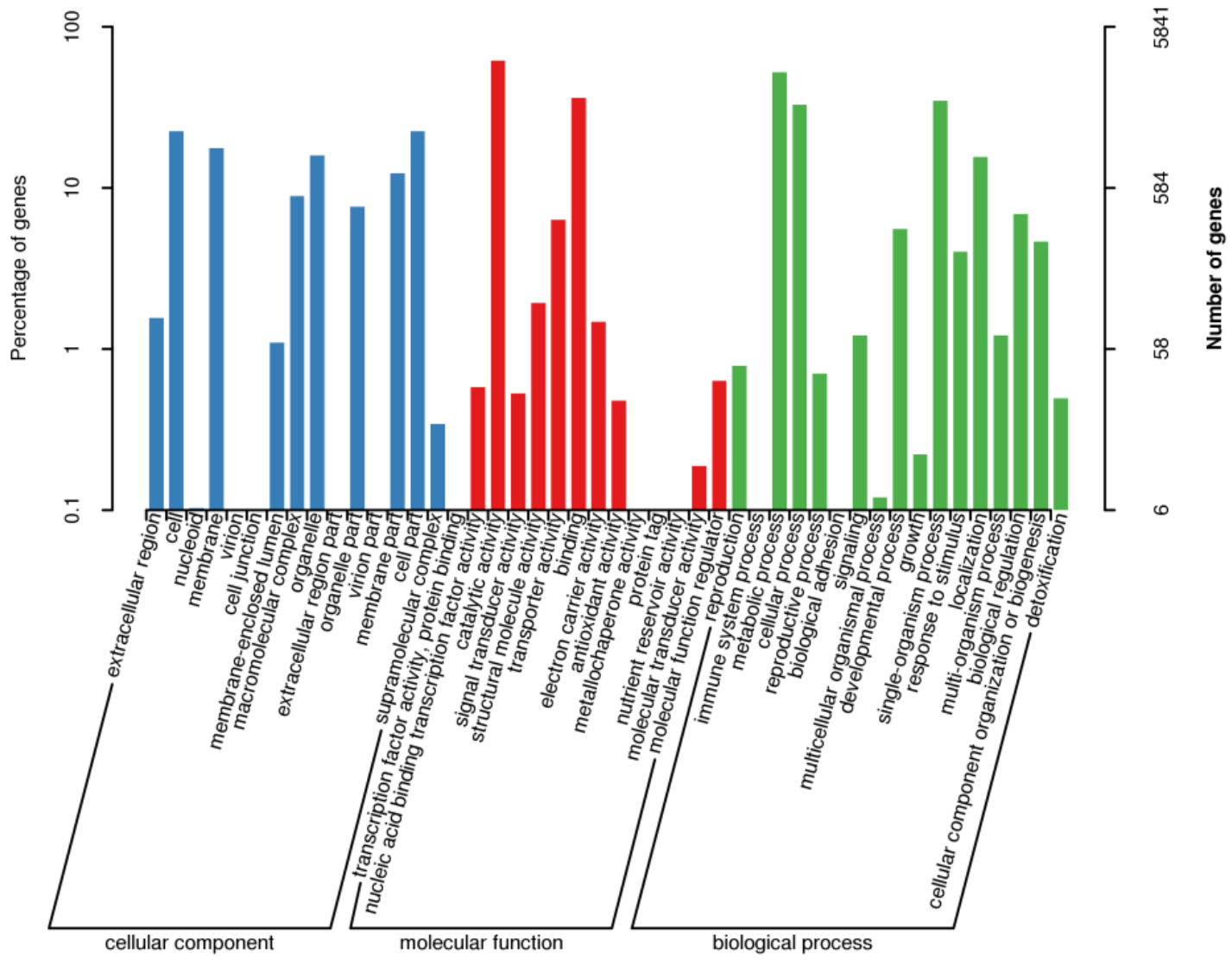

Figure 2

GO annotation classification

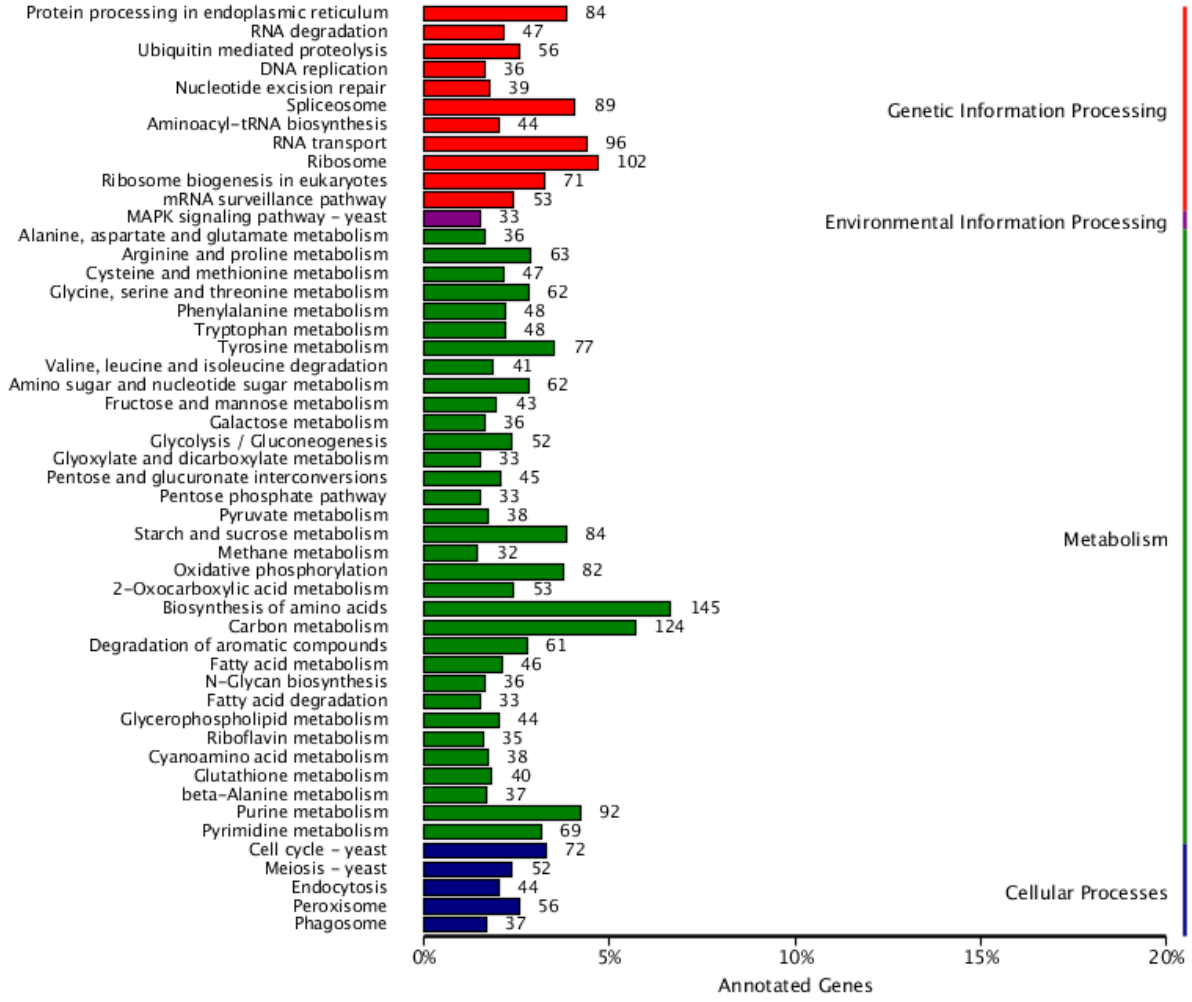

Figure 3

KEGG annotation classification

Supplementary Files

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

- TableS1.xlsx
- TableS2.xls
- TableS3.xls
- TableS4.xlsx