Pan-genome Analyses of 3 Strains of *Inonotus obliquus* and Prediction of Polysaccharide and Terpenoid Genes

Xiaofan Guo\(^1,2\) and Shouming Wang\(^1,2\)

Abstract

*Inonotus obliquus* is a rare, edible and medicinal fungus that is widely used as a remedy for various diseases. Its main bioactive substances are polysaccharides and terpenoids. In this study, we characterized and investigated the pan-genome of three strains of *I. obliquus*. The genome sizes of JL01, HE, and NBRC8681 were 32.04, 29.04, and 31.78 Mb, respectively. There were 6,543 core gene families and 6,197 accessory gene families among the three strains, with 14 polysaccharide-related core gene families and seven accessory gene families. For terpenoids, there were 13 core gene families and 17 accessory gene families. Pan-genome sequencing of *I. obliquus* has improved our understanding of biological characteristics related to the biosynthesis of polysaccharides and terpenoids at the molecular level, which in turn will enable us to increase the production of polysaccharides and terpenoids by this mushroom.

Keywords

*Inonotus obliquus*, medicinal fungus, pan-genome analyses, polysaccharide, terpenoid

Received: September 13th, 2021; Accepted: November 1st, 2021.

Introduction

*Inonotus obliquus* (Persoon : Fries) Pilat also called chaga in Russia, is a type of white-rot fungus in the family Hymenochaetaceae, Basidiomycetes\(^7\) that is edible and has medicinal uses. It is widely distributed across the Northern Hemisphere, mainly in Europe, Asia, and North America.\(^2\) The mushroom is used as an alternative and traditional medicine to manage certain diseases in China, Russia, Japan, and other Asian countries for centuries due to its perceived health benefits.\(^3\)

Polysaccharides and terpenes have been recognized as the main bioactive compounds in *I. obliquus*, and they are responsible for antitumor,\(^4\) anti-inflammatory,\(^8\) antioxidant,\(^6\) and immunomodulatory activities.\(^7\) Some studies have illustrated that crude polysaccharides from *I. obliquus* ameliorate diabetic symptoms and that they show great antioxidant and hypolipidemic capacities.\(^8,9\) A triterpene called inotodiol has antiproliferative properties, demonstrated in *vitro* with human lung adenocarcinoma (A549), cancer-derived, and HeLa cells.\(^10\)

Genome sequencing has become more rapid and more practical, but genome sequences alone are insufficient to explain biological characteristics. Studies of genetic variation between different organisms can overcome such limitations and broaden the understanding of genes with biological functions. The pan-genome is defined as all genes from all isolated strains of a species.\(^11\) Pan-genome analyses are conducted to account for genomic diversity.\(^12\) The pan-genome is usually subdivided into core genome and accessory genome. The core genome includes the conserved genes across all observed genomes from all species. These genes are essential for the viability of the organisms. The accessory genome includes genes from isolated genomes or individual genomes within one species. These genes could influence phenotypic differences between isolates.\(^13\)

Pan-genome analyses are well-established in comparative prokaryote genomics, but have also been extended to comparative intraspecific studies in eukaryotes. There have been a series of recent studies showing the existence of eukaryotic pan-genomes in some form. Comparative analyses of different cultivars of *Brassica oleracea* have been conducted using the pan-
In fungi, multiple studies of the *Saccharomyces cerevisiae* pan-genome have shown evidence for core and accessory genomes of varying size based on whole-genome sequencing of *S. cerevisiae* isolates, providing an evolutionary picture of the genomic variants in yeast.

There are many strains of *I. obliquus* that are widely distributed in the northern hemisphere. Analysis of their evolutionary relationships has great significance for making full use of the wild resources of this species. A genome-wide analysis of triterpene compounds and polysaccharide genes can reveal their formation mechanisms at the molecular level, in order to improve triterpene and polysaccharide contents in *I. obliquus*.

### Materials and Methods

#### Sample Collection and DNA Extraction

Three strains of *I. obliquus* from different countries were analyzed. JL01 originated from China, HE from Russia, and NBRC8681 from Japan. Hyphae with vigorous growth were cultured for 20 days to extract genomic DNA, which was extracted from mycelium using the cetyltrimethylammonium bromide (CTAB) procedure. The quality and concentration of the DNA samples were quantified using NanoDrop 2000 (Thermo Fisher).

#### DNA Library Preparation and Sequencing

DNA was extracted and randomly interrupted using physical means (ultrasound) to obtain inserted fragments of the required length (i.e., 350 bp). Primers were added at the sticky end of the fragment to construct the library. The constructed sequencing library was used for bridge PCR on a sequencing chip (Illumina flow cell), followed by paired-end sequencing. The resulting data were interpreted as nucleic acid sequences for further analysis.

#### Illumina Sequencing Quality Control

The raw sequence reads were scored (quality score = $-10 \times \log_{10} P$, $P$ was the probability of identifying base errors). Higher quality scores indicate more reliable bases. A quality score of 20 indicates that one in 100 bases will be mistakenly identified, while a score of 30 indicates that one in 1 000 bases will be mistakenly identified. Clean reads were obtained by removing the sequencing adapters and primer sequences of raw reads and filtering of low-quality data.

#### Genomic Component Analysis

Velvet software was used to assemble the sequencing data of each strain. Genemark-ES was used to predict the genes of the assembled genome. Repeat Masker software was used to predict the repeated sequences of the strain genome. The software tRNA scan-SE was used to predict the tRNA in the genome, and the software Infernal 1.1 was used to predict the rRNA in the genome and other ncRNAs, in addition to tRNA and rRNA based on the Rfam database.

#### Gene Function Annotation

The predicted gene sequences were annotated with functional databases such as clusters of orthologous groups (COG) (Kyoto encyclopedia of genes and genome (KEGG)), Swiss-prot, TrEMBL, and the non-redundant (Nr) database by BLAST, and the annotation results of gene functions were obtained. Based on the comparison results of the nr database, Blast2GO software was used to annotate the function in the gene ontology (GO) database. HMMER was used to conduct Pfam function annotation based on the Pfam database. In addition, further COG classification analysis, KEGG metabolic pathway analysis, and GO function classification analysis were conducted according to the annotation results.

#### Analysis of Variation and Differences Among Strains

JL01 was used as the reference genome, and genomes of the other two strains were compared with JL01 using the Applied MUMmer package. These comparisons identified single nucleotide polymorphisms (SNPs), small insertions/deletions (InDels), and genome structure variation. Alien Hunter and HGT-Finder were then used to predict horizontal gene transfer for each strain.

#### Comparative Genomics Analysis

OrthoMCL was used for gene family analysis of predicted protein sequences of each strain and protein sequences of the reference genome to look for gene families that are either common (core gene families) or unique to each strain (accessory gene families). JL01 was used as the reference genome. The software Mugsy was used to compare the genome sequences of all sequenced strains with those of the reference genome. From the comparison results, the common sequences of all strains were identified as the core genome sequences.

| Strain   | Data (Gb) | Depth (X) | Q20 (%) | Q30 (%) |
|----------|-----------|-----------|---------|---------|
| JL01     | 8.64      | 270       | 97.52   | 93.14   |
| HE       | 6.39      | 220       | 93.08   | 97.48   |
| NBRC8681 | 6.42      | 202       | 97.50   | 93.07   |
Results

Genome Sequencing and Assembly

A 350 bp library was constructed for each strain, and the raw data were sequenced using the Illumina platform. The results are shown in Table 1. The sequencing volume of the corresponding sequencing libraries was between 6.39 and 8.64 Gb. The minimum sequencing depth was 202 and the maximum 270. The sequencing quality values Q20 and Q30 were all above 93%.

GC content and genome size for all three strains are summarized in Table 2. Results show that the genome size of JL01 is 32.04 Mb, that of HE is 29.04 Mb, and that of NBRC8681 is 31.78 Mb. GC content for all three strains was >47%, indicating moderate GC content.

Genomic Component Analysis

Gene prediction was carried out for the assembled genomes. The specific statistical information of gene prediction for each strain is shown in Table 3. Strain NBRC8681 had the largest number of genomes (10 491), followed by JL01 (10 399), and HE (9 494). Mean gene length (bp) of all three strains is 1 400-1 500. The percentages of repetitive sequence ranged from 0.84%-1.14% (Table 4). The numbers of rRNA genes were four (JL01), five (HE), and two (NBRC8681), and the numbers of tRNA genes were 83 (JL01), 86 (HE), and 84 (NBRC8681). The numbers of other RNA genes were 21 (JL01), 19 (HE), and 20 (NBRC8681) (Table 5).

Gene Function Annotation

The results of functional annotation of the predicted genes are shown in Table 6. The databases used for functional annotations include COG, GO, KEGG, Pfam, Swiss-Prot, TrEMBL, and Nr. The numbers of predicted genes annotated were 8 908 (JL01), 8 744 (HE), and 8 932 (NBRC8681). The protein sequences of the predicted genes were compared with special databases including the carbohydrate-active enzymes database (CAZyme) and pathogen host interactions (PHI-base) (Table 7). In total, 424 (JL01), 428 (HE), and 419 (NBRC8681) genes were annotated to the CAZyme private database. For the private PHI-base, 2 394 (JL01), 2 429 (HE), and 2 377 (NBRC8681) genes were annotated.

Variation and Differences Among Strains

Each strain was compared with the genome of the reference strain JL01 to find SNPs and small InDels (Table 8). The number of SNP mutations in HE reached 36 973; however, NBRC8681 had 144 348 SNPs. In terms of structural variation (InDels), HE did not differ from JL01, regardless of insertion or deletion. However, there were 21 insertions and 70 deletions in strain NBRC8681. The horizontally transferred genes of each strain were predicted; there were no shared genes between JL01, HE, and NBRC8681.
Comparative Genomics Analysis

The predicted protein sequences of each strain and the protein sequences of the reference genome were analyzed to find the common and unique gene families of each strain. Gene families shared by all strains are core gene families, and any others are considered accessory. The statistical results of core gene and accessory gene families are shown in Figure 1. There were 6,543 core and 6,197 accessory gene families among the three strains.

Note: the area in the middle circle represents the number of gene families shared by all strains. The number in the ellipses around the middle circle represent the number of gene families unique to each strain.

Pan-genome families vary with the number of strains included (Figure 2). JL01 is the reference genome. The number of core gene families decreased as strains were added: there were 6,660 core genes when HE was added and 6,543 core genes when NBRC8681 was added. However, on the contrary, the number of pan-genome gene families increased gradually as strains were added. The number of pan-genome gene families reached 11,383 when HE was added, and 12,740 when NBRC8681 was added.

To characterize the distribution of gene functions on the basis of the GO classification, GO analysis was performed to identify the functions of the strain-specific genes of JL01, HE, and NBRC8681, which were classified into three main categories: cellular component, molecular function, and biological process. The results are shown in Figure 3.

For JL01, there were no strain-specific genes for supramolecular complex, transcription factor activity, protein binding, nucleic acid binding transcription factor activity, signal transducer activity, electron carrier activity, metallochaperone activity, translation regulator activity, molecular transducer activity, molecular function regulator, reproduction, immune system process, reproductive process, signaling, multicellular organismal process, developmental process, multi-organism process, and biological regulation. The largest number of annotated genes (1,061) was enriched in catalytic activity, followed by metabolic process (894) and binding (775).

For HE, there were no strain-specific genes for supramolecular complex, transcription factor activity, protein binding, signal transducer activity, metallochaperone activity, translation regulator activity, molecular transducer activity, molecular function regulator, reproduction, immune system process, reproductive process, signaling, multicellular organismal process, developmental process, multi-organism process, and biological regulation. The largest number of annotated genes (1,061) was enriched in catalytic activity, followed by metabolic process (894) and binding (775).

For NBRC8681, there were no strain-specific genes for extra-cellular region, membrane-enclosed lumen, supramolecular complex, transcription factor activity, protein binding, nucleic acid binding transcription factor activity, electron carrier activity, antioxidant activity, protein tag, translation regulator activity,
molecular function regulator, reproduction, immune system process, reproductive process, multicellular organismal process, developmental process, multi-organism process, and detoxification. All the annotated genes were concentrated in catalytic activity (1,037), metabolic process (868), and binding (789).

**Prediction of Polysaccharide and Terpenoid Genes**

To identify candidate genes responsible for the polysaccharide and terpenoid genes, functional genome comparisons were conducted by analyzing the pan-genome. There were 20 gene families related to polysaccharide biological function, including 50 genes. The number of core gene families was 14 and there were seven accessory gene families among the three strains. There were 30 gene families related to terpenoid biological function, including 71 genes. The number of core gene families was 13 and there were 17 accessory gene families among the three strains (Figure 4, Tables 9 and 10). In the polysaccharide biological synthesis pathway, polysaccharide lyase family protein, endocellulase, cellulase, glucoamylase, lytic polysaccharide monooxygenase, cellobiose dehydrogenase, glycoside hydrolase family protein, and NAD-binding protein were identified (Table 9). In the terpenoid synthesis pathway, terpenoid synthase, terpenoid cyclases, delta(6)-protoilludene synthase, and alpha-muurolene synthase genes were predicted (Table 10).

**Discussion**

*Inonotus obliquus* has a wide geographical distribution, being found from the mountainous areas of meridional zones to subarctic areas in the Northern Hemisphere, including regions within Canada, the United States, Asia, and Europe. In this study, JL01 originated from China, HE from Russia, and NBRC8681 from Japan. The strains used in this pan-genome analysis span the typical geographical region of *I. obliquus*, and can, therefore, provide accurate and broad genomic information.

Genome sequencing provides the opportunity to better understand and confirm the genetic relationships among fungal isolates, and can be used to identify gene candidates.
Figure 3. Annotation classification of GO secondary nodes for specific genes.
that could contribute to biological activity. For *I. obliquus* strains, our results showed that the genome size of JL01 is 32.04 Mb, for HE 29.04 Mb, and for NBRC8681 31.78 Mb, all of which are consistent with the genome sizes of edible and medicinal fungi sequenced previously, which are between 30 and 50 Mb. The genome size of *Ganoderma lucidum* is 43.3 Mb, encoding 16 113 predicted genes.42 The genome of *Volvariella volvacea* has been sequenced and assembled into 62 scaffolds with a total genome size of 35.7 Mb, containing 11 084 predicted gene models.43 The complete *Pleurotus ostreatus* genome contains 35 Mb.44 Although JL01 had the largest genome, it has fewer predicted genes than strain NBRC8681, which had 10 491.

Comparing the genome sequences of the three strains, the number of SNPs/small InDels was found to be significantly different (Table 8). SNPs/small InDels of NBRC8681 had the biggest differences compared to JL01. Structural variation of NBRC8681, including InDels, was obviously different from JL01. Although HE and JL01 have some differences in SNPs/small InDels, there is no difference in InDel structural variation between the two strains. This may be due to the relatively close geographical locations of JL01 and HE. However, NBRC8681 has the most predicted genes (10 491), which may be related to it having the most variation.

Sequencing has been widely used to predict gene function. Li et al. used *Ganoderma lucidum* at three different developmental stages as materials and conducted transcriptome analysis on the long non-coding RNA (lincRNA), identifying 402 lincRNA with an average length of 609 bp.45 Muraguchi et al. identified a large number of genes related to secondary metabolism by sequencing.46 In our study, functional genome comparisons were conducted by analyzing the pan-genome. In total, 20 gene families related to polysaccharide biological function were identified. The number of core gene families related to polysaccharide biological function was 14; however, there were only seven gene families for polysaccharide biological function belonging to accessory gene families among the three strains. Researchers should therefore focus on core gene families when studying polysaccharide biosynthesis genes. On the contrary, there were 13 core gene families related to terpenoid biological function and 17 accessory gene families related to terpenoid biological function (Figure 4). When studying the functional genes of terpenes of *I. obliquus*, it may thus be more important to focus on accessory gene families than on core gene families.

In this study, we predicted some polysaccharide related genes (Table 9). Polysaccharide lyases cleave the polymer chain by a β-elimination mechanism. Most of the enzymes in this class randomly cleave the main chain of polysaccharide structures. The products in a few cases may be monosaccharides, but are more commonly oligosaccharides.47 Cellulases break down cellulose into monosaccharides, polysaccharides, and oligosaccharides.48 Endocellulase (EGPl) is one kind of hyperthermophilic cellulase, and EGPf combined with β-glucosidase (BGLPf) degrades glucose at high temperature.49 Glucoamylase hydrolyzes polysaccharides by consecutive cleavage of α-1,4 and α-1,6 glycosidic bonds.50 Regulation of glycoside hydrolase is important during polysaccharide biosynthesis.51 Lytic polysaccharide monooxygenases (LPMO10 s) use redox chemistry to cleave glycosidic bonds in polysaccharides, such as cellulose and chitin.52

![Figure 4. Polysaccharide (left) and terpenoid (right) gene family Venn diagrams with three circles.](image-url)
| Genes | Nr annotation | Species | Genes | Nr annotation | Species |
|-------|---------------|---------|-------|---------------|---------|
| GE02963_g(HE) | Polysaccharide lyase family 14 protein  | *Gymnopus lucarium* | FD-317M1 | GE07739_g(NBRC8681) | Cellobiose dehydrogenase  | mediterranea MF3/22 |
| GE05277_g(HE) | Polysaccharide lyase family 3 protein  | *Gymnopus lucarium* | FD-317M1 | GE09325_g(HE) | Cellobiose dehydrogenase  | mediterranea MF3/22 |
| GE08278_g(HE) | Polysaccharide lyase family 1 protein  | *Tulasnella calospora* | MUT 4182 | GE06865_g(JL01) | Cellobiose dehydrogenase  | mediterranea MF3/22 |
| GE04038_g(HE) | Polysaccharide lyase family 1 protein  | *Plicaturopsis crispa* | FD-325 SS-3 | GE08930_g(HE) | Glycoside hydrolase family 15 protein  | *Serpula lacrymans* var. *lacrymans* S7.9 |
| GE04401_g(NBRC8681) | Polysaccharide lyase family 1 protein  | *Plicaturopsis crispa* | FD-325 SS-3 | GE06777_g(NBRC8681) | Glycoside hydrolase family 15 protein  | *Serpula lacrymans* var. *lacrymans* S7.9 |
| GE01958_g(JL01) | Polysaccharide lyase family 1 protein  | *Plicaturopsis crispa* | FD-325 SS-3 | GE04981_g(JL01) | Glycoside hydrolase family 15 protein  | *Serpula lacrymans* var. *lacrymans* S7.9 |
| GE04074_g(HE) | Endocellulase | *Fomitiporia mediterranea* | MF3/22 | GE09186_g(HE) | Hypothetical protein  | *Fomitiporia mediterranea* MF3/22 |
| GE00984_g(NBRC8681) | Endocellulase | *Fomitiporia mediterranea* | MF3/22 | GE05573_g(NBRC8681) | Hypothetical protein  | *Fomitiporia mediterranea* MF3/22 |
| GE00018_g(JL01) | Endocellulase | *Fomitiporia mediterranea* | MF3/22 | GE07531_g(JL01) | Hypothetical protein  | *Fomitiporia mediterranea* MF3/22 |
| GE01852_g(JL01) | Cellulase | *Fomitiporia mediterranea* | MF3/22 | GE00786_g(HE) | Hypothetical protein  | *Fomitiporia mediterranea* MF3/22 |
| GE02649_g(JL01) | Cellulase | *Fomitiporia mediterranea* | MF3/22 | GE09192_g(NBRC8681) | Hypothetical protein  | *Fomitiporia mediterranea* MF3/22 |
| GE01325_g(NBRC8681) | Cellulase | *Fomitiporia mediterranea* | MF3/22 | GE08442_g(JL01) | Hypothetical protein  | *Fomitiporia mediterranea* MF3/22 |
| GE04681_g(NBRC8681) | Cellulase | *Fomitiporia mediterranea* | MF3/22 | GE05292_g(JL01) | NAD-binding protein  | *Fomitiporia mediterranea* MF3/22 |
| GE05314_g(HE) | Cellulase | *Fomitiporia mediterranea* | MF3/22 | GE08414_g(NBRC8681) | NAD-binding protein  | *Fomitiporia mediterranea* MF3/22 |
| GE05615_g(HE) | Glucoamylase | *Fomitiporia mediterranea* | MF3/22 | GE01195_g(HE) | NAD-binding protein  | *Fomitiporia mediterranea* MF3/22 |
| GE03127_g(NBRC8681) | Glucoamylase | *Fomitiporia mediterranea* | MF3/22 | GE06244_g(JL01) | Polysaccharide lyase family 14 protein  | *Fomitiporia mediterranea* MF3/22 |
| GE04675_g(JL01) | Glucoamylase | *Fomitiporia mediterranea* | MF3/22 | GE04179_g(NBRC8681) | Polysaccharide lyase family 14 protein  | *Fomitiporia mediterranea* MF3/22 |
| GE07012_g(HE) | Polysaccharide lyase family 8 protein  | *Stereum hirsutum* | FP-91666 SS1 | GE01990_g(HE) | Polysaccharide lyase family 14 protein  | *Fomitiporia mediterranea* MF3/22 |
| GE03717_g(NBRC8681) | Polysaccharide lyase family 8 protein  | *Stereum hirsutum* | FP-91666 SS1 | GE06715_g(JL01) | Polysaccharide lyase family 14 protein  | *Fomitiporia mediterranea* MF3/22 |
| GE03964_g(JL01) | Polysaccharide lyase family 8 protein  | *Stereum hirsutum* | FP-91666 SS1 | GE04559_g(NBRC8681) | Polysaccharide lyase family 1 protein  | *Tulasnella calospora* MUT 4182 |
| GE07304_g(HE) | Glycoside hydrolase family 61 protein | *Trametes versicolor* | FP-101664 SS1 | (Continued) | (Continued) | (Continued) |
NAD-binding domain is found in diverse bacterial polysaccharide biosynthesis proteins. It is an import protein for polysaccharide biosynthesis.53 Some terpenoid related genes were also predicted (Table 10). Most have been annotated to terpenoid synthase. Terpene synthase (TPS) enzymes are recognized as the gatekeepers of species-specific terpenoid pathways. TPS gene families include more than 100 members that function in metabolic networks.54 Because terpenoid synthases catalyze cyclization reactions, they are also known as terpenoid cyclases.55 Terpene cyclase catalyzes the cyclization of farnesyl diphosphate (FPP) to delta(6)-protoilludene. In the presence of Ca2+, delta(6)-protoilludene synthase catalyzes diverse cyclization reactions.56 Alpha-muurolene synthase is encoded by the COP3 gene. Cop3 synthesizes seven different sesquiterpenes with alpha muurolene 15 and germacrene A 13 as the major products.57 Analysis of these identified genes has great significance to elucidate the synthesis mechanism of polysaccharides and terpenoids of *I. obliquus*.

### Author Contributions
Xiaofan Guo and Shouming Wang designed experiments, collected and analyzed data. Shouming Wang wrote the manuscript.

### Data Availability Statement
The genome assembly of Inonotus obliquus in this study has been deposited at NCBI with accession no. PRJNA743015.

**Table 9. Continued**

| Genes                  | Nr annotation     | Species                  |
|------------------------|-------------------|--------------------------|
| GE01814_g(HE)          | Polysaccharide lyase family 1 protein | *Tulasnella calospora* MUT 4182 |
| GE02264_g(JL01)        | Lytic polysaccharide monoxygenase | *Hydnomorpha pinastri* MD-312 |
| GE09761_g(NBRC8681)    | Lytic polysaccharide monoxygenase | *Tulasnella calospora* MUT 4182 |
| GE03045_g(JL01)        | Polysaccharide lyase family 1 protein | *Tulasnella calospora* MUT 4182 |
| GE01081_g(HE)          | Polysaccharide lyase family 1 protein | *Fomitiporia mediterranea* MF3/22 |
| GE01333_g(JL01)        | Polysaccharide lyase family 14 protein | *Fomitiporia mediterranea* MF3/22 |
| GE02583_g(HE)          | Cellulase CEL6B   | *Fomitiporia mediterranea* MF3/22 |
| GE00446_g(NBRC8681)    | Cellulase CEL6B   | *Fomitiporia mediterranea* MF3/22 |
| GE09864_g(JL01)        | Cellulase CEL6B   | *Fomitiporia mediterranea* MF3/22 |

**Table 10. Prediction of Terpenoid Genes Family.**

| Genes                  | Nr annotation     | Species                  |
|------------------------|-------------------|--------------------------|
| GE03794_g(HE)          | Terpenoid synthase | *Fomitiporia mediterranea* |
| GE05504_g(NBRC8681)    | Terpenoid synthase | *Fomitiporia mediterranea* |
| GE08223_g(JL01)        | Terpenoid synthase | *Fomitiporia mediterranea* |
| GE04171_g(HE)          | Terpenoid cyclases | *Fomitiporia mediterranea* |
| GE05339_g(NBRC8681)    | Terpenoid cyclases | *Fomitiporia mediterranea* |
| GE01777_g(JL01)        | Terpenoid cyclases | *Fomitiporia mediterranea* |
| GE04248_g(HE)          | delta(6)-protoilludene synthase | *Inonotus obliquus* |
| GE03437_g(NBRC8681)    | delta(6)-protoilludene synthase | *Inonotus obliquus* |
| GE01863_g(JL01)        | delta(6)-protoilludene synthase | *Inonotus obliquus* |
| GE04995_g(HE)          | Terpenoid synthase | *Fomitiporia mediterranea* |
| GE08561_g(NBRC8681)    | Terpenoid synthase | *Fomitiporia mediterranea* |
| GE07732_g(JL01)        | Terpenoid synthase | *Fomitiporia mediterranea* |
| GE02607_g(HE)          | alpha-muurolene synthase | *Inonotus obliquus* |
| GE00672_g(NBRC8681)    | alpha-muurolene synthase | *Inonotus obliquus* |
| GE00673_g(NBRC8681)    | alpha-muurolene synthase | *Inonotus obliquus* |
| GE05391_g(JL01)        | alpha-muurolene synthase | *Inonotus obliquus* |
| GE05392_g(JL01)        | alpha-muurolene synthase | *Inonotus obliquus* |
| GE06333_g(HE)          | Terpenoid cyclases | *Fomitiporia mediterranea* |
| GE05333_g(NBRC8681)    | Terpenoid cyclases | *Fomitiporia mediterranea* |
| GE05832_g(JL01)        | Terpenoid cyclases | *Fomitiporia mediterranea* |
| GE06571_g(HE)          | alpha-muurolene synthase | *Inonotus obliquus* |
| GE05054_g(NBRC8681)    | alpha-muurolene synthase | *Inonotus obliquus* |
| GE07751_g(JL01)        | alpha-muurolene synthase | *Inonotus obliquus* |

(Continued)
Table 10. Continued

| Genes            | Nr annotation | Species                  |
|------------------|---------------|--------------------------|
| GE00313_g(HE)    | Terpenoid synthase | Fomitiporia mediterranea |
| GE00314_g(HE)    | Terpenoid synthase | Fomitiporia mediterranea |
| GE02426_g(NBRC8681) | Terpenoid synthase | Fomitiporia mediterranea |
| GE02427_g(NBRC8681) | Terpenoid synthase | Fomitiporia mediterranea |
| GE02678_g(JL01)  | Terpenoid synthase | Fomitiporia mediterranea |
| GE08174_g(HE)    | Terpenoid synthase | Sanghuangporusbaumii     |
| GE04419_g(NBRC8681) | Terpenoid synthase | Sanghuangporusbaumii     |
| GE06407_g(JL01)  | Terpenoid synthase | Sanghuangporusbaumii     |
| GE08442_g(HE)    | Terpenoid cyclases | Fomitiporia mediterranea |
| GE01971_g(NBRC8681) | Terpenoid cyclases | Fomitiporia mediterranea |
| GE01672_g(JL01)  | Terpenoid cyclases | Fomitiporia mediterranea |
| GE00906_g(HE)    | Terpenoid synthase | Fomitiporia mediterranea |
| GE00907_g(HE)    | Terpenoid synthase | Fomitiporia mediterranea |
| GE09461_g(NBRC8681) | Terpenoid synthase | Fomitiporia mediterranea |
| GE06089_g(JL01)  | Terpenoid synthase | Fomitiporia mediterranea |
| GE00746_g(HE)    | Terpenoid synthase | Fomitiporia mediterranea |
| GE09334_g(NBRC8681) | Terpenoid synthase | Fomitiporia mediterranea |
| GE02815_g(JL01)  | Terpenoid synthase | Fomitiporia mediterranea |
| GE01592_g(JL01)  | Terpenoid synthase | Fomitiporia mediterranea |
| GE02617_g(JL01)  | Terpenoid synthase | Fomitiporia mediterranea |

(Continued)
Table 10. Continued

| Genes                  | Nr annotation | Species                  |
|------------------------|---------------|--------------------------|
| GE07733_g(L01)         | Terpenoid synthase | *Fomitiporia* mediterranea |
| GE09597_g(L01)         | Terpenoid synthase | *Fomitiporia* mediterranea |
| GE00208_g(HE)          | Terpenoid synthase | *Fomitiporia* mediterranea |
| GE00747_g(HE)          | Terpenoid synthase | *Fomitiporia* mediterranea |
| GE08508_g(HE)          | Terpenoid synthase | *Fomitiporia* mediterranea |
| GE05710_g(HE)          | Terpenoid synthase | *Fomitiporia* mediterranea |
| GE09460_g(NBRC8681)    | Terpenoid synthase | *Fomitiporia* mediterranea |

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Natural Science Foundation of Xiaogan city (grant number [XGKJ2020010055]).

ORCID iD

Shouming Wang  https://orcid.org/0000-0001-7122-0375

References

1. Taji S, Yamada T, Wada SI, Tokuda H, Sakuma K, Tanaka R. Lanostane-type triterpenoids from the sclerotia of *Inonotus obliquus* possessing anti-tumor promoting activity. *Eur J Med Chem*. 2008;43(11):2373-2379.
2. Sun JF, Ao ZH, Lu ZM, et al. Anthyglycemic and antilipid per-oxidative effects of dry matter of culture broth of *Inonotus obliquus* in submerged culture on normal and alloxan diabetes mice. *J Ethnopharmacol*. 2008;118(1):7-13.
3. Zhang N, Chen HX, Ma LS, Zhang Y. Physical modifications of polysaccharide from *Inonotus obliquus* and the antioxidant properties. *Int J Biol Macromol*. 2013;54:209-215.
4. Zhao F, Mai Q, Ma J, et al. Triterpenoids from *Inonotus obliquus* and their antitumor activities. *Fitoterapia*. 2015;101:34-40.
5. Ma L, Chen H, Dong P, Lu X. Anti-inflammatory and anticancer activities of extracts and compounds from the mushroom *Inonotus obliquus*. *Food Chem*. 2013;139:503-508.
6. Glamoclija J, Ciric A, Nikolic M, et al. Chemical characterization and biological activity of chaga (*Inonotus obliquus*), a medicinal "mushroom". *J Ethnopharmacol*. 2015;162:323-332.
7. Fan L, Ding S, Ai L, Deng K. Antitumor and immunomodulatory activity of water-soluble polysaccharide from *Inonotus obliquus*. *Carbohydr Polym*. 2012;90:870-874.
8. Diao BZ, Jin WR, Yu XJ. Protective effect of polysaccharides from *Inonotus obliquus* on streptozotocin-induced diabetic symptoms and their potential mechanisms in rats. *Evid-based Complement. Altern. Med*. 2014. doi:10.1155/2014/841496
9. Liang LY, Zhang ZS, Sun W, Wang YB. Effect of the *Inonotus obliquus* polysaccharides on blood lipid metabolism and oxidative stress of rats fed high-fat diet in vivo. *ICBEI*. 2009;1-4. doi: 10.1109/ICBEI.2009.530591
10. Zhong XH, Wang LB, Sun DZ. Effects of inotodiol extracts from *Inonotus obliquus* on proliferation cycle and apoptotic gene of human lung adenocarcinoma cell line A549. *Chin J Integr Med*. 2011;17:218-223.
11. Charley GP, McCarthy DAF. Pan-genome analyses of model fungal species. Microb. Genom. 2019;5(2):e000243.
12. Tettelin H, Masignani V, Cieslewicz MJ, et al. Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae* implications for the microbial "pan-genome". *Proc. Natl. Acad. Sci.U.S.A*. 2005;102:13950-13955.
13. Rouli L, Merhej V, Fournier PE, Raoult D. The bacterial pan-genome as a new tool for analysing pathogenic bacteria. *New Microbes New Infect*. 2015;7:72-85.
14. Golicz AA, Bayer PE, Barker GC, et al. The pangenome of an agronomically important crop plant *Brassica oleracea*. *Microbes New Infect*. 2015;7:13390.
15. Dunn B, Richter C, Kvitk P, Pugh T, Sherlock G. Analysis of the *Saccharomyces cerevisiae* pan-genome reveals a pool of copy number variants distributed in diverse yeast strains from differing industrial environments. *Genome Res*. 2012;22:908-924.
16. Ewing B, Hillier L, Wendel MC, Green P. Base-calling of automated sequencer traces using phred. *I. Accuracy assessment. Genome Res*. 1998;8:175-185.
17. Zerbino DR, Birney E. Velvet: algorithms for de novo short read assembly using deBrujin graphs. *Genome Res*. 2008;18(5):821-829.
18. Borodovsky M, Lomsadze A. Eukaryotic gene prediction using GeneMark. *BMC Bioinformatics*. 2011. doi:10.1186/1471-2105-12-53. Chapter: Unit 17.3.
19. Tarailo-Graovac M, Chen N. Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr Protoc Bioinformatics*. 2009. doi:10.1002/0471250953.bi0410s25. Chapter 4: Unit 4.10.
20. Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res*. 1997;25(5):955-964.
23. Nawrocki EP, Eddy SR. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics*. 2013;29(22):2933-2935.

24. Nawrocki EP, Burge SW, Bateman A, et al. Rfam 12.0: updates to the RNA families database. *Nucleic Acids Res*. 2015;43:D130-7. doi: 10.1093/nar/gku1063

25. Tatusov RL, Galperin MY, Natale DA, et al. The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res*. 2000;28(1):33-36.

26. Kanehisa M, Goto S, Kawashima S, et al. The KEGG resource for genomes. *Nucleic Acids Res*. 2017;45:D354-D365.

27. Boeckmann B, Bairoch A, Apweiler R, et al. The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Res*. 2003;31(1):365-370.

28. Deng YY, Li JQ, Wu SF, et al. Integrated nr database in protein annotation system and its localization. *Comput Eng*. 2006;32(5):71-74.

29. Altschul SF, Madden TL, Schäffer AA, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*. 1997;25(17):3389-3402.

30. Conesa A, Götz S, García-Gómez JM, et al. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*. 2005;21(18):3674-3676.

31. Ashburner M, Ball CA, Blake JA, et al. Gene ontology: tool for the unification of biology. *Nature genetics*. 2000;25(1):25-29.

32. Eddy SR. Profile hidden Markov models. *Bioinformatics*. 1998;14(9): 755-763.

33. Finn RD, Coggill P, Eberhardt RY, et al. The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res*. 2016;44(Database issue):D279-D285.

34. Winnenburg R, Baldwin TK, Urban M, et al. PHI-base: a new database for pathogen host interactions. *Nucleic Acids Res*. 2002;30(11):2478-2483.

35. Delcher AL, Phillippy A, Carlton J, et al. Fast algorithms for large-scale genome alignment and comparison. *Nucleic Acids Res*. 2002;30(11):2478-2483.

36. Cantarel BL, Coutinho PM, Rancurel C, et al. The carbohydrate-active enzymes database (CAZy): an expert resource for glycogenomics. *Nucleic Acids Res*. 2009;37:D233-D238.

37. Vernikos GS, Parkhill J. Interpolated variable order motifs for identification of horizontally acquired DNA: revisiting the *Salmonella* pathogenicity islands. *Bioinformatics*. 2006;22(18):2196-2203.

38. Nguyen M, Ekstrom A, Li X, et al. HGT-Finder: A new tool for horizontal gene transfer finding and application to Aspergillus genomes. *Taxon*. 2015;7(10):4035-4053.

39. Li L, Stoeckert CJ, Roos DS. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome research*. 2003;13(9):2178-2189.

40. Angiuoli SV, Salzberg SL. Mugsy: fast multiple alignment of closely related whole genomes. *Bioinformatics*. 2011;27(3):334-342.

41. Lee MW, Hur H, Chang KC, Lee TS, Ka KH, Jankovsky L. Introduction to distribution and ecology of sterile conks of *Inonotus obliquus*. *Mycolology*. 2008;36(4):199-202.

42. Chen SL, Xu J, Liu C, et al. Genome sequence of the model medicinal mushroom *Ganoderma lucidum*. *Nature Communications*. 2012;3:913. https://doi.org/10.1038/ncomms1923

43. Bao D, Gong M, Zheng H, et al. Sequencing and comparative analysis of the straw mushroom (*Volvariella volvacea*) genome. *Plas One*. 2013;8(3):e58294. https://doi.org/10.1371/journal.pone.0058294

44. Ramírez L, Oguiza JA, Pérez G, et al. Genomics and transcriptomics characterization of genes expressed during postharvest at 4°C by the edible basidiomycete *Pleurotus ostreatus*. *Intl Microbiol*. 2011;14:111-120.

45. Li J, Wu B, Xu J, Liu C. Genome-wide identification and characterization of long intergenic non-coding RNAs in *Ganoderma lucidum*. *PLoS ONE*. 2014;9(6):e99442. doi: 10.1371/journal.pone.0099442

46. Muraguchi H, Umezawa K, Niikura M, et al. Strand-specific RNA-Seq analyses of fruiting body development in *Carpophilus cinerea*. *PLoS ONE*. 2015;10(10):e0141586. doi:10.1371/journal.pone.0141586.

47. Sutherland IW. Polysaccharide lyases. *FEBS Microbiology Reviews*. 1995;16:323-347.

48. Barkalow DG, Whistler RL. *Cellulose*. *Access-Science McGraw-Hill Education*. 2019. doi:10.1036/1097-8542.118200.

49. Kataoka M, Ishikawa K. Complete saccharification of β-glucan using hyperthermophilic endocellulase and β-glucosidase from *Pyrrococcus furiosus*. *Biosci Biotechnol Biochem*. 2014;78(9):1537-1541. doi: 10.1080/09168451.

50. Smith JA, Lee BH. Glucosamylases: microbial sources, industrial applications and molecular biology - A Review. *J Food Biochem*. 1997;21:1-52.

51. Baker P, Whitfield GB, Hill PJ, et al. Characterization of the *Pseudomonas aeruginosa* glycoside hydrolase PaG reveals that its levels are critical for PaI polysaccharide biosynthesis and biofilm formation. *J Biot Chem*. 2015;290(47):28374-28387. https://doi.org/10.1074/jbc.M115.674929

52. Forbsberg Z, Bissaro B, Gullesen J, et al. Structural determinants of bacterial lytic polysaccharide monoxygenase functionality. *J Biot Chem*. 2018;293(4):1397-1412.

53. Lin WS, Cunneen T, Lee CY. Sequence analysis and molecular characterization of genes required for the biosynthesis of type 1 capsular polysaccharide in *Staphylococcus aureus*. *J. Bacteriol*. 1994;176:7005-7016.

54. Karunanithi PS, Zerbe P. Terpene syntheses as metabolic gatekeepers in the evolution of plant terpenoid chemical diversity. *Front. Plant Sci*. 2019;1. https://doi.org/10.3389/fpls.2019.01166

55. Christianson DW. Structural and chemical biology of terpenoid cyclases. *Chem Rev*. 2017;117(17):11570-11648.

56. Quin MB, Michel SN, Schmidt-Dannert C. Moonlighting metals: insights into regulation of cyclization pathways in fungal Δ(6)-protoilidene sesquiterpene syntheses. *Chembiochem*. 2015;16(15):2191-2199. doi: 10.1002/cbic.201500308

57. López-Gallego F, Warwryt GT, Schmidt-Dannert C. Selectivity of fungal sesquiterpene syntheses: role of the active site’s H-1 alpha loop in catalysis. *Appl Environ Microb*. 2010;76(23):7723-7733. doi: 10.1128/AEM.01811-10