

Analysis of the Genome Sequence of Strain GiC-126 of *Gloeostereum incarnatum* with Genetic Linkage Map

Wan-Zhu Jiang^a, Fang-Jie Yao^{a,b} , Ming Fang^b, Li-Xin Lu^b, You-Min Zhang^b, Peng Wang^c, Jing-Jing Meng^b, Jia Lu^a, Xiao-Xu Ma^a, Qi He^a, Kai-Sheng Shao^b, Asif Ali Khan^a and Yun-Hui Wei^d

^aInternational Cooperation Research Center of China for New Germplasm Breeding of Edible Mushrooms, Jilin Agricultural University, Changchun, China; ^bCollege of Horticulture, Jilin Agricultural University, Changchun, China; ^cEconomic Plants Research Institute, Jilin Academy of Agricultural Sciences, Gongzhuling, China; ^dInstitute of Agricultural Applied Microbiology, Jiangxi Academy of Agricultural Sciences, Nanchang, China

ABSTRACT

Gloeostereum incarnatum has edible and medicinal value and was first cultivated and domesticated in China. We sequenced the *G. incarnatum* monokaryotic strain GiC-126 on an Illumina HiSeq X Ten system and obtained a 34.52-Mb genome assembly sequence that encoded 16,895 predicted genes. We combined the GiC-126 genome with the published genome of *G. incarnatum* strain CCMJ2665 to construct a genetic linkage map (GiC-126 genome) that had 10 linkage groups (LGs), and the 15 assembly sequences of CCMJ2665 were integrated into 8 LGs. We identified 1912 simple sequence repeat (SSR) loci and detected 700 genes containing 768 SSRs in the genome; 65 and 100 of them were annotated with gene ontology (GO) terms and KEGG pathways, respectively. Carbohydrate-active enzymes (CAZymes) were identified in 20 fungal genomes and annotated; among them, 144 CAZymes were annotated in the GiC-126 genome. The A mating-type locus (*MAT-A*) of *G. incarnatum* was located on scaffold885 at 38.9 cM of LG1 and was flanked by two homeodomain (HD1) genes, *mip* and *beta-fg*. Fourteen segregation distortion markers were detected in the genetic linkage map, all of which were skewed toward the parent GiC-126. They formed three segregation distortion regions (SDR1–SDR3), and 22 predictive genes were found in scaffold1920 where three segregation distortion markers were located in SDR1. In this study, we corrected and updated the genomic information of *G. incarnatum*. Our results will provide a theoretical basis for fine gene mapping, functional gene cloning, and genetic breeding the follow-up of *G. incarnatum*.

ARTICLE HISTORY

Received 29 March 2021
Revised 6 July 2021
Accepted 7 July 2021

KEYWORDS

Gloeostereum incarnatum;
simple sequence repeat;
carbohydrate-active
enzymes; genetic linkage
map; mating-type

1. Introduction

The fruiting bodies of *Gloeostereum incarnatum* (family Cyphellaceae) are flower- or fan-shaped, pinkish-white or red, covered with a layer of hair, with verrucous protuberances, sterile handles, and wavy edges. *G. incarnatum* contains polysaccharides, sesquiterpenoids, phenols, and alcohols, and has antibacterial, anti-inflammatory, and antioxidant properties [1–4]. It was first domesticated and cultivated successfully in China and is a well-known wild edible and medicinal fungus in Northeast China. It has been popularized and is widely cultivated in China, and excellent new varieties have been bred [5,6]. At present, although it is widely cultivated in China, few cultivated strains are available, and most of the cultivated strains are domesticated from wild strains. Due to the long-term wild state, natural disasters and environmental selection, wild strains have a large number of excellent genes

such as disease resistance, insect resistance and stress resistance, which are important genetic resources for genetic improvement of *G. incarnatum*. In the process of evolution, rich genetic diversity was formed, which preserved the excellent genetic genes that the cultivated strains did not have or disappeared. The whole genome sequencing of wild strains is an important material for molecular breeding and basic genetic research [7,8]. Simple sequence repeats (SSRs) are widely distributed in genomes and are commonly used molecular markers in genetic diversity research. SSRs have been used in a variety of fungi, including *Auricularia heimuer*, *Auricularia cornea*, and *Trametes versicolor* [9–12]. *G. incarnatum* is a rare edible and medicinal fungus, so there are few reports about the application of molecular markers in *G. incarnatum*. Analysis of the number and distribution of SSRs in the *G. incarnatum* genome will be useful for applications in

CONTACT Fang-Jie Yao  yaofj1965@sina.com

 Supplemental data for this article can be accessed [here](#).

G. incarnatum germline identification, genetic diversity analysis, genetic linkage map construction, and assisted breeding. Sexual reproduction is an important part of its sexual life history, and classical genetic studies have shown that the *G. incarnatum* mating system is dipolar. However, little is known about the structure and coding sequences of this mating type, which has severely restricted the industrial development of *G. incarnatum*.

The wide application of whole-genome sequencing technologies has provided a theoretical basis for genetic breeding of fungi [12–14] and basic biological research [15,16]. Because of its high throughput and low cost [17], second-generation sequencing has been used widely to sequence the genomes of many fungi, including *Volvariella volvacea* [16], *Lentinula edodes* [18], *Flammulina filiformis* [19], and *Ganoderma lucidum* [15]. Third-generation sequencing does not require PCR amplification, which can reduce the difficulty of data assembly and produce larger fragments [20]. However, because of its high cost and a large number of sequencing errors, third-generation sequencing has been used in only a few fungi, such as *Agaricus bisporus* [21] and *A. heimuer* [22]. The whole genome of the monokaryotic strain CCMJ2665 of *G. incarnatum* has been sequenced using the PacBio Sequel technology [14].

On the basis of the research on germplasm evaluation, breeding of superior strains, nutritional physiology, and efficient cultivation [5,6,23,24], we selected the monokaryotic strain GiC-126 of *G. incarnatum* for whole-genome sequencing using Illumina HiSeq technology. The composition of SSR loci in the obtained *G. incarnatum* genome assembly sequence was analyzed, SSR-containing genes were screened against the GO and KEGG database. Genes were screened against the CAZy database (<http://www.cazy.org/>), carbohydrate-active enzymes of various fungi were compared, the mating-type structure was analyzed, and the coding sequences of mating-type genes were determined. We compared the GiC-126 genome assembly sequence with that of the published CCMJ2665 genome sequence [14] and integrated and corrected the assembly. The aim of this study was to update and improve the *G. incarnatum* genome assembly sequence and provide a theoretical basis for the genetic breeding of *G. incarnatum*.

2. Materials and methods

2.1. Strains and culture conditions

The monokaryotic strain GiC-126 was from wild strain GIC of *G. incarnatum* grown in Baishan City, Jilin Province, and the monokaryotic strain GiD-15 was from the cultivated strain GID of *G. incarnatum*

grown in Siping City, Jilin Province. All strains were preserved at the College of Horticulture, Jilin Agricultural University. The monokaryotic strains GiC-126 and GiD-15 were isolated from the mature fruiting bodies of *G. incarnatum* [25]. They were the starting strains for genome sequencing, genetic linkage map construction and QTL mapping of agronomic traits, and GiC-126 was used for de novo genome sequencing. GiC-126 and GiD-15 were inoculated in the same dish (90-mm diameter) containing PDA medium (1000 mL water, 200 g peeled potato, 20 g glucose, and 10 g agar powder) for hybridization; the interval was 2 cm and the incubation time was 10 days at 25 °C. F₁ hybrid strains were obtained and a genetic linkage map was constructed with the F₁ monokaryotic strains as the mapping population [26].

2.2. Genomic DNA extraction and assembly

A 0.5 mm × 0.5 mm section of GiC-126 was inoculated into a flask containing PD medium (1000 mL water, 200 g peeled potatoes, and 20 g glucose), and incubated in the dark for 20 days with shaking at 120 rpm. Genomic DNA of GiC-126 was extracted using the improved CTAB method [27]. Velvet v1.2 (<https://www.ebi.ac.uk/~zerbino/velvet/>) was used to assemble the genome sequence. The genome sequencing library contained 450-bp long reads and the sequencing depth was 176×. To ensure the accuracy of the analysis, low-complexity and low-quality adapters were removed from the original raw data.

2.3. Gene prediction and function annotation

We used GeneMark software [28] to predict potential protein-coding regions in the *G. incarnatum* genome sequence. To functionally annotate the predicted genes, we used BLAST to align the predicted gene sequences with known sequences in the following databases: SwissProt [29], National Center for Biotechnology Information's non-redundant (Nr), Gene Ontology (GO) [30], Kyoto Encyclopedia of Genes and Genomes (KEGG) [31,32], Karyotic Orthology Groups (KOG), Protein family database of alignments and hidden Markov models (Pfam) (<http://pfam.xfam.org/>), and Pathogen Host Interactions (PHI-base) (<http://www.phi-base.org/>).

2.4. SSR loci statistics

MISA software (<http://pgrc.ipk-gatersleben.de/misa/>) and the method of Murat et al. [33] were used to screen SSR loci in the *G. incarnatum* genome, with minimum number of mononucleotide repeats set as

10, minimum number of dinucleotide repeats as 6, and minimum numbers of trinucleotide to hexanucleotide repeats as 5.

2.5. Functional annotation and primer design of SSR-containing genes

We used the GO [30] and KEGG [31,32] databases to annotate and classify the SSR-containing genes in the *G. incarnatum* genome. Primer 3 software [34] with default parameters was used to design primers that can be used for germline identification, genetic diversity analysis, genetic linkage map construction, and assisted breeding.

2.6. Annotation of carbohydrate-active enzyme (CAZymes) genes

Using dbCAN2 software with default parameters (<http://bcb.unl.edu/dbCAN2/>) [35] for searches against the CAZymes database (<http://www.cazy.org/>), we identified and annotated 20 CAZymes of 19 fungi (Supplemental Table S1), including glycoside hydrolases (GHs), glycosyl transferases (GTs), carbohydrate esterases (CEs), polysaccharide lyases (PLs), auxiliary activity (AAs), and carbohydrate-binding modules (CBMs).

2.7. Preparation of population for mapping mating type

We used the whole-genome sequence of *G. incarnatum* GiC-126 (GenBank: GCA_010588315.1) to develop SSR molecular markers and construct a genetic linkage map of *G. incarnatum* [26]. In the F₁ mapping population, a monokaryotic strain Qx was selected randomly to mate with other strains, and the presence of clamp connections was observed under an optical microscope to determine the mating type of the mapping population. The mapping population was divided into two groups: one that produced clamp connections under an optical microscope, indicating affinity with Qx, and the other that did not produce clamp connections, indicating incompatibility with Qx.

2.8. Linkage analysis and genomic structure of the mating-type locus of *G. incarnatum*

JoinMap 4.0 software [36] was used to carry out linkage analysis of mating-type locus and to locate the mating-type locus on the genetic linkage map of *G. incarnatum*. The genetic linkage map was constructed as described in section 2.7 [26].

The mating-type locus sequence was found to be related to the A mating-type factor by blastx, blastp, and tblastn searches (E-value cutoff $<1 \times 10^{50}$). On the basis of the genetic linkage map and genome sequence data, we analyzed the structure of the *G. incarnatum* A mating-type locus and drew a structure map.

2.9. Data availability

The whole-genome sequence of the monokaryotic strain GiC-126 of *G. incarnatum* obtained in this study is available in GenBank under accession number GCA_010588315.1. The genome sequence of the monokaryotic strain CCMJ2665 of *G. incarnatum* is available in GenBank under accession number GCA_004338095.1.

3. Results

3.1. Genomic characteristics of GiC-126

The whole-genome sequence of the monokaryotic strain GiC-126 of *G. incarnatum* was 34.52-Mb long and comprised 5885 contigs, 49.20% GC content, and 16,895 predicted genes (Table 1). Functional annotation of the genome assembly sequence showed that 13,834 (81.88%) predicted genes matched homologous sequences in at least one of the searched databases. The remaining 3061 predicted genes had no obvious homologs in any of these databases (Supplemental Table S2–S9).

3.2. Comparison of the GiC-126 and CCMJ2665 genome assembly sequences

The GiC-126 genome sequence was about 4 Mb smaller than the published *G. incarnatum* CCMJ2665 genome sequence [14] and the GC content was similar in both sequences (Table 2). The

Table 1. Genome assembly statistics of *Gloeostereum incarnatum* strain GiC-126.

General features		Properties of predicted gene models	
Genome size (Mb)	34.52	Nr alignment	13,751
GC content (%)	49.20	KEGG alignment	10,504
Number of contigs	5885	Swissprot alignment	6358
Number of predict gene models	16,895	KOG alignment	5473
Number of exons	30,162	Pfam alignment	4996
Average exon size (bp)	432.54	GO alignment	3244
Number of introns	13,267	PHI alignment	2170
Average introns size (bp)	253.66	CAZy alignment	144

Table 2. Genomic characteristics of *Gloeostereum incarnatum* strains GiC-126 and CCMJ2665.

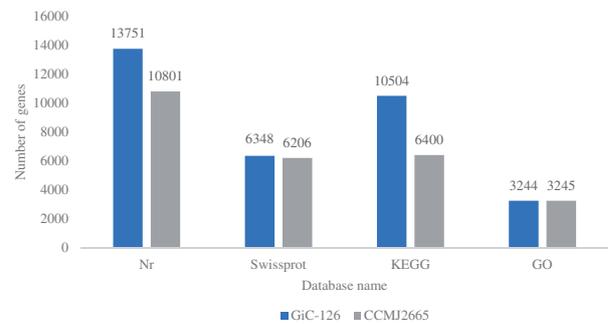
Strain	GiC-126	CCMJ2665
Genome assembly length (Mb)	34.52	38.65
GC content (%)	49.20	49.00
Number of genes	16,895	15,251
Number of scaffolds	1981	20
Number of contig	5885	3609
Genome coverage	176.0x	94.0x
Sequencing technology	Illumina HiSeq	PacBio Sequel
ssembly Method	Velvet v. 1.2	SOAPdenovo v. OCT-2018

numbers of Nr and KEGG annotations for GiC-126 were about 1.3 and 1.6 times those for CCMJ2665, respectively (Figure 1).

The scaffold and contig numbers for GiC-126 were about 100 and 1.6 times those for CCMJ2665. By integrating the data for the two genomes, we were able to integrate 1734 GiC-126 sequences into 19 of the CCMJ2665 sequences (Table 3, S10, Figure S1). Additionally, the *G. incarnatum* genome sequence was integrated using the GiC-126 genome sequence and its anchoring to the constructed genetic linkage map [26]. Ten linkage groups (LGs) and 15 CCMJ2665 sequences were integrated into eight LGs (Figure 2, Table S11). LG6 and LG9 were integrated into SCAFFOLD1, and LG3 and LG10 were integrated into SCAFFOLD5 of the CCMJ2665 genome. Additionally, SCAFFOLD2 and SCAFFOLD3 of the CCMJ2665 genome were integrated into LG1; SCAFFOLD6, SCAFFOLD8, SCAFFOLD14, and SCAFFOLD15 were integrated into LG2; SCAFFOLD7 was integrated into LG4; SCAFFOLD4 was integrated into LG5; SCAFFOLD9, SCAFFOLD11, and SCAFFOLD16 were integrated into LG7; and SCAFFOLD10 and SCAFFOLD12 were integrated into LG8. Most of the assembly sequences of GiC-126 in the same linkage map were anchored in the SCAFFOLDS of CCMJ2665, but there were three assembly sequences that corresponded to SCAFFOLDS in other LGs. Scaffold949 of the GiC-126 genome in LG1 was located in SCAFFOLD12 (LG8) of the CCMJ2665 genome; scaffold2019 in LG4 and scaffold125 of LG8 were located in SCAFFOLD1 (LG6 and 9); and scaffold431 of LG4 was not anchored in the CCMJ2665 genome assembly sequence.

3.3. Composition of the SSR loci in the *G. incarnatum* genome assembly sequence

A total of 1912 SSR loci were detected in the GiC-126 genome assembly sequence. Most of the SSRs were trinucleotides (813, accounting for 42.52% of the total), followed by mononucleotides and dinucleotides (535 (27.98%) and 415 (21.70%) respectively). The numbers of tetranucleotides,

**Figure 1.** Genome annotations of *Gloeostereum incarnatum* strains GiC-126 and CCMJ2665.**Table 3.** Integration of the genome assembly sequences of *Gloeostereum incarnatum* strains GiC-126 and CCMJ2665.

No.	Number of scaffolds	No.	Number of scaffolds
SCAFFOLD1	1046	SCAFFOLD11	41
SCAFFOLD2	4	SCAFFOLD12	38
SCAFFOLD3	42	SCAFFOLD13	10
SCAFFOLD4	59	SCAFFOLD14	30
SCAFFOLD5	70	SCAFFOLD15	27
SCAFFOLD6	51	SCAFFOLD16	12
SCAFFOLD7	92	SCAFFOLD17	14
SCAFFOLD8	40	SCAFFOLD18	5
SCAFFOLD9	66	SCAFFOLD19	1
SCAFFOLD10	16		

pentanucleotides, and hexanucleotides were relatively small, 85, 30, and 34 respectively (Table 4). Except for trinucleotides, the number of SSR loci trended to decrease as the numbers of nucleotides in the repeat increased. The number of SSR repetitions ranged from 5 to >20; five was the most common (552, accounting for 28.87% of the total), followed by 6 and 10 repetitions (414 (21.65%) and 318 (16.63%) respectively) (Table 4).

According to the SSR loci size distribution (Figure 3), most SSR loci were 11–15 bp (970, accounting for 50.73% of the total), followed by 16–20 bp, ≤10 bp, and 21–25 bp (341 (17.83%), 291 (15.22%), and 183 (9.47%) respectively). The number of SSR loci that were 26–30 bp and ≥30 bp was relatively small, 64 and 63 respectively. Long SSRs can have high polymorphism and mutation rates; therefore, in this study, we focused on SSR loci that were ≥30-bp long and designed 51 primer pairs. Using GiC-126 as template, PCR was verified by polyacrylamide gel electrophoresis, and 50 pairs of primers could amplify clear bands (Figure 4, Table S12).

3.4. Go classification of SSR-containing genes in the *G. incarnatum* genome assembly sequence

The GO database contains terms that describe the functions of homologous genes and gene products across species. A total of 700 genes containing 768 SSR loci were detected in the GiC-126 genome. Among the SSR-containing genes, 65 were

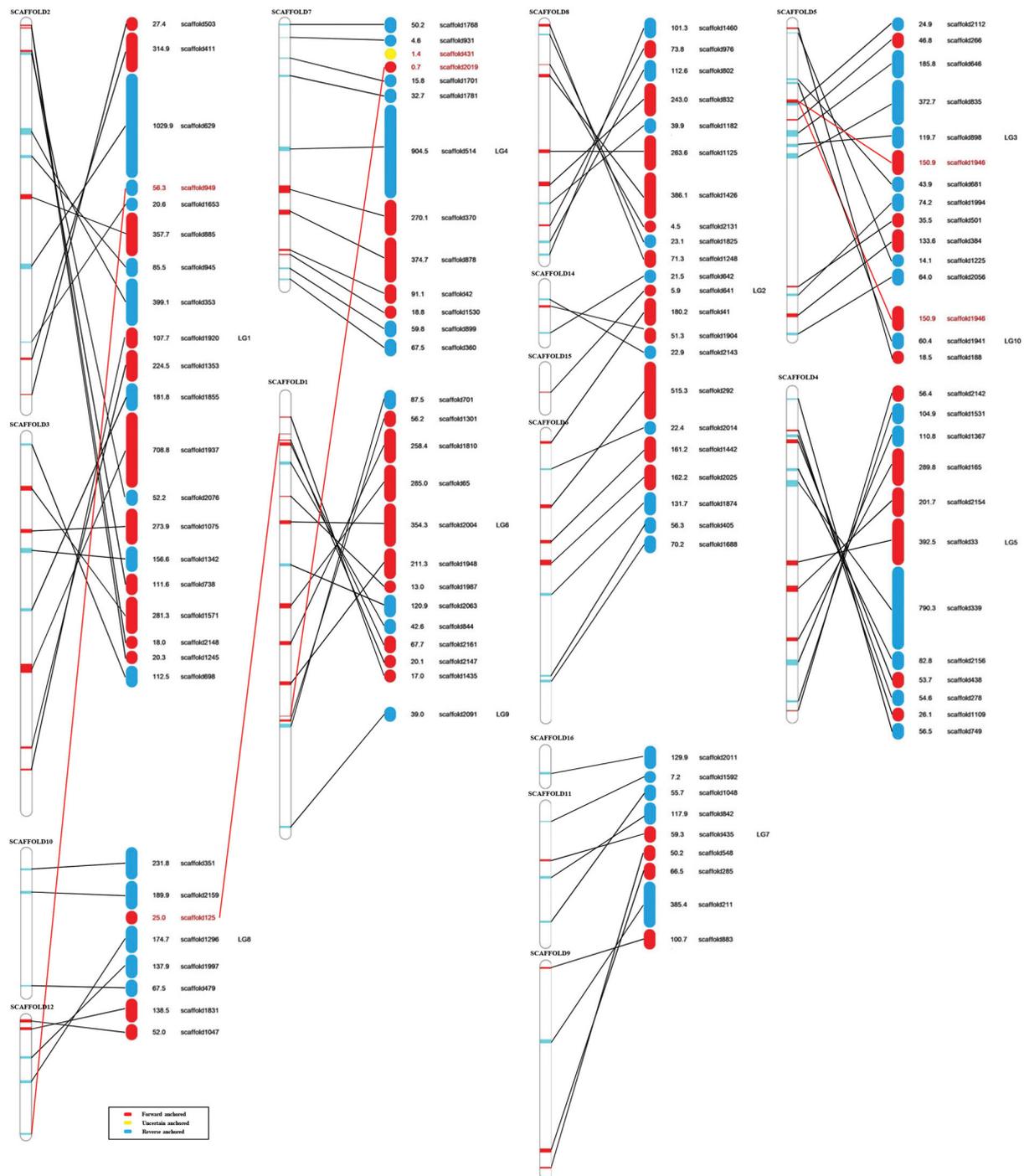


Figure 2. Integration of the *Gloeostereum incarnatum* genome assembly sequence based on genetic linkage map. Left, CCMJ2665 genome assembly sequence (SCAFFOLD) and length (kb); right, GiC-126 genome assembly sequence (scaffold) and length (kb). Number indicated the lengths of the scaffolds (kb). Red numbers and lines indicate the scaffold sequences where an anchoring problem occurred.

Table 4. Composition and number of SSR loci in the *Gloeostereum incarnatum* genome assembly sequence.

SSR type	Repeats																Total	
	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20		>20
Mononucleotide	–	–	–	–	–	291	111	62	32	9	7	3	5	5	2	1	7	535
Dinucleotide	–	207	78	52	29	10	12	8	4	3	2	4	2	1	1	–	2	415
Trinucleotide	464	178	87	42	13	13	6	4	4	–	1	–	–	1	–	–	–	813
Tetranucleotide	56	16	6	5	2	–	–	–	–	–	–	–	–	–	–	–	–	85
Pentanucleotide	17	4	3	3	1	1	–	–	–	1	–	–	–	–	–	–	–	30
Hexanucleotide	14	10	1	2	3	3	1	–	–	–	–	–	–	–	–	–	–	34
Total	551	415	175	104	48	318	130	74	40	13	10	7	7	7	3	1	9	1912

annotated with GO terms. Among them, 83, 45, and 49 genes were annotated with terms under the three main GO categories, biological process, cellular component, and molecular function, respectively (Figure 5). The most enriched terms under molecular function were “catalytic activity” (24 genes) and

“binding” (21 genes), respectively. Under cellular component, the most enriched terms were “cell,” “cell part,” and “macromolecular complex” with 10, 10, and 9 genes, respectively. Under biological process, the most enriched terms were “cellular process,” “metabolic process,” and “single-organism process” with 26, 18, and 17 genes, respectively.

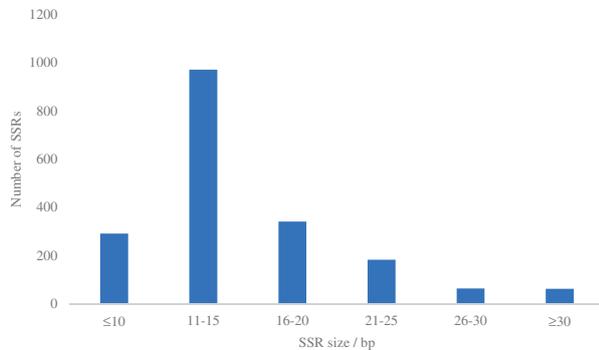


Figure 3. Size distribution of SSR loci in the *Gloeostereum incarnatum* genome assembly sequence.

3.5. KEGG classification of SSR-containing genes in the *G. incarnatum* genome assembly sequence

The KEGG database contains a lot of information, including graphical cell biochemical processes, chemical substances, enzyme molecules, and enzyme reactions. Among the 700 SSR-containing genes, 100 were annotated into 58 KEGG pathways (Figure 6). These pathways are distributed mainly in four categories, “cellular processes” (22 genes), “environmental information processing” (10 genes),

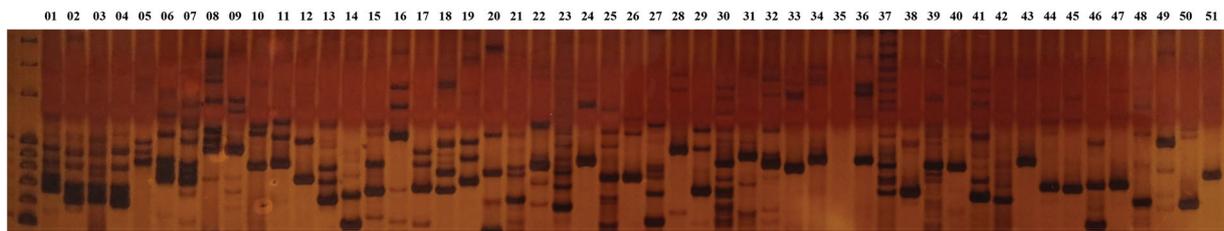


Figure 4. Electrophoresis of 51 pairs of SSR primers ≥ 30 bp.

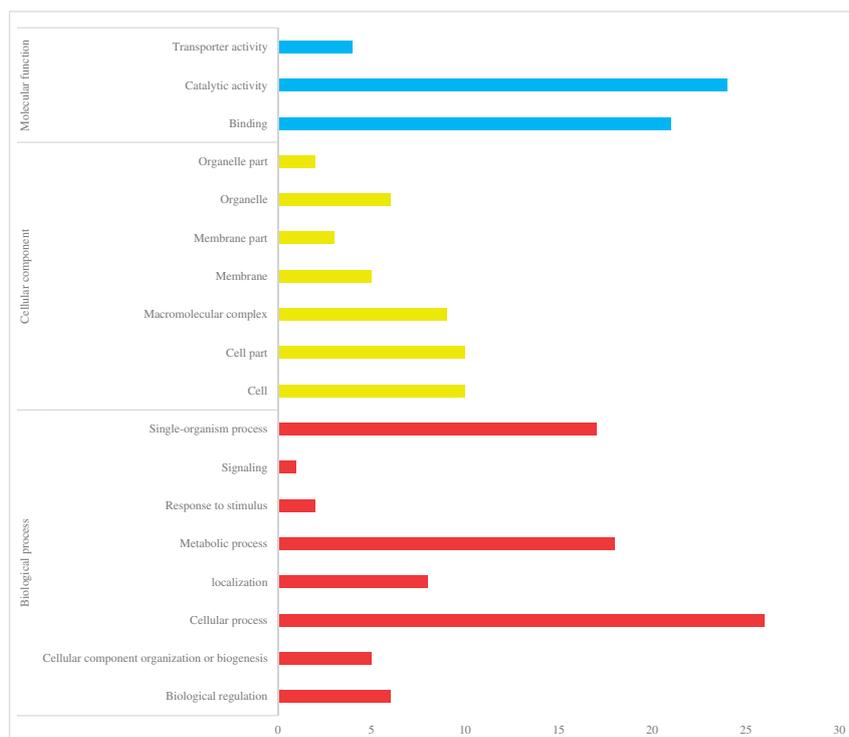


Figure 5. Gene ontology (GO) classification of SSR-containing genes in the *Gloeostereum incarnatum* genome assembly sequence.

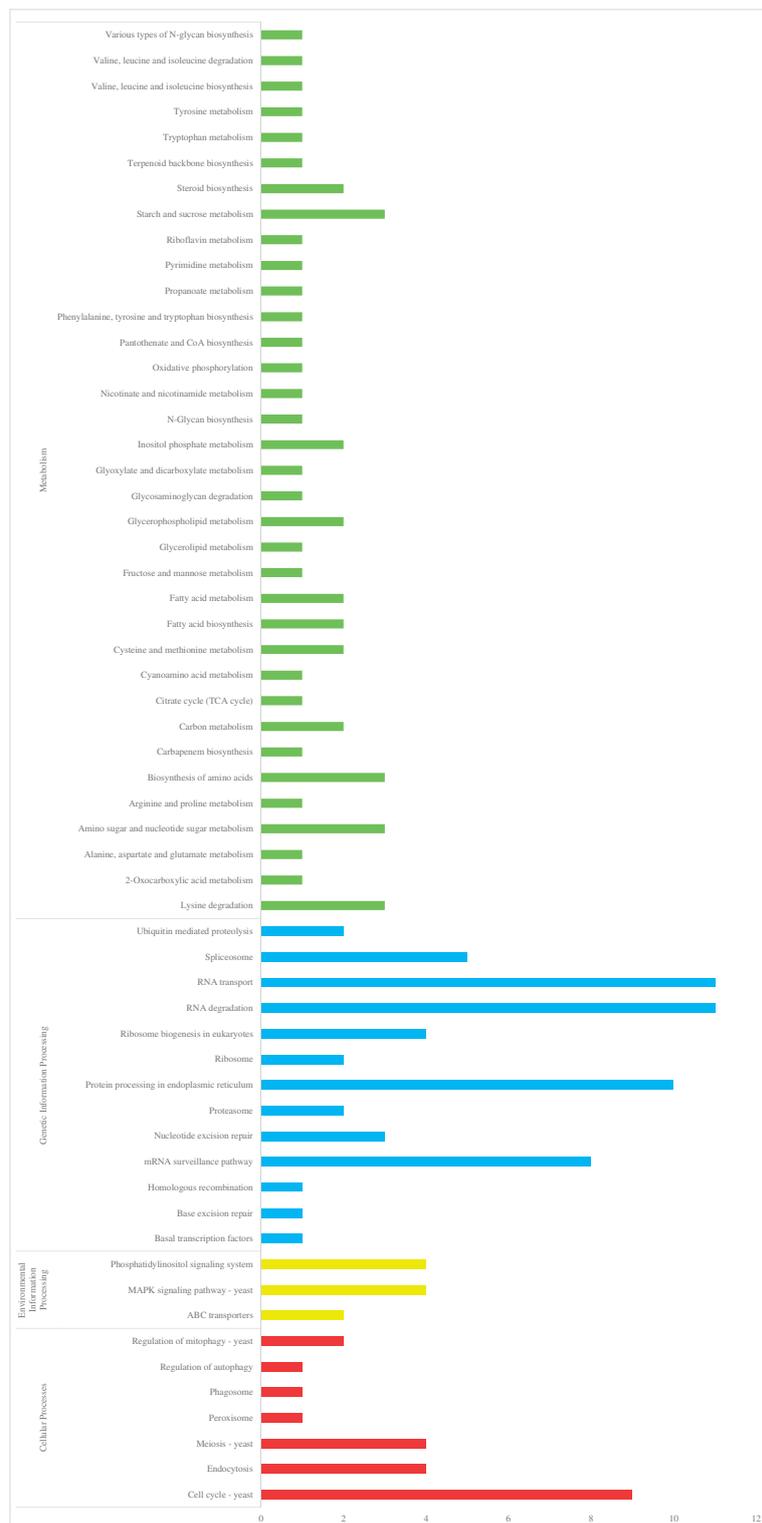


Figure 6. KEGG classification of SSR-containing genes in the *Gloeostereum incarnatum* genome assembly sequence.

“genetic information processing” (61 genes), and “metabolism” (50 genes). The genes assigned to “metabolism” were distributed in 34 subgroups; among them, “lysine degradation,” “amino sugar and nucleotide sugar metabolism,” “biosynthesis of amino acids,” and “starch and sucrose metabolism” had the most genes (3 in each subgroup). The genes assigned to “genetic information processing” were distributed in 13 subgroups; among them, “RNA

degradation,” “RNA transport,” “protein processing in endoplasmic reticulum,” and “mRNA surveillance pathway” had the most genes (11, 11, 10, and 8, respectively). The genes assigned to “cellular processes” were distributed in 7 subgroups; among them, “cell cycle-yeast” had the most genes (9). The genes assigned to “environmental information processing” were distributed in 3 subgroups; among them, “phosphatidylinositol signaling system,” “MAPK

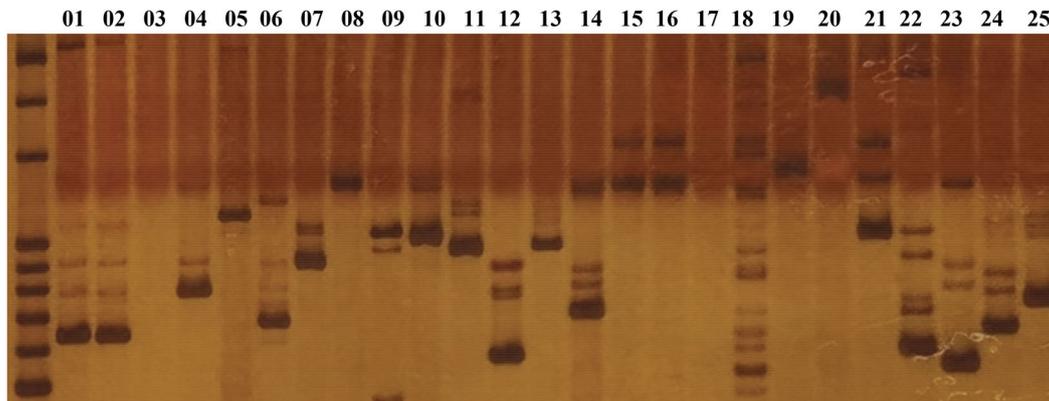


Figure 7. Electrophoresis of 25 pairs of SSR primers related to metabolism.

Table 5. Numbers of CAZymes in the two *Gloeostereum incarnatum* genomes compared with those in other fungi genomes.

Species	GH	GT	CE	PL	AA	CBM	sum	Species	GH	GT	CE	PL	AA	CBM	sum
<i>G. incarnatum</i> Gi-C126	51	31	16	11	30	5	144	<i>G. incarnatum</i> CCMJ2665	203	70	53	24	108	9	467
<i>A. bisporus</i>	159	54	59	10	92	12	386	<i>H. erinaceus</i>	150	66	56	7	88	6	373
<i>A. ostoyae</i>	253	73	86	28	150	25	615	<i>L. bicolor</i>	132	65	35	7	46	6	291
<i>A. cornea</i>	383	84	109	22	155	18	771	<i>L. edodes</i>	229	66	53	8	90	11	457
<i>A. heimuer</i>	301	63	88	19	116	25	612	<i>O. olearius</i>	187	62	54	10	87	7	407
<i>A. subglabra</i>	337	67	98	21	134	20	677	<i>P. ostreatus</i>	209	63	59	23	135	30	519
<i>C. cinereus</i>	175	72	72	13	128	21	481	<i>S. commune</i>	226	73	63	17	81	11	471
<i>F. filiformis</i>	216	75	70	24	96	7	488	<i>T. fuciformis</i>	121	65	50	2	17	4	259
<i>G. lucidum</i>	250	62	57	9	104	6	488	<i>T. mesenterica</i>	67	65	14	2	12	4	164
<i>G. luxurians</i>	302	82	105	13	159	27	688	<i>V. volvacea</i>	201	63	51	28	118	23	484

CAZy Family: GH: glycoside hydrolase; GT: glycosyltransferase; CE: carbohydrate esterases; PL: polysaccharide lyases; AA: auxiliary activities; CBM: carbohydrate-binding modules.

signaling pathway-yeast,” and “ABC transporters” had the most genes (4, 4, and 2 respectively).

G. incarnatum is rich in polysaccharides, amino acids, fatty acids, terpenes, and inorganic salts [2], many of which are related to KEGG “metabolism” subgroups such as “terpenoid backbone biosynthesis,” “fatty acid biosynthesis,” “starch and sucrose metabolism,” and “fructose and mannose metabolism”. We designed 25 primer pairs of primers for the SSR-containing genes annotated to “metabolism” to facilitate subsequent studies on active ingredients, fine gene mapping, and genetic breeding of *G. incarnatum*. Using GiC-126 as template, PCR was verified by polyacrylamide gel electrophoresis, and 23 pairs of primers could amplify clear bands (Figure 7, Table S13).

3.6. Carbohydrate active enzymes (CAZymes)

We compared the numbers and types of CAZymes in the two *G. incarnatum* genomes with those in 20 genomes of 19 species, including three straw-rotting fungi (*A. bisporus*, *Coprinus cinereus*, and *V. volvacea*), one mycorrhizal fungus (*Laccaria bicolor*), and 15 wood-rotting fungi. Among them, the number of CAZymes annotated in the GiG-126 genome was the least; the number in the CCMJ2665 genome was in the middle. Among the CAZymes, the number of AA (108) was higher than the average (97), the number of CBM (9) was lower than the average

(15), and the other numbers were similar to the mean (Table 5, Table S14). A total of 144 CAZymes were detected in the GiC-126 genome: 51 GHs, 31 GTs, 16 CEs, 11 PLs, 30 AAs, and 5 CBMs (Table 5).

A total of 17 enzyme families involved in wood degradation were identified by Floudas [37]. We detected 35 genes from 11 families in the GiC-126 genome, and 102 genes from 13 families in the CCMJ2665 genome; genes encoding GH11 and CE8 were not detected in either of the *G. incarnatum* genomes (Supplemental Table S15). White-rot fungi are rich in crystalline cellulose-degrading enzymes, including GH6, GH7, CBM1, and AA9 (formerly GH61). In the 20 genomes of 19 fungi, we detected numerous CAZymes related to crystalline cellulose degradation (Supplemental Table S16). Among them, *L. bicolor*, *Tremella mesenterica*, and *Tremella fuciformis* had very low levels of crystalline cellulose-degrading enzymes, and were classified as brown-rot fungi, the rest of the fungi were classified as white-rot fungi. Among the wood-rot fungi, *Pleurotus ostreatus* had the highest number of genes related to the degradation of crystalline cellulose, followed by *Auricularia subglabra*, *A. cornea*, and *A. heimuer*. *V. volvacea* and *C. cinereus* had the highest numbers of genes related to the degradation of crystalline cellulose among the wood-rotting fungi and straw-rotting fungi, which is in accordance with the strong ability of straw-rotting fungi to degrade

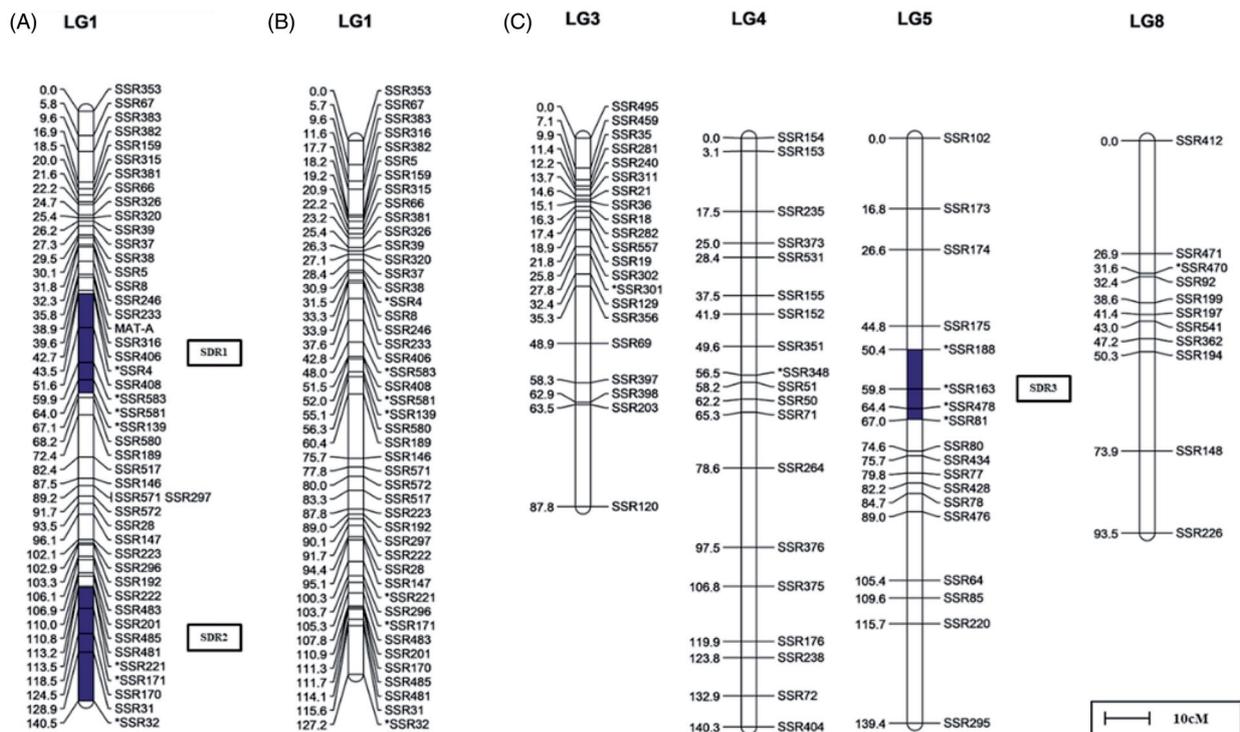


Figure 8. Linkage group of the segregation distortion site. (A) Linkage group with the *MAT-A* locus. (B) Linkage group before the addition of the *MAT-A* locus. (C) Linkage group of the segregation distortion site (except LG1).

crystalline cellulose. These results are consistent with those reported by Fang [38].

The two *G. incarnatum* genomes contain many AA genes. In the GiC-126 genome, we found 30 genes that encode 9 AAs. The highest numbers of AA-encoding genes were for AA3 (8 genes) and AA9 (10 genes), accounting for 60.00% of the total number of AAs. In the CCMJ2665 genome, we detected 108 genes that encode 11 AAs. The highest numbers of AA-encoding genes were for AA1, AA3, AA7, and AA9 with 14, 31, 12, and 28 genes, respectively, accounting for 78.70% of the total number of AAs. In the two *G. incarnatum* genomes, the highest number of AA-encoding genes were for the AA9 family; no AA11-encoding genes were found.

3.7. Construction of genetic linkage map and location of mating-type locus

The F_1 generation of Gi126-15 (hybrid of GiC-126 \times GiD-15) was observed under an optical microscope in clamp connections, and 114 monokaryotic strains were isolated. To determine the mating type of the mapping population, the 114 strains (mapping population) were mated with two mapping parents GiC-126 (A_1) and GiD-15 (A_2). Among them, 61 members of the mapping population were incompatible with GiC-126 and compatible with GiD-15, the mating-type was A_1 , and 53 members of the mapping population were compatible with GiC-126 and incompatible with GiD-15,

the mating-type was A_2 (Supplemental Table S17). A genetic linkage map was reconstructed using JoinMap software. The A mating-type locus (named *MAT-A*) marker was added to LG1 and the position of the LG changed from 127.2 cM to 140.5 cM. The *MAT-A* locus marker was located between markers SSR233 and SSR316, and the distances between SSR233 and *MAT-A* and *MAT-A* and SSR316 were 3.1 cM and 0.7 cM, respectively (Figure 8(A)).

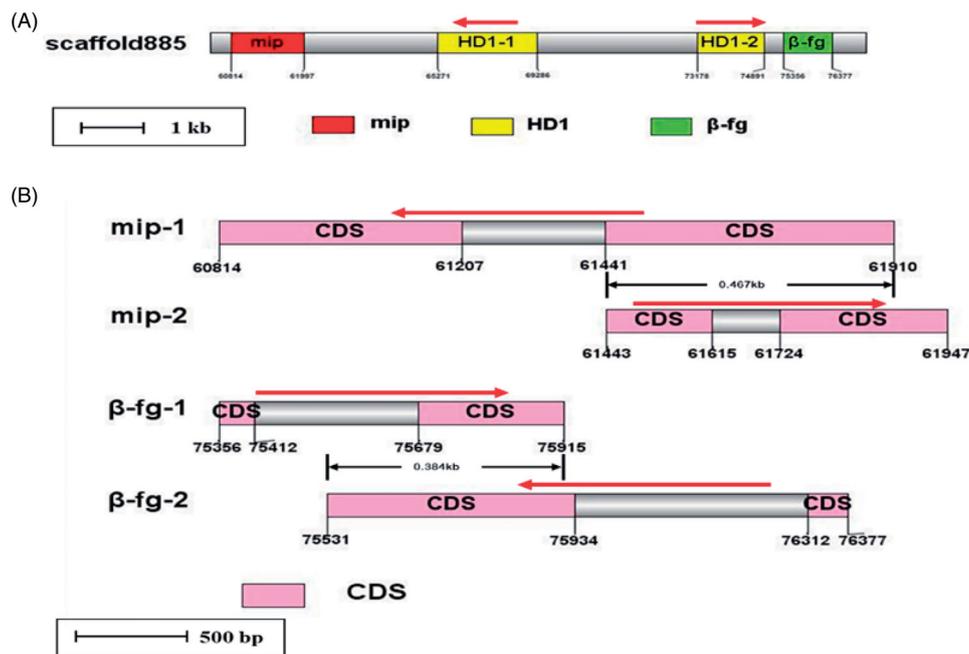
3.8. Segregation distortion

In the original genetic linkage map of *G. incarnatum* [26], a total of 14 segregation distortion markers were found, accounting for 7.6% of the total number of markers. All 14 markers were skewed toward the parent GiC-126. Among the seven segregation distortion markers of LG1, three were clustered between 48.0 cM and 55.1 cM, forming a segregation distortion region (SDR), and the other four were scattered. LG3, LG4, and LG8 had one segregation marker each. The four segregation distortion markers on LG5 were clustered between 50.4 cM and 67.0 cM, with a span of 26.6 cM. There also was a SDR on LG5 (Figure 8(B), 8(C)). After the addition of the *MAT-A* locus marker, LG1 changed from 127.2 cM to 140.5 cM and two SDRs were formed on LG1. SDR1 contained four segregation distortion markers with a span of 23.5 cM, and SDR2 contained three segregation distortion markers with a span of 27.0 cM (Figure 8(A), Table 6).

Table 6. Segregation distortion sites of SSRs in the *Gloeostereum incarnatum* linkage map after adding the *MAT-A* locus.

Locus name	Genotype		χ^2 Value	Direction of skewed	Linkage group		scaffold		
	A	B			LG	Pos/cM	Pos/cM	Start	End
SSR4	74	40	10.14	A	LG1	43.5	scaffold949	45.227	45.258
SSR583	72	42	7.89	A	LG1	59.9	scaffold1920	104.676	104.687
SSR581	70	44	5.93	A	LG1	64	scaffold1920	99.686	99.697
SSR139	68	46	4.25	A	LG1	67.1	scaffold1920	60.271	60.303
SSR221	68	46	4.25	A	LG1	113.5	scaffold1075	20.503	20.526
SSR171	69	45	5.05	A	LG1	118.5	scaffold1571	249.249	249.272
SSR32	72	42	7.89	A	LG1	140.5	scaffold698	93.528	93.549
SSR301	68	46	4.25	A	LG3	27.8	scaffold681	29.16	29.171
SSR348	68	46	4.25	A	LG4	56.5	scaffold514	416.806	416.823
SSR188	68	46	4.25	A	LG5	50.4	scaffold1367	13.386	13.409
SSR163	69	45	5.05	A	LG5	59.8	scaffold165	155.683	155.7
SSR478	72	42	7.89	A	LG5	64.4	scaffold2154	199.829	199.849
SSR81	72	42	7.89	A	LG5	67	scaffold33	143.301	143.327
SSR470	68	46	4.25	A	LG8	31.6	scaffold2159	40.365	40.385

(A) Skewed toward GiC-126. (B) Skewed toward GiD-15.

**Figure 9.** Structure of the *MAT-A* in the GiC-126 genome. (A) Structure of *MAT-A*. (B) Structures of the *mip* and *beta-fg* genes in *MAT-A*.**Table 7.** Genes in the *MAT-A* in the *Gloeostereum incarnatum* genome assembly sequence.

No.	Gene	Mating-type locus	Length (bp)
1	HDTF-1 (C12611655)	scaffold353	693
2	HD1-1 (C12616611)	scaffold885	2436
3	HD1-2 (C12616617)	scaffold885	1644
4	mip-1 (C12616608)	scaffold885	864
5	mip-2 (C12616609)	scaffold885	447
6	beta-fg-1 (C12616618)	scaffold885	294
7	beta-fg-2 (C12616619)	scaffold885	510

Among the 14 segregation distortion markers, SSR139, SSR581, and SSR583 were located between 60.271 kb and 104.687 kb in scaffold1920. We found 22 predictive genes in this region, and 19 of them were annotated from at least one of the databases (Nr, KEGG, SwissProt, Pfam, KOG) as follows: 9 aryl-alcohol dehydrogenases, 7 hypothetical proteins, 2 glucooligosaccharide oxidases, and 1 ribosome biogenesis regulatory. The results of the blastx,

blastp, and tblastn of 22 genes were basically consistent with the annotation results (Supplemental Table S18).

3.9. Structure of the *a* mating-type locus (*Mat-A*)

By comparing the blastx, blastp, and tblastn results, we found one homeodomain transcription factor gene C12611655 (HDTF-1) on scaffold353 and two homeodomain genes C12616611 (HD1-1) and C12616617 (HD1-2) on scaffold855 (Supplemental Table S19). According to the genome data and Figure S1, scaffold629 and scaffold945 were connected to the *MAT-A* locus, and scaffold885 was located exactly between scaffold629 and scaffold945, and scaffold353 was near scaffold945. By comparing the sequence of the mating-type factor in the GiC-126 scaffolds and the location of the *MAT-A* on the genetic linkage map, we found that the *MAT-A* in

the *G. incarnatum* genome contained two HD1 genes with a distance between them of 3.89 kb and length ranging from 1644 to 2436 bp (Figure 9(A), Table 7).

The red arrow indicates the transcription direction of the gene. The gray bar indicates the region containing the mating-type genes in scaffold885 and the numbers below show the positions of the two ends of the mating-type gene on scaffold885. There were 0.467 kb and 0.384 kb overlaps between the two *mip* and two *beta-fg* genes. *mip*, mitochondrial intermediate peptidase; HD, homeodomain; *beta-fg*, beta-flanking protein; CDS, coding sequence.

The genes encoding mitochondrial intermediate peptidase (*mip*) and *beta-fg* are located on both sides *MAT-A* and are closely linked to the *MAT-A* in the GiC-126 genome (Figure 9(A)). The *mip-1* (C12616608) and *mip-2* (C12616609) genes were 864-bp and 447-bp long, respectively, and there was a 467-bp overlap between *mip-1* and *mip-2*. The *beta-fg-1* (C12616618) and *beta-fg-2* (C12616619) genes were 294-bp and 510-bp long, respectively, and there was a 384-bp overlap between *beta-fg-1* and *beta-fg-2* (Table 7, Supplemental Table S20). In the *G. incarnatum* genome, *mip* and *beta-fg* were located on both sides of the *MAT-A*; *mip-1* and *mip-2* were located at 3.36 kb and 3.27 kb upstream of HD1-1, and *beta-fg-1* and *beta-fg-2* were located at 0.46 kb and 0.64 kb downstream of HD1-2, respectively (Figure 9(B)).

4. Discussion

4.1. Comparative analysis of the second-generation and third-generation sequencing data of *G. incarnatum*

Strain GiC-126 used in this study is from the wild strain GiC in Baishan City, Jilin Province. Strain CCMJ2665 is a commercially cultivated strain, so the two genomes can be regarded as representatives of different strains. The GiC-126 genome was about 4 Mb shorter than the CCMJ2665 genome, and there were some differences in the functional annotations of the predicted genes, which are related to the different sequencing methods used. These findings are similar to those reported in *G. lucidum* [15,39] and *V. volvacea* [16,40]. In this study, 1734 assembly sequences of GiC-126 were initially integrated into 19 assembly sequences of CCMJ2665, and the two genomes were further integrated through a genetic linkage map. Ten LGs and fifteen assembly sequences of CCMJ2665 were integrated into eight LGs. One LG was integrated into one to four CCMJ2665 assembly sequences, and different LGs were integrated into one CCMJ2665 assembly sequence. In future studies, increased numbers of SNP and Indel

markers with good genetic stability and high polymorphism are needed to improve the distribution density of the markers and to integrate the genomes more accurately.

4.2. Characterization of the SSR markers

We used bioinformatics methods to comprehensively analyze the number, distribution, and functions of the SSR in the *G. incarnatum* genome. We obtained 1912 SSRs, among which trinucleotide SSRs were the most common, which is consistent with the results for *G. lucidum*, *T. versicolor*, and *F. filiformis*, but different for *A. bisporus* in which mononucleotide are the most [9,41]. We also detected 700 SSR-containing genes in the *G. incarnatum* genome. These data provide a basis for exploring the role of SSRs in the genome structure, for the development of *G. incarnatum* molecular markers, and for population evolution and genetic diversity research.

The detected SSR-containing genes were enriched in GO terms, including “binding” and “catalytic activity” under molecular function and “metabolic process” and “cellular process” under biological process, which is consistent with the GO annotation results for *G. lucidum* [41], *V. volvacea* [42], *P. ostreatus* [43], and *T. versicolor* [9]. Additionally, the SSR-containing genes were enriched mainly in two KEGG categories, “metabolism” and “genetic information processing,” which is consistent with the results for *T. versicolor* [9]. These results indicate that the functions of the SSR-containing genes in *G. incarnatum* genome were related mainly to basal metabolism. On the basis of these annotations, we identified genes related to the “terpenoid backbone biosynthesis,” “fatty acid biosynthesis,” “starch and sucrose metabolism,” and “fructose and mannose metabolism” pathways and designed primers for them. To study the roles of the SSR-containing genes in the biosynthesis of medicinal active substances in *G. incarnatum*, in the future, fine gene mapping and cloning can be carried out according to the annotation results and designed primers.

4.3. Analysis of CAZymes

We analyzed the CAZymes in the genomes of GiC-126 using the dbCAN2 database, and the results of Wang [10] based on the Blastp and CAZy database, the number of genes identified in the GiC-126 genome was one-third of that reported by Wang in the CCMJ2665 genome [10]. The number of genes annotated as GH differed the most, followed by AA with differences of 152 and 78, respectively, between the two *G. incarnatum* genomes. The types of

CAZymes also differed greatly between the two genomes. The number of GHs and GTs annotated in the GiC-126 genome was about one-half that of the CCMJ2665 genome; the numbers of the other types were similar. The results of the CAZymes analysis in this study also differed from the results for *A. heimuer* [38] and *F. filiformis* [19], likely because of the different annotation pipelines that were used. In this study, we also used the Pfam database to compare and analyze these genomes and found that, compared with dbCAN2, fewer CAZymes families were detected, so we did not report these data (Supplemental Table S21). Although the correspondence between the CAZymes families and the Pfam/Interpro or dbCAN2 families is known to be far from perfect [44], the differences in these results provide additional information for the development of enzymes and offer more candidate genes for further research.

In most of the fungal genomes that we compared, the numbers of genes in the AA3, AA7, and AA9 families were significantly higher than in the other AA families, accounting for about 55%–71% of the total number of AA families, followed by the AA1 and AA2 (class II lignin-modifying peroxidases, POD) families. White-rot and brown-rot fungi are classified according to their ability as wood-rotting fungi to degrade plant cell wall lignin. The POD content and number of crystalline cellulose-degrading enzymes are higher in white-rot fungi than in brown-rot fungi [45]. In this study, we identified PODs and a large number of crystalline cellulose-degrading enzymes in the genome of *G. incarnatum*, white-rot fungi, which is consistent with the results of Wang [10]. Additionally, the genes encoding members of the CBM1, GH6, GH7, AA9, and AA1 families play important roles in the degradation of lignocellulose in *G. incarnatum* and can be used as core candidate genes for nutrient utilization and molecular marker development of *G. incarnatum*. In future studies, transcriptome analysis of the different developmental stages of *G. incarnatum* can be used to detect the most active lignocellulose degradation genes.

4.4. Mating-type locus of *G. incarnatum*

The *MAT-A* was located at 38.9 cM of LG1 on the constructed genetic linkage map. By searching the genome assembly sequence between SSR233 (scaffold945) and SSR316 (scaffold629), we found that two genes that control the *MAT-A* were in scaffold885. The anchoring results located scaffold885 between scaffold945 and scaffold629, confirming the accuracy of the results and showing that the two methods, genome sequencing and genetic linkage

map, produced the same results. The bioinformatics analysis identified eight pheromone receptor genes controlling B mating-type locus in scaffold1997 (LG8) of the GiC-126 genome, and no pheromone precursor genes controlling B mating-type locus were detected. The genes controlling the *MAT-A* were located in scaffold885 (LG1) of the linkage map constructed based on the GiC-126 genome, and on SCAFFOLD2 of the CCMJ2665 genome. The genes controlling the B mating-type locus were on scaffold1997 (LG8) of the linkage map constructed based on the GiC-126 genome, and on SCAFFOLD12 of the CCMJ2665 genome, showing that the scaffolds or LGs that control the A and B mating-type locus genes are not the same. Further, LG1 and LG8 were not integrated in the genetic linkage map integration of the *G. incarnatum* genome assembly sequence; therefore, we consider that the A and B mating-type loci are not on the same chromosome. Separate A and B mating-type loci also have been found in *Coprinellus disseminatus*, *Pholiota microspora*, *Phanerochaete chrysosporium*, and *A. heimuer* [38,46–48].

We also found 467-bp and 384-bp overlaps between the adjacent *mip* (C12616608 and C12616609) and *beta-fg* (C12616618 and C12616619) genes, respectively, possibly because the transcription directions of the genes are different. The *mip* and *beta-fg* genes of *G. incarnatum* flanked *MAT-A* and were closely related to homologous genes of *F. filiformis*, *C. cinereus*, *L. bicolor*, *Coprinellus disseminatus*, *P. chrysosporium*, *A. bisporus*, *V. volvacea*, and *Postia placenta* [16,46,49–51]. In basidiomycetes, the distance between the *MAT-A* gene and *mip* locus has been reported to be closely linked and conserved within a range of <1 kb. However, the distance between the *MAT-A* gene and *mip* locus in the *G. incarnatum* genome is not as conserved as it is in the genomes of the model mushrooms *C. cinereus* [52] and *Schizophyllum commune* [53], but it is much shorter than it is in the genomes of *A. heimuer* strains A14-8 (59.5 kb) and Dai 13,782 (67.6 kb) [38]. Mating-type genes are essential for the formation of binucleated mycelium and the development of fruiting bodies. Understanding the structure and synchronization of *MAT-A* will provide a reference for the evolution of different species of *G. incarnatum* and other fungi.

4.5. Segregation distortion

Segregation distortion is a common phenomenon in genetic mapping, and the proportion of segregation distortion varies greatly with species, population types, and marker types. The 14 segregation distortion markers produced in this study are all skew

toward the parent strain GiC-126. Segregation distortion markers skew toward one parent have been reported in many crops [54,55]. In this study, 22 predictive genes were found in scaffold1920 where three segregation distortion markers also were located, which indicates that these predictive genes may be related to segregation distortion. In future studies, more segregation distortion markers could be detected by increasing the marker types (SRAPs, SNPs, Indels) and/or mapping population types (F_2 , BC, DH, RIL) to provide a theoretical basis for the study of segregation distortion.

5. Conclusions

Complete fungal genome sequences provide an effective basis for the study of genetic mechanisms and for genetic breeding. In this study, we obtained a 34.52-Mb genome assembly sequence of strain GiC-126 of *G. incarnatum*, and compared it with the published genome sequence of strain CCMJ2665 of *G. incarnatum*. The two genomes were combined with a genetic linkage map to integrate and update the genome assembly sequence of *G. incarnatum*. The composition of SSR loci in the GiC-126 genome was analyzed, and SSR-containing genes were screened out and annotated using the GO and KEGG databases. Our results will provide a theoretical basis for research on genetic diversity and assisted breeding. The CAZymes of 20 fungal genomes were compared, and the genes that encode lignocellulosic enzymes were analyzed and screened to lay a foundation for the cultivation of edible fungi that efficiently degrade lignocellulose. The structure of the *MAT-A* was clarified, the coding sequence of the mating-type gene was determined, and the structure of *MAT-A* was analyzed. The results will help to improve the efficiency of genetic breeding and new variety selection. This study aimed to correct and update the *G. incarnatum* genome sequence and provide a theoretical basis for genetic breeding and improving the medicinal value of *G. incarnatum*.

Acknowledgment

We thank Margaret Biswas, PhD, from Liwen Bianji, Edanz Editing China (www.liwenbianji.cn/ac), for editing the English text of a draft of this manuscript.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was funded by the project of “one thousand talents plan” in Jiangxi Province and the National Key Research and Development Plan of China [2020YFD1000300].

ORCID

Fang-Jie Yao  <http://orcid.org/0000-0001-8581-1838>

References

- [1] Li YT, Song H, Li YQ, et al. Study on antioxidant properties of alcohol extract of *Gloeostereum incarnatum* fermentation. *J Fungal Res.* 2010;8:90–92.
- [2] Zhang ZF, Lv GY, Jiang X, et al. Extraction optimization and biological properties of a polysaccharide isolated from *Gloeostereum incarnatum*. *Carbohydr Polym.* 2015;117:185–191.
- [3] Wang D, Li Q, Qu YD, et al. The investigation of immunomodulatory activities of *Gloeostereum incarnatum* polysaccharides in cyclophosphamide-induced immunosuppression mice. *Exp Ther Med.* 2018;15(4):3633–3638.
- [4] Li X, Liu X, Zhang YF, et al. Protective effect of *Gloeostereum incarnatum* on ulcerative colitis via modulation of Nrf2/NF- κ B signaling in C57BL/6 mice. *Mol Med Rep.* 2020;22(4):3418–3428.
- [5] Yu Y, Yao FJ, Sun ML, et al. A new *Gloeostereum incarnatum* cultivar ‘Qirou 1. *Acta Horticulturae Sin.* 2013;40:15–15.
- [6] Yu Y, Yao FJ, Zhang YM. A new *Gloeostereum incarnatum* cultivar ‘Jirou 1. *Acta Horticulturae Sin.* 2016;43:1013–1014.
- [7] Gao YL, He D, Wei YY, et al. Research progress on fungal genetic methods. *J Fungal Res.* 2019;17:173–179.
- [8] Zhang Y, Huang CY, Gao W. Research advances on molecular mushroom breeding. *J Fungal Res.* 2019;17:229–239.
- [9] Wang DD, Li LQ, Ma LY, et al. Progress in development and applications of SSR molecular marker in macrofungi. *Microbiology.* 2013;40:646–654.
- [10] Cao Y, Zhang YZ, Cheng SJ, et al. Genome-wide distributional and comparative analysis of SSR loci in *Trametes versicolor*. *Mycosystema.* 2017;36:1524–1542.
- [11] Lu LX, Yao FJ, Wang P, et al. Construction of a genetic linkage map and QTL mapping of agronomic traits in *Auricularia auricula-judae*. *J Microbiol.* 2017;55(10):792–799.
- [12] Wang P, Yao FJ, Lu LX, et al. Map-based cloning of genes encoding key enzymes for pigment synthesis in *Auricularia cornea*. *Fungal Biol.* 2019;123(11):843–853.
- [13] Yao FJ, Lu LX, Wang P, et al. Development of a molecular marker for fruiting body pattern in *Auricularia auricula-judae*. *Mycobiology.* 2018;49:72–78.
- [14] Wang XX, Peng JY, Sun L, et al. Genome sequencing illustrates the genetic basis of the pharmacological properties of *Gloeostereum incarnatum*. *Genes.* 2019;10(3):188.

- [15] Chen SL, Xu J, Liu C, et al. Genome sequence of the model medicinal mushroom *Ganoderma lucidum*. *Nat Commun.* 2012;3:913.
- [16] Bao DP, Gong M, Zheng HJ, et al. Sequencing and comparative analysis of the straw mushroom (*Volvariella volvacea*) genome. *PLOS One.* 2013; 8(3):e58294.
- [17] Bau T, Lu T. Advance of macro-fungal genomes Ssequencing. *J Fungal Res.* 2017;15:151–165.
- [18] Au CH, Wong MC, Qin J, et al. Genome sequence and genetic linkage analysis of Shiitake mushroom *Lentinula edodes*. *Nat Prec.* 2012. DOI:10.1038/npre.2012.6855.1
- [19] Park YJ, Baek JH, Lee S, et al. Whole genome and global gene expression analyses of the model mushroom *Flammulina velutipes* reveal a high capacity for lignocellulose degradation. *PLoS One.* 2014;9(4):e93560.
- [20] Eid J, Fehr A, Gray J, et al. Real-time DNA sequencing from single polymerase molecules. *Science.* 2009;323(5910):133–138.
- [21] Sonnenberg AS, Gao W, Lavrijsen B, et al. A detailed analysis of the recombination landscape of the button mushroom *Agaricus bisporus* var. *bisporus*. *Fungal Genet Biol.* 2016;93:35–45.
- [22] Yuan Y, Wu F, Si J, et al. Whole genome sequence of *Auricularia heimuer* (Basidiomycota, Fungi), the third most important cultivated mushroom worldwide. *Genomics.* 2019;111:50–58.
- [23] Song H, Yao FJ, Tang J, et al. Research overview of *Gloeostereum incarnatum*. *Chin Edible Fungi.* 2008;27:3–4.
- [24] Yu Y, Yao FJ, Sun ML, et al. Spring and autumn cultivation management technology of *Gloeostereum incarnatum*. *Northern Horticulture.* 2013;10:145–146.
- [25] Choi YW, Hyde KD, Ho WWH. Single spore isolation of fungi. *Fungal Divers.* 1999;3:29–38.
- [26] Jiang WZ, Yao FJ, Lu LX, et al. Genetic linkage map construction and quantitative trait loci mapping of agronomic traits in *Gloeostereum incarnatum*. *J Microbiol.* 2021;59(1):41–50.
- [27] Watanabe M, Lee K, Goto K, et al. Rapid and effective DNA extraction method with bead grinding for a large amount of fungal DNA. *J Food Prot.* 2010;73(6):1077–1084.
- [28] Borodovsky M, Lomsadze A. Eukaryotic gene prediction using GeneMark.hmm-E and GeneMark-ES. *Curr Protocols Bioinform.* 2011;4:1–10.
- [29] Boeckmann B, Bairoch AM, Apweiler R, et al. The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Res.* 2003;31(1):365–370.
- [30] Ashburner M, Ball CA, Blake JA, et al. Gene ontology: tool for the unification of biology. *Nat Genet.* 2000;25(1):25–29.
- [31] Kanehisa M, Goto S, Kawashima S, et al. The KEGG resource for deciphering the genome. *Nucleic Acids Res.* 2004;32(90001):277D–2280.
- [32] Kanehisa M, Goto S, Hattori M, et al. From genomics to chemical genomics: new developments in KEGG. *Nucleic Acids Res.* 2006;34(90001):D354–D357.
- [33] Murat C, Riccioni C, Belfiori B, et al. Distribution and localization of microsatellites in the Perigord black truffle genome and identification of new molecular markers. *Fungal Genet Biol.* 2011;48(6): 592–601.
- [34] Untergasser A, Cutcutache I, Koressaar T, et al. Primer3-new capabilities and interfaces. *Nucleic Acids Res.* 2012;40(15):e115–e115.
- [35] Zhang H, Yohe T, Huang L, et al. dbCAN2: A meta server for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res.* 2018; 46(W1):W95–W101.
- [36] Van Ooijen J, Voorrips R. JoinMap 4.0: Software for the calculation of genetic linkage maps in experimental populations. 2006; Kyazma BV, Wageningen.
- [37] Floudas D, Binder D, Riley R, et al. The paleozoic origin of enzymatic lignin decomposition reconstructed from 31 fungal genomes. *Science.* 2012; 336(6089):1715–1719.
- [38] Fang M, Wang XE, Chen Y, et al. Genome sequence analysis of *Auricularia heimuer* combined with genetic linkage map. *J Fungi.* 2020;6(1):37.
- [39] Liu DB, Gong J, Dai WK, et al. The genome of *Ganoderma lucidum* provide insights into triterpene biosynthesis and wood degradation. *PLOS One.* 2012;7(5):e36146.
- [40] Chen BZ, Gui F, Xie BG, et al. Composition and expression of genes encoding carbohydrate-active enzymes in the straw-degrading mushroom *Volvariella volvacea*. *PLoS One.* 2013;8(3):e58780.
- [41] Qian J, Xu HB, Song JY, et al. Genome-wide analysis of simple sequence repeats in the model medicinal mushroom *Ganoderma lucidum*. *Gene.* 2013;512(2):331–336.
- [42] Wang Y, Chen MJ, Wang H, et al. Microsatellites in the genome of the edible mushroom, *Volvariella volvacea*. *BioMed Res Int.* 2014;2014:28912.
- [43] Qu JB, Huang CY, Zhang JX. Genome-wide functional analysis of SSR for an edible mushroom *Pleurotus ostreatus*. *Gene.* 2016;575(2 Pt 2): 524–530.
- [44] Lombard V, Golaconda Ramulu H, Drula E, et al. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* 2014;42:D490–D495.
- [45] Riley R, Salamov AA, Brown DW, et al. Extensive sampling of basidiomycete genomes demonstrates inadequacy of the white-rot/brown-rot paradigm for wood decay fungi. *Proc Natl Acad Sci USA.* 2014;111(27):9923–9928.
- [46] James TY, Srivilai P, Kües U, Vilgalys, et al. Evolution of the bipolar mating system of the mushroom *Coprinellus disseminatus* from its tetrapolar ancestors involves loss of mating-type-specific pheromone receptor function. *Genetics.* 2006; 172(3):1877–1891.
- [47] Yi R, Tachikawa T, Mukaiyama H, Mochida, et al. DNA-mediated transformation system in a bipolar basidiomycete, *Pholiota microspora* (*P. nameko*). *Mycoscience.* 2009;50(2):123–129.
- [48] James TY, Lee M, van Diepen LTA. A Single Mating-Type Locus Composed of Homeodomain genes promotes nuclear migration and heterokaryosis in the white-rot fungus *Phanerochaete chrysosporium*. *Eukaryot Cell.* 2011;10(2):249–261.
- [49] Van Peer AF, Park SY, Shin PG, et al. Comparative genomics of the mating-type loci of the mushroom *Flammulina velutipes* reveals widespread synteny and recent inversions. *PLoS One.* 2011;6(7):e22249.

- [50] Bakkeren G, Kronstad JW. Linkage of mating-type loci distinguishes bipolar from tetrapolar mating in basidiomycetous smut fungi. *Proc Natl Acad Sci USA*. 1994;91(15):7085–7089.
- [51] Au CH, Wong MC, Bao DP, et al. The genetic structure of the A mating-type locus of *Lentinula edodes*. *Gene*. 2014;535(2):184–190.
- [52] Casselton LA, Asante-Owusu RN, Banham AH, et al. Mating type control of sexual development in *Coprinus cinereus*. *Can J Bot*. 1995;73(S1):266–272.
- [53] Specht CA, Stankis MM, Giasson L, et al. Functional analysis of the homeodomain related proteins of the A alpha locus of *Schizophyllum commune*. *Proc Natl Acad Sci USA*. 1992;89(15):7174–7178.
- [54] Liu HY, Cui JT, Gao YM. Progress of segregation distortion. *J Plant Genet Res*. 2009;10:613–617.
- [55] Liu ZF, Zhu ZP, Huang DR, et al. Genetic analysis on segregation distortion of molecular markers in F2 population of *C. Moschata* Duch. *Mol Plant Breed*. 2019;17:3993–3999.