**BMC Genomics**

# De novo sequencing, assembly and functional annotation of *Armillaria borealis* genome

Vasilina S. Akulova[1,2], Vadim V. Sharov[1,2,3], Anastasiya I. Aksyonova[1], Yuliya A. Putintseva[1], Natalya V. Oreshkova[1,2,4], Sergey I. Feranchuk[1,5,6], Dmitry A. Kuzmin[1,3], Igor N. Pavlov[1,7,8], Yulia A. Litovka[7,8] and Konstantin V. Krutovsky[1,9,10,11,12]*

## Abstract

**Background:** Massive forest decline has been observed almost everywhere as a result of negative anthropogenic and climatic effects, which can interact with pests, fungi and other phytopathogens and aggravate their effects. Climatic changes can weaken trees and make fungi, such as *Armillaria* more destructive. *Armillaria borealis* (Marxm. & Korhonen) is a fungus from the Physalacriaceae family (*Basidiomycota*) widely distributed in Eurasia, including Siberia and the Far East. Species from this genus cause the root white rot disease that weakens and often kills woody plants. However, little is known about ecological behavior and genetics of *A. borealis*. According to field research data, *A. borealis* is less pathogenic than *A. ostoyae*, and its aggressive behavior is quite rare. Mainly *A. borealis* behaves as a secondary pathogen killing trees already weakened by other factors. However, changing environment might cause unpredictable effects in fungus behavior.

**Results:** The de novo genome assembly and annotation were performed for the *A. borealis* species for the first time and presented in this study. The *A. borealis* genome assembly contained ~ 68 Mbp and was comparable with ~ 60 and ~ 79.5 Mbp for the *A. ostoyae* and *A. mellea* genomes, respectively. The N50 for contigs equaled 50,544 bp. Functional annotation analysis revealed 21,969 protein coding genes and provided data for further comparative analysis. Repetitive sequences were also identified. The main focus for further study and comparative analysis will be on the enzymes and regulatory factors associated with pathogenicity.

(Continued on next page)

* Correspondence: kkrutov@gwdg.de
[1]Laboratory of Forest Genomics, Genome Research and Education Center, Institute of Fundamental Biology and Biotechnology, Siberian Federal University, 660036 Krasnoyarsk, Russia
[9]Department of Forest Genetics and Forest Tree Breeding, Georg-August University of Göttingen, 37077 Göttingen, Germany
Full list of author information is available at the end of the article

Akulova *et al. BMC Genomics* 2020, **21**(Suppl 7):534

Page 2 of 8

(Continued from previous page)

**Conclusions:** Pathogenic fungi such as *Armillaria* are currently one of the main problems in forest conservation. A comprehensive study of these species and their pathogenicity is of great importance and needs good genomic resources. The assembled genome of *A. borealis* presented in this study is of sufficiently good quality for further detailed comparative study on the composition of enzymes in other *Armillaria* species. There is also a fundamental problem with the identification and classification of species of the *Armillaria* genus, where the study of repetitive sequences in the genomes of basidiomycetes and their comparative analysis will help us identify more accurately taxonomy of these species and reveal their evolutionary relationships.

**Keywords:** Annotation, *Armillaria*, De novo sequencing, Fungal forest pathogen, Genome

## Background

Massive forest decline as a result of negative anthropogenic and climatic effects, often aggravated by pests, fungi, and other phytopathogens, has been observed almost everywhere. Environmental changes, such as increased average annual temperatures, decreased precipitation, more frequent droughts, can weaken trees and make fungi much more destructive. Forest conservation has become a serious issue since the scale of plant death caused by phytopathogenic fungi is enormous. For instance, tree diseases have caused the loss of approximately 100 million elm trees in the United Kingdom and the United States, and the list can be continued. Among all phytopathogens, fungi cause 64% of infection-related species extinction and regional extirpation events [1].

The basidiomycete genus *Armillaria* plays a very important role in forest ecosystems worldwide and currently includes more than 40 officially described species [2, 3]. *Armillaria* species differ significantly in virulence, for example, some species, such as *A. ostoyae*, are the main cause of tree death while other species colonize plants already damaged by various factors (drought, pests, etc.) [4, 5]. Difference in pathogenicity has also been observed in *A. ostoyae*, however virulence variation of *A. borealis* has not been studied yet [6].

*Armillaria borealis* (Marxm. & Korhonen) is a fungus from the *Physalacriaceae* family (*Basidiomycota*) widely distributed in Eurasia, including Siberia and the Far East [1]. Species from this genus cause the root white rot disease that weakens and often kills woody plants [7]. Several phylogenetic and genomic studies on *A. ostoyae* have been carried out due to its high pathogenic potential and common occurrence [4, 8, 9], while little is known about ecological behavior of *A. borealis*. According to field research data, *A. borealis* is less pathogenic than *A. ostoyae*, and its aggressive behavior is rare. Mainly *A. borealis* behaves as a secondary pathogen killing trees already weakened by biotic and abiotic factors [10–13]. However, changing environment might cause unpredictable effects in fungi behavior.

*Armillaria spp.* impact on forest populations has both economic and ecological significance. They attack hundreds of different tree species (e.g., *Abies, Picea, Pinus, Betula, Sorbus, Juglans, Malus*, etc.) in both hemispheres under different climatic conditions, and are among the most destructive forest pathogens [2, 14, 15].

Identification of species and pathogenicity levels of *Armillaria* is crucial for forest conservation. The genomic data are needed to study the pathogenicity of pathogenic species and to better understand their impact on trees and the host-pathogen interactions. In addition, comparative genomics can help to resolve complex phylogeny of *Armillaria* species. It is worth noting that fungi genomic data are also important for industrial applications. For example, white rot *Armillaria* fungi are capable of lignin and cellulose decomposition, and they can be used to utilize the wood and paper production waste [16].

*A. borealis* is very important for the vast boreal forest ecosystems. However, despite the enormous influence of *Armillaria* species on forestry, horticulture, and agriculture, fungi of this genus and their pathogenicity are still not well-studied in this large region, which makes the presented genomic study very much needed.

There are already published genomic and proteomic data for *A. mellea*, *A. solidipes*, and *A. ostoyae* revealing the presence of plant cell wall degradation enzymes (PCWDE) and some secreted proteins [17–19]. Genomic analysis of other pathogenic basidiomycetes, such as *Moniliophthora* [20, 21], *Heterobasidion* [22], and *Rhizoctonia* [23], also revealed genes encoding PCWDE, as well as secreted enzymes and secondary metabolism effector proteins as putative pathogenicity factors. However, the life cycle and the distribution strategy of *Armillaria* members indicate that they may have evolved other additional mechanisms for pathogenicity, which along with other potential genomic mechanisms are not yet studied [24]. It is worth noting that the role and functional significance of mobile and highly repetitive elements (REs) are still not completely clear. Gradually accumulated data suggest that REs can play an important role in the evolutionary development of organisms, replication, and formation of nucleoprotein complexes, as well as affect gene expression [17]. Genomes of fungi are densely packed containing effector genes and transposable elements (TEs) [25–27]. It

Akulova *et al. BMC Genomics* 2020, **21**(Suppl 7):534

Page 3 of 8

was reported that different fungal pathogens, such as *Fusarium* [28] and *Verticillium* [29], have similar genome architecture. So, it is expected that TEs may play important roles in host switching and adaptation to new ecological niches [30]. It was found in *Magnaporthe oryzae* that genes involved in host specialization were associated with TEs [31].

## Results

### Genome assembly

The *A. borealis* genome assembly contained ~ 67 Mbp and was comparable with ~ 60 and ~ 79.5 Mbp for the *A. ostoyae* and *A. mellea* genomes, respectively. The N50 for contigs equaled 50,554 bp (Table 1).

### Completeness of the genome assembly

The BUSCO v3.1.0 software [32] was used to evaluate the completeness of the genome assembly. It showed that 94.7% of reference genes were captured as complete single-copy genes (Table 2). In addition, the RNA sequence reads were mapped to the genome assembly by TopHat v. 2.1.0 [33].

### Content of the RNA-seq reads

The SortMeRNA v2.1 program [34] with the pre-installed eight rRNA databases (silva-bac-16 s-id90, silva-arc-16 s-id95, silva-euk-18 s-id95, silva-bac-23 s-id98, silva-arc-23 s-id98, silva-euk-28 s-id98, rfam-5 s-id98, and rfam-5.8 s-id98) from the SILVA rRNA database project (SILVA SSU and LSU Ref NR v.119; https://www.arb-silva.de) was used to check the content of the RNA sequences. The results are presented in Table 3. Both RNA samples contained high number of rRNA reads – 32.5 and 72.9%, respectively.

### Genome gene annotation

Functional annotation revealed 21,969 protein coding genes, which was also comparable with 22,705 and 14,473 genes in *A. ostoyae* and *A. mellea*, respectively. Their gene ontology (GO) functional annotation is presented in Fig. 1 and Additional files 1, 2 and 3. The greatest number of annotated sequences was related to the functioning of the cell nucleus (Fig. 1).

The distribution of enzyme genes across main classes is represented in Fig. 2. Oxidoreductases and hydrolases were among the most abundant enzymes.

**Table 1** Assembly parameters of the *Armillaria borealis* genome

| Parameter | Contig |
| --- | --- |
| Number | 44,412 |
| Total length, bp | 66,792,428 |
| Maximum length, bp | 2,136,877 |
| N50, bp | 50,544 |
| N90, bp | 346 |

**Table 2** Results of the *Armillaria borealis* genome assembly assessment using BUSCO

| Genes | Number | Percentage |
| --- | --- | --- |
| Complete (single-copy) | 1286 | 96.3 (94.7) |
| Fragmented | 17 | 1.3 |
| Missing | 32 | 2.4 |
| Total number | 1335 | 100 |

### Repetitive element (RE) annotation

In total, 886 RE sequences were identified in the *A. borealis* genome assembly. However, 839 (94.7%) of them remained unrecognized based on the initial classification using RepeatModeler program (Table 4).

The following RE types were identified: LTR retrotransposons (Copia, Gypsy, Pao), LINE-retrotransposons (Tad1), and retrotransposons with a tyrosine recombinase (Ngaro).

The TEclass classifier software allowed us to further partition sequences including initially unclassified into four main groups (Table 5). The additional comparative studies of repetitive sequences in the genomes of basidiomycetes are needed to further classify REs. The variety of the detected TE families was not very wide, excep the LTR TE family, which was widely represented in *A. borealis*. The Ty3/Gypsy and Ty1/Copia elements have been identified, and Gypsy was the most abundant among them.

## Discussion

Typical fungal nuclear genome sizes occupy an intermediate position between prokaryotes and other eukaryotes. On average, the size of the fungal genome is two orders of magnitude smaller than that of higher plants and varies from ~ 2.19 to ~ 3706 Mbp ( [35], see also DOE JGI Fungi Portal: https://mycocosm.jgi.doe.gov/mycocosm/home). The average genome size of *Ascomycota* and *Basidiomycota* divisions is ~ 36.91 and ~ 46.48 Mb, respectively [36]. The genome size of *A. borealis* (~ 66.79 Mbp) was within a range of genome sizes in closely related species. For example, the genomes of *A. ostoyae* and *A. mellea* were ~ 60 and ~ 79.5 Mbp, respectively [19, 37].

The genome sizes of our assembly and draft assembly of *Armillaria borealis* FPL87.14 v1.0 available in the US DoE JGI fungal genomics resource database (https://mycocosm.jgi.doe.gov/Armbor1) are quite comparable,

**Table 3** Content of the RNA-seq reads

| Parameter | Sample 1 | Sample 2 |
| --- | --- | --- |
| Total number of reads (%) | 6,428,260 (100%) | 8,444,300 (100%) |
| Number of non-rRNA reads (%) | 4,339,634 (67.5%) | 2,288,932 (27.1%) |
| Number of rRNA reads (%) | 2,088,626 (32.5%) | 6,155,368 (72.9%) |
| Average read length, bp | 153 | 151 |

Akulova *et al. BMC Genomics* 2020, **21**(Suppl 7):534
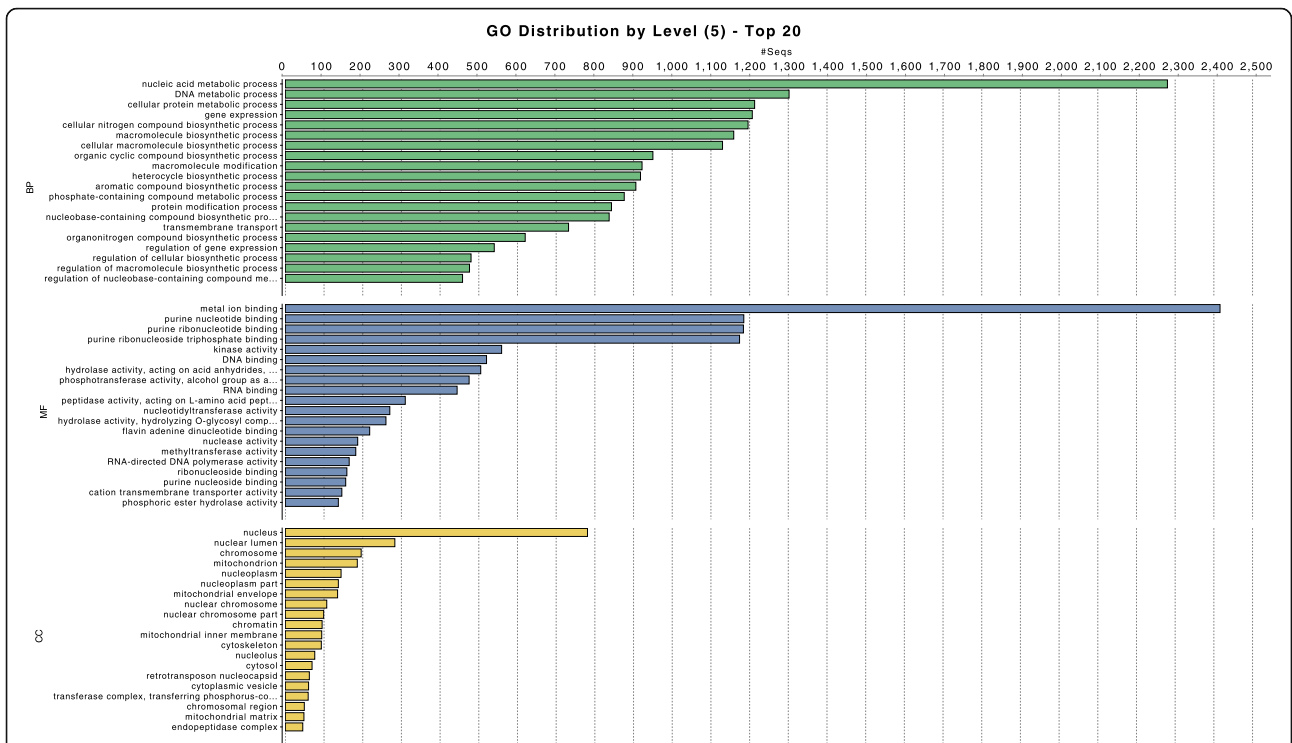
Page 4 of 8



**Fig. 1** Distribution of 21,969 protein coding genes found in the *Armillaria borealis* genome assembly at the three levels - molecular function (MF), biological processes (BP), and cellular components (CC), respectively, based on the GO functional annotation
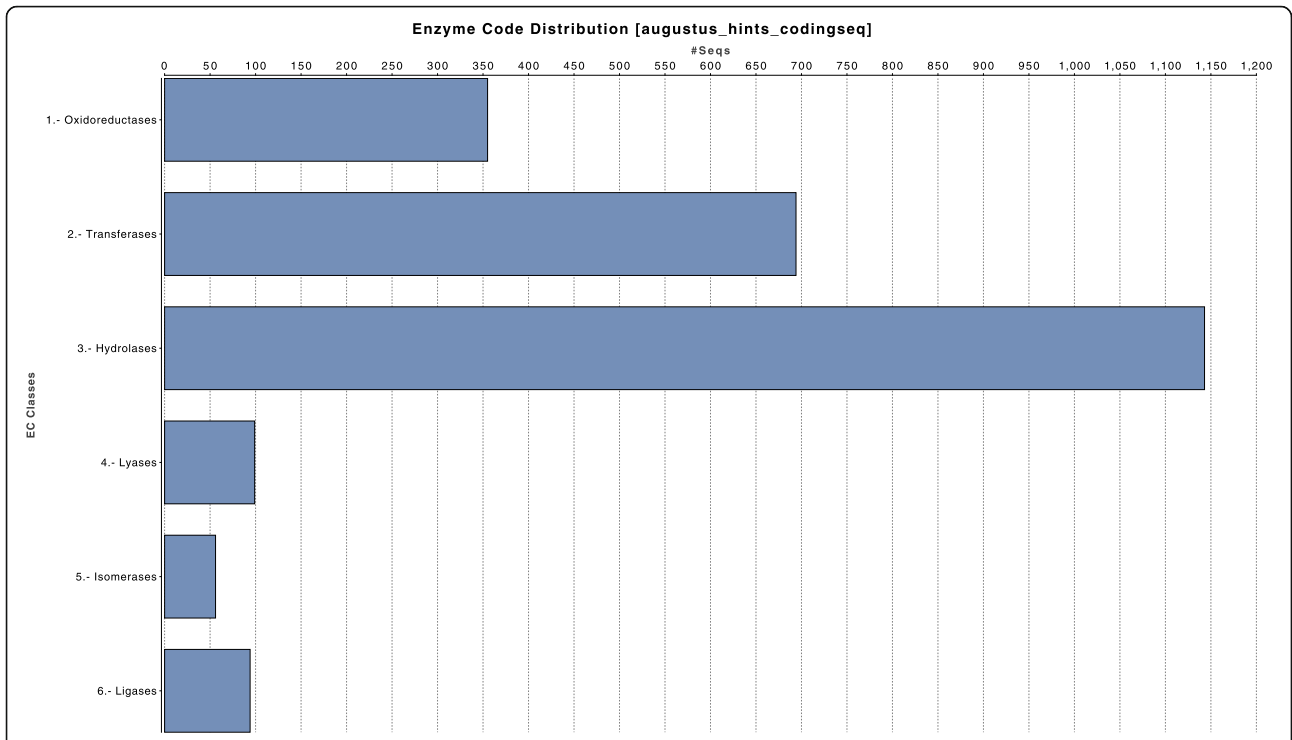


**Fig. 2** Distribution of enzyme genes found in the *Armillaria borealis* genome assembly across main classes

Akulova *et al. BMC Genomics* 2020, **21**(Suppl 7):534

Page 5 of 8

**Table 4** Initial classification of the repetitive elements (REs) identified in the *A. borealis* genome assembly

| RE type | RE family | Number |
|---|---|---|
| LTR retrotransposons | Copia | 5 |
| | Gypsy | 32 |
| | Pao | 1 |
| LINE-retrotransposons | Tad1 | 5 |
| Retrotransposons with a tyrosine recombinase (YR) | Ngaro | 3 |
| Simple sequence repeats | Simple Repeats | 1 |
| **Classified** | | **47 (5.3%)** |
| **Unclassified** | | **839 (94.7%)** |
| **Total** | | **886** |

~ 66.8 vs. ~ 71.69 Mbp, respectively. The difference can be explained by the use of different sequencing technologies (Illumina MiSeq vs. PacBio) that resulted in significant differences in the number of scaffolds between two assemblies but did not influence much the number of coding sequences identified, 21,969 coding sequences in our assembly vs. 19,984 in *Armillaria borealis* FPL87.14 v1.0. Meanwhile, the genome size of *A. gallica* [19] is almost 19 Mbp bigger than in *A. borealis*. The number of genes identified in the *A. gallica* genome was also larger (~ 25,000 genes).

Enzymatic activities of plant cell wall degrading fungi are performed by complex mixtures of cellulases, hemicellulases, and ligninases [38]. Oxidoreductases and hydrolases are the most interesting among enzymes in *Armillaria* because they are involved in lignin (oxidoreductases) and cellulose (hydrolases) decomposition.

Among identified TE classes LTR elements were the most frequent, particularly in basidiomycete fungi [39]. It is also true for *A. borealis*. The effect of TE on lifestyle of white-rot fungi has not been studied in depth yet, but some studies revealed considerable difference in TE number among different *Armillaria* species [19]. Their detailed comparative analysis will be presented in a separate paper.

**Table 5** Additional classification based on TEclass including initially unclassified repetitive elements (REs) identified in the *A. borealis* genome assembly

| RE type | Number | % |
|---|---|---|
| DNA-transposons | 79 | 9 |
| Retrotransposons: | | |
| LTR | 735 | 83 |
| LINE | 27 | 3 |
| Unclear | 18 | 2 |
| Total | 886 | 100 |

## Conclusions

The destruction of forests by pathogenic fungi is one of the main problems of forest conservation. Further comprehensive studies of these fungi at genomic, transcriptomic, proteomic, and metabolomic levels are very much needed to identify causes and mechanisms of their increasing pathogenicity. Our study provides important genomic resources of sufficiently good quality for further detailed work on studying the genetics of pathogenicity of *Armillaria* and other fungi species. It should also help with in-depth evolutionary and phylogenomic analyses and better identification and classification of *Armillaria* species genus.

## Methods

### Sample collection and DNA sequencing

The active mycelia of *A. borealis* were collected from dead trees of *Abies sibirica* in 2015 in a mixed forest consisted mainly of Siberian fir, silver birch, Norway spruce, and Siberian stone pine and located 40 km to the northwest of Krasnoyarsk City, Russia (56.175847°N, 92.184933°E). The fresh mycelium was isolated from under the bark of the infested stems 50 cm above the soil surface using sterile tweezers and gloves to avoid contamination. Before DNA extraction, mycelium was fixed for 2 days at 4°C in RNAlater (Thermo Fisher Scientific Company, Waltham, Massachusetts, USA). Then RNAlater fixed mycelium was quickly ground in acid-washed and autoclaved mortar. DNA was isolated using a modified version of the hot-CTAB extraction at 65 °C [40], followed by chloroform (double washing). Total DNA was precipitated for an hour with isopropanol at 4 °C, centrifuged at 6500 g for 30 min at 4 °C, washed twice with 70% ethanol, and was eluted in 50 μl nuclease-free water. Integrity and amount of the isolated total DNA were examined by 1.5% (wt/vol) agarose gel electrophoresis, and using the NanoDrop 1000 Spectrophotometer (Thermo Fisher Scientific Company, Waltham, Massachusetts, USA). The amount of extracted DNA was also measured on the Invitrogen Qubit 4 Fluorometer (Thermo Fisher Scientific Company, Waltham, Massachusetts, USA).

The paired-end sequencing libraries with 500 bp long genomic DNA inserts were generated using Truseq DNA Sample Prep Kit according to the manufacturer's instructions (Illumina, Inc., San Diego, CA, USA). MiSeq Reagent Kit v2 (500-cycles) was used to sequence on the Illumina MiSeq platform with $2 \times 250$ cycles at the Laboratory of Forest Genomics of Siberian Federal University (Genome Research and Education Center, Siberian Federal University, Krasnoyarsk, Russia).

### RNA isolation and sequencing

Two samples were isolated from the active *A. borealis* mycelium from two dead trees of *Abies sibirica*, respectively, in 2015 and fixed for 2 days at 4°C in RNAlater

Akulova *et al. BMC Genomics* 2020, **21**(Suppl 7):534

Page 6 of 8

(Thermo Fisher Scientific Company, Waltham, Massachusetts, USA). The distance between the trees was 2–10 m. The RNA was isolated using Qiagen RNeasy Mini Kit (Qiagen, Valencia, CA, USA).

The quality and concentration of the RNA were measured on Agilent 2100 Bioanalyzer using Agilent RNA 6000 Nano kit (Agilent Technologies, Inc., Santa Clara, CA, USA). Purified RNA with high quality was selected for further cDNA library construction. Purification of mRNA from total RNA was performed using Oligo (dT) magnetic beads. The mRNA treated with fragmentation buffer was used as a template for cDNA synthesis. A double-stranded cDNA library was constructed with the TruSeq RNA Library Prep Kit v2 (Illumina, Inc., San Diego, CA, USA). End-repair, A-tailing, adapter ligation, and library amplification were performed during cDNA library construction followed by cluster generation and sequencing on the Illumina MiSeq platform in the Laboratory of Forest Genomics (Genome Research and Education Center, Siberian Federal University, Krasnoyarsk, Russia) using MiSeq Reagent Kit v2 (2 × 150 cycles).

### Genome assembly
The de novo genome assembly was performed using SPAdes 3.13.0 genome assembler (http://cab.spbu.ru/software/spades) on high-performance computing (HPC) system IBM × 3950 X6 with 96 CPU and 3 TB RAM using an iterative genome assembly module for short reads. *K*-mer values were automatically selected based on read length and data type [41].

### Gene annotation
The gene annotation of the *A. borealis* genome was performed using BRAKER2 [42], which is a combination of GeneMark-ET [43] and AUGUSTUS [44], that uses genomic and RNA-Seq data to automatically generate full gene structure annotations in novel genomes. AUGUSTUS integrates the extrinsic evidence from protein homology information into the prediction. There were no protein data for *A. borealis* before this study, therefore protein sequences of a close relative *A. ostoyae* have been used [19].

The gene ontology (GO) functional annotation was carried out using Blast2GO [45]. This program worked directly with coding sequences in fasta format. First, the nucleotide sequences homologues to the *A. borealis* sequences were searched for in the BLAST database. Then, sequence mapping and annotation were carried out. In parallel, protein domains were detected using InterProScan [46].

### Repetitive element (RE) annotation
Repetitive sequences were identified initially using the RepeatModeler program designed to perform de novo search [47]. The additional classification of RE sequences was done using TEclass program [48].

## Supplementary information
**Supplementary information** accompanies this paper at https://doi.org/10.1186/s12864-020-06964-6.

> **Additional file 1: Figure S1.** GO distribution of coding sequences found in the *Armillaria borealis* genome assembly at the biological processes (BP) level based on the GO functional annotation.
>
> **Additional file 2: Figure S2.** GO distribution of coding sequences found in the *Armillaria borealis* genome assembly at the molecular function (MF) level based on the GO functional annotation.
>
> **Additional file 3: Figure S3.** GO distribution of coding sequences found in the *Armillaria borealis* genome assembly at the cellular components (CC) level based on the GO functional annotation.

### Abbreviations
BLAST: Basic Local Alignment Search Tool; bp: Base pair; CPU: Central processing unit; DNA: Deoxyribonucleic acid; LINE: Long interspersed nuclear element; LTR: Long terminal repeat; Mbp: Million base pair; N50: A weighted median statistic such that 50% of the entire assembly is contained in contigs or scaffolds equal to or larger than this value; N90: A weighted median statistic such that 90% of the entire assembly is contained in contigs or scaffolds equal to or larger than this value; PCWDE: Plant cell wall degradation enzymes; RAM: Random access memory; RE: Repetitive element; RNA: Ribonucleic acid; rRNA: ribosomal RNA; SINE: Short interspersed nuclear element; TE: Transposable element

### Availability of data and materials
The *A. borealis* genome assembly and all sequences described in this study are available in GenBank under the accession number JAAGUC000000000, sample SAMN13920850, project PRJNA603128 (https://www.ncbi.nlm.nih.gov/bioproject/603128).

### Ethics approval and consent to participate
Not applicable.

Akulova *et al. BMC Genomics* 2020, **21**(Suppl 7):534

Page 7 of 8

## Author details
[1]Laboratory of Forest Genomics, Genome Research and Education Center, Institute of Fundamental Biology and Biotechnology, Siberian Federal University, 660036 Krasnoyarsk, Russia. [2]Laboratory of Genomic Research and Biotechnology, Federal Research Center "Krasnoyarsk Science Center of the Siberian Branch of the Russian Academy of Sciences", 660036 Krasnoyarsk, Russia. [3]Department of High Performance Computing, Institute of Space and Information Technologies, Siberian Federal University, 660074 Krasnoyarsk, Russia. [4]Laboratory of Forest Genetics and Selection, V. N. Sukachev Institute of Forest, Siberian Branch of Russian Academy of Sciences, 660036 Krasnoyarsk, Russia. [5]Department of Informatics, National Research Technical University, 664074 Irkutsk, Russia. [6]Limnological Institute, Siberian Branch of Russian Academy of Sciences, 664033 Irkutsk, Russia. [7]Laboratory of Reforestation, Mycology and Plant Pathology, V. N. Sukachev Institute of Forest, Siberian Branch of Russian Academy of Sciences, 660036 Krasnoyarsk, Russia. [8]Department of Chemical Technology of Wood and Biotechnology, Reshetnev Siberian State University of Science and Technology, Krasnoyarsk 660049, Russia. [9]Department of Forest Genetics and Forest Tree Breeding, Georg-August University of Göttingen, 37077 Göttingen, Germany. [10]Center for Integrated Breeding Research, George-August University of Göttingen, 37075 Göttingen, Germany. [11]Laboratory of Population Genetics, N. I. Vavilov Institute of General Genetics, Russian Academy of Sciences, 119333 Moscow, Russia. [12]Department of Ecosystem Science and Management, Texas A&M University, College Station, TX 77843-2138, USA.

## References
1. Fisher MC, Henk DA, Briggs CJ, Brownstein JS, Madoff LC, McCra SL, Gurr SJ. Emerging fungal threats to animal, plant and ecosystem health. Nature. 2012;484:186–94.
2. Coetzee M, Wingfield BD, Wingfield MJ. Armillaria root-rot pathogens: species boundaries and global distribution. Pathogens. 2018;7(4):83.
3. Baumgartner K, Coetzee MPA, Hoffmeister D. Secrets of the subterranean pathosystem of *Armillaria*. Mol Plant Pathol. 2011;12(6):515–34.
4. Prospero S, Holdenrieder O, Rigling D. Comparison of the virulence of *Armillaria cepistipes* and *Armillaria ostoyae* on four Norway spruce provenances. For Pathol. 2004;34(1):1–14.
5. Marcais B, Breda N. Role of an opportunistic pathogen in the decline of stressed oak trees. J Ecol. 2006;94(6):1214–23.
6. Morrison DJ, Pellow KW. Variation in virulence among isolates of *Armillaria ostoyae*. For Pathol. 2002;32(2):99–107.
7. Sahu N, Merényi Z, Bálint B, Kiss B, Sipos G, Owens R, Nagy LG. Hallmarks of basidiomycete soft- and white-rot in wood-decay-omics data of *Armillaria*. bioRxiv preprint. 2020. https://doi.org/10.1101/2020.05.04.075879.
8. Legrand P, Ghahari S, Guillaumin JJ. Occurrence of genets of *Armillaria spp.* in four mountain forests in Central France: the colonization strategy of *Armillaria ostoyae*. New Phytol. 1996;133(2):321–32.
9. Omdal DW, Shaw CG III, Jacobi WR, Wager TC. Variation of pathogenicity and virulence of isolates of *Armillaria ostoyae* on eight tree species. Plant Dis. 1995;79(9):939–44.
10. Bendel M, Kienast F, Rigling D. Genetic population structure of three *Armillaria* species at the landscape scale: a case study from Swiss *Pinus mugo* forests. Mycol Res. 2006;110(6):705–12.
11. Keča N, Solheim H. Ecology and distribution of *Armillaria* species in Norway. For Pathol. 2011;41(2):120–32.
12. Marxmüller H, Holdenrieder O. *Armillaria mellea* sl in Southern Bavaria. In: Frontiers in mycology, vol. 36. Wallingford, Oxon: CAB International; 1990. p. 9–32.
13. Pavlov IN. Biotic and abiotic factors as causes of coniferous forests dieback in Siberia and Far East. Contemp Probl Ecol. 2015;8(4):440–56.
14. Cromey M, Drakulic J, Beal L, Waghorn I, Perry J, Clover GR. Susceptibility of garden trees and shrubs to *Armillaria* root rot. Plant Dis. 2020;104(2):483–92.
15. Drakulic J, Gorton C, Perez-Sierra A, Clover G, Beal L. Associations between *Armillaria* species and host plants in UK gardens. Plant Dis. 2017;101(11):1903–9.
16. Abdel-Hamid AM, Solbiati JO, Cann IKO. Insights into lignin degradation and its potential industrial applications. Adv Appl Microbiol. 2013;82:1–28.
17. Hirsch CD, Springer NM. Transposable element influences on gene expression in plants. Biochim Biophys Acta. 2017;1860(1):157–65.
18. Ross-Davis AL, Steward JE, Hanna JW, Kim M-S, Knaus BJ, Cronn R, Rai H, Richardson BA, GI MD, Klopfenstein NB. Transcriptome of an *Armillaria* root disease pathogen reveals candidate genes involved in host substrate utilization at the host-pathogen interface. For Path. 2013;43(6):468–77.
19. Sipos G, Prasanna AN, Walter MC, O'Connor E, Bálint B, Krizsán K, et al. Genome expansion and lineage-specific genetic innovations in the forest pathogenic fungi *Armillaria*. Nat Ecol Evol. 2017;1(12):1931–41.
20. Meinhardt LW, Costa GGL, Thomazella DPT, Thomazella DP, Teixeira PJP, Carazzolle MF, et al. Genome and secretome analysis of the hemibiotrophic fungal pathogen, *Moniliophthora roreri*, which causes frosty pod rot disease of cacao: mechanisms of the biotrophic and necrotrophic phases. BMC Genomics. 2014;15(1):164.
21. Mondego JMC, Carazzolle MF, Costa GG, Formighieri EF, Parizzi LP, Rincones J, Cotomacci C, Carraro DM, Cunha AF, Carrer H, Vidal RO, Estrela RC, Garcia O, Thomazella DPT, de Oliveira BV, ABL P, MCS R, MRR A, de Moraes MH, Castro LAB, Gramacho KP, Goncalves MS, Neto JPM, Neto AG, Barbosa LV, Guiltinan MJ, Bailey BA, Meinhardt LW, Cascardo JCM, Pereira GAG. A genome survey of *Moniliophthora perniciosa* gives new insights into Witches' Broom disease of cacao. BMC Genomics. 2008;9(1):548.
22. Olson Å, Aerts A, Asiegbu F, Belbahri L, Bouzid O, Broberg A, Canback B, Coutinho PM, Cullen D, Dalman K, Deflorio G, van Diepen LT, Dunand C, Duplessis S, Durling M, Gonthier P, Grimwood J, Fossdal CG, Hansson D, Henrissat B, Hietala A, Himmelstrand K, Hoffmeister D, Hogberg N, James TY, Karlsson M, Kohler A, Kues U, Lee YH, Lin YC, et al. Insight into trade-off between wood decay and parasitism from the genome of a fungal forest pathogen. New Phytol. 2012;194(4):1001–13.
23. Hane JK, Anderson JP, Williams AH, Sperschneider J, Singh KB. Genome sequencing and comparative genomics of the broad host-range pathogen *Rhizoctonia solani* AG8. PLoS Genet. 2014;10(5):e1004281.
24. Raffaele S, Kamoun S. Genome evolution in filamentous plant pathogens: why bigger can be better. Nat Rev Microbiol. 2012;10(6):417–30.
25. Muszewska A, Steczkiewicz K, Stepniewska-Dziubinska M, Ginalski K. Transposable elements contribute to fungal genes and impact fungal lifestyle. Sci Report. 2019;9(1):4307.
26. Moller M, Stukenbrock EH. Evolution and genome architecture in fungal plant pathogens. Nat Rev Microbiol. 2017;15:756–71.
27. Dong S, Raffaele S, Kamoun S. The two-speed genomes of filamentous pathogens: waltz with plants. Curr Opin Genet Dev. 2015;35:57–65.
28. Sperschneider J, et al. Genome-wide analysis in three *Fusarium* pathogens identifies rapidly evolving chromosomes and genes associated with pathogenicity. Genome Biol Evol. 2015;7:1613–27.
29. Faino L, Seidl MF, Shi-Kunne X, Pauper M, van den Berg GC, Wittenberg AH, Thomma BP. Transposons passively and actively contribute to evolution of the two-speed genome of a fungal pathogen. Genome Res. 2018;26(8):1091–100.
30. Ma L-J, et al. Comparative genomics reveals mobile pathogenicity chromosomes in *Fusarium*. Nature. 2010;464:367–73.
31. Yoshida K, Saunders DG, Mitsuoka C, Natsume S, Kosugi S, Saitoh H, et al. Host specialization of the blast fungus *Magnaporthe oryzae* is associated with dynamic gain and loss of genes linked to transposable elements. BMC Genomics. 2016;17(1):370.
32. Simao FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. 2015;31(19):3210–2.
33. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. Bioinformatics. 2009;25(9):1105–11.
34. Kopylova E, Noé L, Touzet H. SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. Bioinformatics. 2012;28(24):3211–7.
35. Talhinhas P, Tavares D, Ramos AP, Gonçalves S, Loureiro J. Validation of standards suitable for genome size estimation of fungi. J Microbiol Methods. 2017;142:76–8.
36. Mohanta TK, Bae H. The diversity of fungal genome. Biol Proced Online. 2015;17:8.

Akulova *et al. BMC Genomics* 2020, **21**(Suppl 7):534

Page 8 of 8

37. Collins C, Keane TM, Turner DJ, O'Keeffe G, Fitzpatrick DA, Doyle S. Genomic and proteomic dissection of the ubiquitous plant pathogen, *Armillaria mellea*: toward a new infection model system. J Proteome Res. 2013;12(6): 2552–70.

38. Adejumo TO, Coker ME, Ogundeji JS, Adejoro DO. Qualitative determination of lignocellulolytic enzymes in eight wood-decomposing fungi. J Nat Sci Res. 2015;5(14):1–8.

39. Castanera R, Borgognone A, Pisabarro AG, Ramírez L. Biology, dynamics, and applications of transposable elements in basidiomycete fungi. Appl Microbiol Biotechnol. 2017;101(4):1337–50.

40. Devey ME, Bell JC, Smith DN, Neale DB, Moran GF. A genetic linkage map for *Pinus radiata* based on RFLP, RAPD, and microsatellite markers. Theor Appl Genet. 1996;92(6):673–9.

41. Bankevich A, Nurk S, Antipov D, Gurevich A, Dvorkin M, Kulikov AS, Lesin V, Nikolenko S, Pham S, Prjibelski A, Pyshkin A, Sirotkin A, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 2012;19(5):455–77.

42. Hoff KJ, Lange S, Lomsadze A, Borodovsky M, Stanke M. BRAKER1: unsupervised RNA-seq-based genome annotation with GeneMark-ET and AUGUSTUS. Bioinformatics. 2015;32(5):767–9.

43. Lomsadze A, Burns PD, Borodovsky M. Integration of mapped RNA-seq reads into automatic training of eukaryotic gene finding algorithm. Nucleic Acids Res. 2014;42(15):e119.

44. Stanke M, Keller O, Gunduz I, Hayes A, Waack S, Morgenstern B. AUGUSTUS: *ab initio* prediction of alternative transcripts. Nucleic Acids Res. 2006; 34(Suppl. 2):W435–9.

45. Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics. 2005;21(18):3674–6.

46. Quevillon E, Silventoinen V, Pillai S, Harte N, Mulder N, Apweiler R, Lopez R. InterProScan: protein domains identifier. Nucleic Acids Res. 2005;33(Suppl. 2):W116–20.

47. . Smit A, Hubley R. RepeatModeler-1.0.11. Institute for Systems Biology. http://www.repeatmasker.org.

48. Abrusán G. TEclass—a tool for automated classification of unknown eukaryotic transposable elements. Bioinformatics. 2009;25(10):1329–30.

## Publisher's Note