Genome Size Versus Genome Assemblies: Are the Genomes Truly Expanded in Polyploid Fungal Symbionts?

Pepijn W. Kooij and Jaume Pellicer

1Department of Comparative Plant and Fungal Biology, Royal Botanic Gardens, Kew, Richmond, United Kingdom
2Center for the Study of Social Insects, São Paulo State University (UNESP), Rio Claro, São Paulo, Brazil
3Institut Botànic de Barcelona (IBB, CSIC-Ajuntament de Barcelona), Barcelona, Spain

*Corresponding author: E-mail: pepijn.kooij@gmail.com.
Accepted: 10 October 2020

Abstract

Each day, as the amount of genomic data and bioinformatics resources grows, researchers are increasingly challenged with selecting the most appropriate approach to analyze their data. In addition, the opportunity to undertake comparative genomic analyses is growing rapidly. This is especially true for fungi due to their small genome sizes (i.e., mean 1C = 44.2 Mb). Given these opportunities and aiming to gain novel insights into the evolution of mutualisms, we focus on comparing the quality of whole genome assemblies for fungus-growing ants cultivars (Hymenoptera: Formicidae: Attini) and a free-living relative. Our analyses reveal that currently available methodologies and pipelines for analyzing whole-genome sequence data need refining. By using different genome assemblers, we show that the genome assembly size depends on what software is used. This, in turn, impacts gene number predictions, with higher gene numbers correlating positively with genome assembly size. Furthermore, the majority of fungal genome size data currently available are based on estimates derived from whole-genome assemblies generated from short-read genome data, rather than from the more accurate technique of flow cytometry. Here, we estimated the haploid genome sizes of three ant fungal symbionts by flow cytometry using the fungus Pleurotus ostreatus (Jacq.) P. Kumm. (1871) as a calibration standard. We found that published genome sizes based on genome assemblies are 2.5- to 3-fold larger than our estimates based on flow cytometry. We, therefore, recommend that flow cytometry is used to precalibrate genome assembly pipelines, to avoid incorrect estimates of genome sizes and ensure robust assemblies.

Key words: evolution, fungi, fungus-growing ants, genome assembly, genome size, mutualism.

Significance

Genome sequencing and analyses are increasing daily due to decreasing costs; however, analyzing the data can be difficult at times due to a large availability of software potentially leading to erroneous genome assemblies. In our study, we show that different software can lead to different conclusion for the same genome data, that is, when the genome assembly is longer the number of genes one can predict from that assembly increases as well. We show that by accurately measuring the genome size using flow cytometry, the resulting data can help as a quality control for the genome assemblies.

Results and Discussion

Mutualistic symbioses, in which both partners benefit from living with each other, are an important driver for the evolution of biodiversity (Thompson 2005). To further understand the mechanisms underpinning this biodiversity as well as the evolution of mutualisms themselves, it is important to compare mutualistic species with their nonmutualistic relatives. Thanks to the decreasing costs of whole-genome sequencing (Wetterstrand 2019) it has become increasingly easy to tackle
these kinds of comparisons at the genomic level (Nygaard et al. 2016; Knapp et al. 2018). For example, in a recent study using comparative genomics it was shown that the genomes of ericoid mycorrhizal fungi, that form a mutualistic relationship with plants from the family Ericaceae, are more similar to saprotrophic fungi than other mutualistic ectomycorrhizal fungi (Martino et al. 2018). In another study, comparing rates of sequence divergence in the genomes of ants that form mutualistic relationships with plants with those of their non-mutualistic relatives, it was shown that rates of molecular evolution were higher in mutualistic ants, suggesting similar selective pressures to those of parasites (Rubin and Moreau 2016).

Although the decreasing sequencing costs have enabled more researchers to generate genomic data (Stephens et al. 2015), it has also led to the development of an increasing number of bioinformatic tools for analyzing the data (Clément et al. 2018). Researchers are thus faced with the challenge of selecting the most appropriate bioinformatic pipelines to ensure accurate genome assemblies with sufficient sequencing depth to accurately capture the full genomic diversity (Sims et al. 2014). As a result, it has become a relatively common practice to use the default parameters (e.g., k-mer size, PHRED quality offset, bubble algorithms, and overlap sizes) in these bioinformatic resources (Yang et al. 2011; Lai et al. 2014; Larriba et al. 2014; Liu et al. 2014; Wang et al. 2014; Quandt et al. 2015; Dentinger et al. 2016; de Man et al. 2016; Nygaard et al. 2016; Teixeira et al. 2017; Sun et al. 2018; Pucker 2019), rather than checking whether the fine-tuning of these parameters is necessary to increase the accuracy of the genome assembly.

Research into the role of the fungal symbiont of fungus-growing ants has a long history and has become a model system for studying mutualisms (Wheeler 1907; Weber 1979; Mueller et al. 2005). Briefly, the fungus provides the ants with a stable food source whereas the ants in return provide the fungus with active dispersal vectors in the form of flying queens, protection and grooming, and suitable growth substrate (Mueller et al. 2005). Generally, this type of mutualism can be divided into several “agricultural” systems; 1) Basal agriculture, the oldest group with small ant-fungal colonies of ~100 workers, and comprising fungal species/cultivars mostly from the genus Leucoagaricus Pat. (1888) that are typically dikaryotic (i.e., each cell is functionally diploid as it contains two different haploid nuclei); 2) Domesticated agriculture, with colony sizes typically ranging from a few hundred to a few thousand workers and containing heterokaryotic fungi (i.e., each cell is functionally autopolyploid, since although it contains multiple haploid nuclei, there are still just two distinct genomes), mostly from the genus Leucoagaricus Locq. ex Singer (1948); 3) Leaf-cutting agriculture, with colony sizes ranging from thousands to several millions of workers associated with just one type of fungus (Leucoagaricus gongylophorus [Möller] Singer (1986)) with cells that are multinucleate and multigenomic (i.e., functionally allopolyploid, with each cell containing multiple nuclei with, on average, seven distinct genomes) (Mehdiabadi and Schultz 2010; Kooij et al. 2015). Because these fungi are genetically highly heterozygous and, in some cases, as described above, functionally polyploid, it has been difficult to assemble their genomes, the assemblies are, therefore, often highly fragmented (Aylward et al. 2013). Furthermore, published genome assembly sizes for these fungi are much larger (i.e., >100 Mb) compared with the average fungal genome size of 44.2 Mb for all fungi (1,850 species analyzed, Ramos et al. 2015) and ~50 Mb for Agaricales (based on data for 11 species, Gupta et al. 2018), the order to which these fungi belong. In order to see whether the existing fragmented genome assemblies have overestimated the genome sizes of these fungi, and to help optimize the parameters used by genome assemblers to increase genome contiguity, we generated robust genome size estimates with flow cytometry and compared them with genome sizes estimated from different genome assemblers.

We sampled material from the Fungarium at the Royal Botanic Gardens, Kew, for Leucoagaricus barssii (Zeller) Vellinga (2000) (KM164561), a free-living relative of the fungus-growing ant fungi and the type species for the genus Leucoagaricus. The fungal symbiont of the ant Cyphomyrmex costatus Mann 1922 (MS140512-07) was isolated and grown on Potato Dextrose Agar (Sigma-Aldrich, St Louis, MO) with added Yeast Extract (Thermo Fisher Scientific Oxoid Ltd, Basingstoke, UK). DNA was extracted from both samples using the Qiagen DNeasy Plant Tissue kit (Qiagen, Hilden, Germany) following manufacturer’s protocols. 2× 300 bp libraries for the Illumina MiSeq (Illumina, San Diego, CA) were prepared using the Illumina library preparation kit (see Supplemental Experimental Procedures, Supplementary Material online for detailed methods), and sequence data were checked for quality using FastQC (Andrews 2015). We supplemented our sequence data with previously published genomic data available for a C. costatus symbiont (Nygaard et al. 2016; Bioproject PRJNA295288 – 100610-02). We then assembled the genomes for each of the three data sets using four different software packages with default settings: 1) ABySS 2.0.0 (Jackman et al. 2017), 2) SGA 0.10.15 (Simpson and Durbin 2012), 3) SPAdes 3.7.1 (Nurk et al. 2015), and 4) SOAPdenovo 2.04 complemented with GapCloser 1.12 (Luo et al. 2012). All assemblies were further corrected for heterozygous regions and divergent diploid genomes using the package Redundans 0.12a (Przybycz and Gabaldón 2016) (supplementary fig. S1, Supplementary Material online). This resulted in a total of eight assemblies for each data set. All assemblies were assessed for quality using several different genome statistics (table 1) which were generated using the ContigStats.pl script (written by Heath E. O’Brien, available at https://github.com/hobrien/...
Perl/blob/master/ContigStats.pl and the software package BUSCO v4 (Simão et al. 2015).

Our analyses revealed considerable differences in the total length of the assemblies between the four assembler packages, with ABySS creating the smallest and most consistent assemblies (largest N50 and longest contig) for the Leucocoprinus sp. data sets, ranging in size from 37.1 to 41.0 Mb, both with and without Redundans. The other assemblers resulted in assemblies ranging from 41 Mb for the previously published genomes for other fungi belonging to Agaricales. One possible explanation could lie in the fact that the fungi grown by ants show similarities to endosymbionts, that is, the fungus is protected by the ants in underground chambers with barely any contact with the outside world. Endosymbiotic bacteria have been shown to have reduced genomes, both in size and number of genes (McCutcheon and Moran 2012) and a similar reduction could be happening in these fungi. A more plausible explanation, however, could be that the software used (AUGUSTUS) is not able to predict all genes using just DNA data, and that other techniques, such as the use of transcriptome data, or an increased sequence coverage, may be necessary to recover all genes (Sims et al. 2014). Even so, these lower gene numbers do not detract from the overall observed pattern of higher numbers of genes with increased assembly length. As previously shown (Earl et al. 2011; Bradnam et al. 2013; Abbas et al. 2014), the assembly quality, based on N50, number of contigs and number of N’s, was also shown to depend on both the sample type (i.e.,

| Species/ID       | Assembler ± Redundans | Total Length (bp) | No. of Contigs (bp) | N50 (bp) | Longest Contig (bp) | N’s | No. of Predicted Genes | Total BUSCO Genes (% | BUSCO Genes Duplicated (%) |
|------------------|-----------------------|-------------------|---------------------|----------|---------------------|----|-----------------------|----------------------|-------------------------|
| Leucocoprinus sp. | ABySS –               | 37,868,966        | 7,777               | 51,836   | 742,525             | 1,695,483 | 6,730                | 94.9 | 0.6                  |
|                  | ABySS +               | 41,010,461        | 2,091               | 111,840  | 1,288,898           | 96,942  | 7,054                | 97.5 | 0.7                  |
|                  | SGA –                 | 94,552,020        | 48,075              | 3,297    | 92,288              | 0      | 14,239               | 83.6 | 0.4                  |
|                  | SGA +                 | 64,486,262        | 5,151               | 31,855   | 524,303             | 95,670  | 9,385                | 96.5 | 1.5                  |
|                  | SOAPdenovo –          | 171,922,479       | 20,433              | 19,233   | 200,771             | 24,532,942 | 23,180               | 88.8 | 19.8                 |
|                  | SOAPdenovo +          | 190,471,173       | 10,147              | 30,961   | 285,616             | 2,602,102 | 26,171               | 96.0 | 26.1                 |
|                  | SGA –                 | 101,100,038       | 27,307              | 9,377    | 213,758             | 533,869 | 14,882               | 91.4 | 0.4                  |
|                  | SGA +                 | 79,165,892        | 4,326               | 75,775   | 496,909             | 104,891 | 11,069               | 97.5 | 1.0                  |
| Leucocoprinus sp. | ABySS –               | 37,642,602        | 9,676               | 16,119   | 234,386             | 223,574 | 6,794                | 94.4 | 0.5                  |
|                  | ABySS +               | 37,057,723        | 4,778               | 24,356   | 234,386             | 4,684  | 6,715                | 95.7 | 0.5                  |
|                  | SGA –                 | 150,310,303       | 151,153             | 1,024    | 54,556              | 0      | 22,568               | 69.3 | 5.2                  |
|                  | SGA +                 | 58,439,858        | 32,056              | 2,491    | 54,556              | 3,648  | 9,703                | 70.5 | 0.3                  |
|                  | SOAPdenovo –          | 80,942,739        | 30,870              | 4,169    | 42,629              | 450,122 | 11,837               | 67.3 | 0.3                  |
|                  | SOAPdenovo +          | 63,729,975        | 12,609              | 8,493    | 130,081             | 116,914 | 9,226                | 82.3 | 0.5                  |
|                  | SGA –                 | 124,500,266       | 31,228              | 14,053   | 190,955             | 438,978 | 19,134               | 65.3 | 6.8                  |
|                  | SGA +                 | 101,793,281       | 12,552              | 21,931   | 190,955             | 7,959  | 15,194               | 74.2 | 2.6                  |
| Leucoagaricus barssii | ABySS –           | 32,671,039        | 4,005               | 28,088   | 430,464             | 255,022 | 6,552                | 94.5 | 0.9                  |
|                  | ABySS +               | 33,544,974        | 2,248               | 42,155   | 558,264             | 4,497  | 6,641                | 95.7 | 0.5                  |
|                  | SGA –                 | 78,616,683        | 74,089              | 1,143    | 54,770              | 0      | 15,165               | 66.3 | 4.9                  |
|                  | SGA +                 | 43,854,981        | 25,027              | 2,392    | 54,770              | 3,356  | 9,083                | 67.8 | 0.5                  |
|                  | SOAPdenovo –          | 39,685,961        | 4,810               | 19,634   | 245,171             | 104,891 | 11,069               | 97.5 | 1.0                  |
|                  | SOAPdenovo +          | 79,165,892        | 4,326               | 75,775   | 490,909             | 104,891 | 11,069               | 97.5 | 1.0                  |

Table 1

| Genome Statistics for the Different Assemblies |
|-----------------------------------------------|

Note.—Genome statistics as extracted using the ContigStats.pl script and the BUSCO pipeline. Assemblies with and without Redundans optimization are marked with – or +, respectively. Full BUSCO results are presented in supplementary table S1, Supplementary Material online.
genetically heterozygous symbionts vs. free-living fungi) and, potentially, the genome size. It is therefore important to have prior knowledge of the genome size when embarking on any genome assembly project.

Because the different assemblers generated different assembly sizes, we estimated the genome sizes for the sequenced species using flow cytometry (Dolézel et al. 2007). We isolated mycelium from three different ant colonies previously collected in Gamboa, Panama (Leucocoprinus sp. from C. costatus, and Leucocoprinus sp. from M. ednaella), and also from the oyster mushroom Pleurotus ostreatus. All of the ant symbionts were found to have genome sizes which fell close to the global fungal average of 44.2 Mb/1C, with L. gongylophorus at 39.86 Mb, Leucocoprinus sp. from C. costatus at 47.17 Mb and Leucocoprinus sp. from M. ednaella at 49.10 Mb (table 2).

Flow cytometry has been extensively used to estimate genome sizes over the last decades providing tens of thousands of estimates for eukaryotic organisms (e.g., Dell’Amico et al. 2005 [fungi], Pellicer and Leitch 2020 [plants], Gregory 2020 [animals]). The advent and growth of sequencing technologies has meant that there is now an increasing amount of whole-genome short-read sequence data available, which are also increasingly being used to estimate genome sizes based either on k-mer analyses or by mapping short reads to contiguous assemblies (e.g., Sun et al. 2018; Pucker 2019). These novel estimates, however, sometimes differ with those obtained by flow cytometry, although the underlying causes are still somewhat unclear. The nature of the genome (i.e., presence of polyploidy, abundance and composition of repetitive DNA, and level of heterozygosity) could impose challenges when using standard bioinformatic pipelines. For this reason, large comparative analyses using both flow cytometry and genomic approaches on the same specimens are urgently needed.
Table 2

| Species                                             | ID          | 1C-value (Mb) | Standard Deviation (Mb) | CV% (Standard) | CV% (Target) |
|-----------------------------------------------------|-------------|---------------|-------------------------|----------------|--------------|
| *Pleurotus ostreatus*                                | KM237125    | 24.17         | 0.39                    | 4.95           | 6.75         |
| *Leucoagaricus gongylophorus* (Atta colombica)       | Ac-2009-42  | 39.86         | 0.43                    | 4.00           | 5.29         |
| *Leucocoprinus* sp. (Cyphomyrmex costatus)          | MS140512-07 | 47.17         | 0.10                    | 3.94           | 4.45         |
| *Leucocoprinus* sp. (Myrmicocrypta ednaella)        | MS140507-01 | 49.10         | 0.79                    | 4.24           | 5.23         |

NOTE.—Name of the ant species is given in parentheses below the fungal species name. The 1C-value represents the DNA content of the unreplicated haploid chromosome complement (i.e., the holoploid genome size sensu Greilhuber et al. 2005). CV% is the fluorescence peak width expressed as coefficient of variation.

![Flow Cytometry Histograms](https://academic.oup.com/gbe/article-lookup/12/12/2384)

**FIG. 2.—** Examples of flow cytometry histograms used to estimate genome size. An example of the flow cytometry histograms obtained showing results for *Leucoagaricus gongylophorus* (A and C) and *Leucocoprinus* sp. (B and D) either without the standard *Pleurotus ostreatus* (A and B) or with (C and D).
needed (Pellicer and Leitch 2020) to understand what factors are responsible for the discrepancies observed and hence determine whether sequence data can ever reliably be used to provide robust genome size estimates.

Our flow cytometry histograms both showed high 1C (haploid) fluorescence peaks and low 2C (diploid) peaks (fig. 2). Because fungal nuclei are normally haploid and bearing in mind that we used fresh mycelium for analysis, the presence of 2C peaks indicates that some cells were in a premitotic division status (i.e., G2 phase of the cell cycle) at the time of measurement. An earlier study showing that the leaf-cutting ant fungus L. gongylophorus is functionally polyploid raised the possibility that each cell either contained multiple genomes within each nucleus (i.e., forming polyploid nuclei) or each individual nucleus contained just a single genome (i.e., forming haploid nuclei; Kooij et al. 2015). By extrapolating microsatellite data, it was suggested that each nucleus within this fungus was polyploid, however, our results suggest that the majority of nuclei analyzed are indeed haploid and hence the different genomes in each cell are present in different nuclei throughout the mycelium. Our results, therefore, suggest that the heterozygosity found in this fungus is most likely caused by SNPs in orthologs (i.e., genes from two different species with a common gene ancestor), possibly coming from divergent species, rather than paralogs (i.e., genes with different function arising from a gene or genome duplication event).

The genome size estimated by flow cytometry for the Leucocoprinus sp. from C. costatus is smaller than that estimated by most of the assemblies we obtained for this sample (table 1). In general, most assemblers using short-read sequence data have problems in compiling and assembling reads from long tandem repeat regions of the genome. Therefore, one might expect a smaller assembly size than the actual genome size, as has been reported in a previous study (Tavares et al. 2014). Based on the new genome size data generated using flow cytometry and various genome metrics such as total genome assembly length, N50, longest contig size and total percentage of BUSCO genes (see table 1), we conclude that out of the four assembly pipelines tested here, ABySS is the most accurate assembler for our samples.

In conclusion, our study has highlighted the importance of estimating genome size using flow cytometry prior to undertaking a whole genome sequencing and assembly project. This is essential given that prior knowledge of the genome size is essential to evaluate the quality of the genome assemblies and will avoid inferring incorrect gene number expansions, given the correlation observed between gene number and assembly size.

**Supplementary Material**

Supplementary data are available at Genome Biology and Evolution online.

**Acknowledgments**

We thank the Smithsonian Tropical Research Institute (STRI), Panama, for providing logistic help and facilities to work in Gamboa, the Autoridad Nacional del Ambiente y el Mar (ANAM) for permission to sample ant colonies in Panama and export them to Denmark. We also thank Jacobus J. Boomsma for providing the fungal cultures from the fungus-growing ants and Morten Schiøtt and Susanne den Boer for collecting the colonies. We also thank Jacobus J. Boomsma, Michael Chester, Ester Gaya, Ilija J. Leitch, and Morten Schiøtt for their helpful comments, discussion, and English editing, and László Nagy and two anonymous reviewers for their constructive comments, which improved previous versions of this manuscript. P.W.K. received funding from CAPES-PrInt (Grant #88887.468939/2019-00) and J.P. benefited from a Ramón y Cajal Fellowship (RYC-2017-2274) from the government of Spain. This research received no external funding.

**Author Contributions**

P.W.K. and J.P. conceived of the study; P.W.K. collected and identified the samples; P.W.K. carried out the genomic lab work and analyses; J.P. carried out the flow cytometry analyses; and P.W.K. and J.P. wrote the manuscript.

**Data Availability**

Sequence reads for the samples used in this research are available at NCBI under project number PRJNA560224: Leucoagaricus barrsii (KM164561) with BioSample number SAMN12572442 and Leucocoprinus sp. (MS140512-07) with BioSample number SAMN12572441. All assemblies generated are available on Figshare (DOI: 10.6084/m9.figshare.8153204). Scripts used for genome analyses are available on Github: [https://github.com/pwkooij/genome-scripts/blob/master/genome_analyses_scripts.sh](https://github.com/pwkooij/genome-scripts/blob/master/genome_analyses_scripts.sh) and [https://github.com/hobrien/Perl/blob/master/ContigStats.pl](https://github.com/hobrien/Perl/blob/master/ContigStats.pl).

**Literature Cited**

Abbas MM, Malluhi QM, Balakrishnan P. 2014. Assessment of de novo assemblers for draft genomes: a case study with fungal genomes. BMC Genomics 15(S9):S10.
Andrews S. 2015. FastQC – a high throughput sequence QC analysis tool. Available from: [http://www.bioinformatics.babraham.ac.uk/projects/fastqc/](http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Accessed October 26, 2020.
Aylward FO, et al. 2013. Leucoagaricus gongylophorus produces diverse enzymes for the degradation of recalcitrant plant polymers in leaf-cutter ant fungus gardens. Appl Environ Microbiol. 79(12):3770–3778.
Bennett MD, Leitch II, Price HJ, Johnston JS. 2003. Comparisons with Caenorhabditis (approximately 100 Mb) and Drosophila (approximately 175 Mb) using flow cytometry show genome size in Arabidopsis to be approximately 157 Mb and thus approximately
25% larger than the Arabidopsis genome initiative estimate of approxi-
mately 125 Mb. Ann Bot. 91(5):547–557.
Bradnam KR, et al. 2013. Assemblathon 2: evaluating de novo methods of
genome assembly in three vertebrate species. Giga Sci. 2:10.
Clément L, et al. 2018. A data-assembled history of bioinformatics tools.
arXiv. 1807.06808.
De Man TB, et al. 2016. Small genome of the fungus Escoovopsis weberi,
a specialized disease agent of ant agriculture. Proc Natl Acad Sci USA.
113(13):3567–3572.
Dentinger BTM, et al. 2016. Tales from the crypt: genome mining from
fungarium specimens improves resolution of the mushroom tree of
life. Biol J Linn Soc. 117(1):11–32.
Doležel J, Greilhuber J, Suda J. 2007. Estimation of nuclear DNA content in
plants using flow cytometry. Nat Protoc. 2(9):2233–2244.
Earl D, et al. 2011. Assemblathon 1: a competitive assessment of de novo
short read assembly methods. Genome Res. 21(12):2224–2241.
Galbraith DW, et al. 1983. Rapid flow cytometric analysis of the cell cycle
in intact plant tissues. Science 220(4601):1049–1051.
Gregory TR. 2020. Animal genome size database. http://www.genome-
size.com. Accessed October 26, 2020.
Greilhuber J, Doležel J, Lysák MA, Bennett MD. 2005. The origin, evolution and
proposed stabilization of the terms ‘genome size’ and “C-value”
to describe nuclear DNA contents. Ann Bot. 95(1):255–260.
Gremme G, Steinbiss S, Kurtz S. 2013. GenomeTools: a comprehensive
software library for efficient processing of structured genome anno-
ations. IEEE/ACM Trans Comput Biol Bioinf. 10(3):645–656.
Gupta DK, et al. 2018. The genome sequence of the commercially culti-
vated mushroom Agrocybe aegerita reveals a conserved repertoire of
fruiting-related genes and a versatile suite of biopolymer-degrading
enzymes. BMC Genomics 19(1):48.
Jackman SD, et al. 2017. ABySS 2.0: resource-efficient assembly of large
genomes using a Bloom filter. Genome Res. 27(5):768–777.
Knapp DG, et al. 2018. Comparative genomics provides insights into the
lifestyle and reveals functional heterogeneity of dark septate endo-
phytic fungi. Sci Rep. 8(1):6321.
Kooij PW, Aanen DK, Schiøtt M, Boomsma JJ. 2015. Evolutionarily advanced
ant farmers rear polyploid fungal crops. J Evol Biol. 28(11):1911–1924.
Kullman B, Tamm H, Kullman K. 2005. Fungal genome size database.
Available from: http://www.zbi.ife/fungal-genomes/. Accessed
June 12, 2019.
Lai Y, et al. 2014. Comparative genomics and transcriptomics analyses
reveal divergent lifestyle features of nematode endoparasitic fungus
Hirsutella minnesotaensis. Genome Biol Evol. 6(11):3077–3093.
Larriba E, et al. 2014. Sequencing and functional analysis of the genome of
a nematode egg-parasitic fungus, Pochonia chlamydosporia. Fungal
Genet Biol. 65:69–80.
Liu K, et al. 2014. Drechslerella stenodrocha genome illustrates the mech-
anism of constricting rings and the origin of nematode predation in
fungi. BMC Genomics 15(1):114.
Luo R, et al. 2012. SOAPdenovo2: an empirically improved memory-
efficient short-read de novo assembler. Giga Sci. 1:18.
Martino E, et al. 2018. Comparative genomics and transcriptomics depict
ericoid mycorrhizal fungi as versatile saprotophs and plant mutualists.
New Phytol. 217(3):1213–1229.
McCuteachon JP, Moran NA. 2012. Extreme genome reduction in symbi-
ocic bacteria. Nat Rev Microbiol. 10(1):13–26.
Mehdiabadi NJ, Schultz TR. 2010. Natural history and phylogeny of the
fungus-farming ants (Hymenoptera: Formicidae: Myrmicinae: Attini).
Myrmecol News. 13:37–55.
Mueller UG, Gerardo NM, Aanen DK, Six DL, Schultz TR. 2005. The evo-
lution of agriculture in insects. Annu Rev Ecol Evol Syst. 36(1):563–595.
Nurk S, et al. 2013. Assembling single-cell genomes and mini-
etagenomes from chimeric MDA products. J Comput Biol.
20(10):714–737.
Nygaard S, et al. 2016. Reciprocal genomic evolution in the ant-fungus
agricultural symbiosis. Nat Commun. 7(1):12233.
Pellicer J, Leitch U. 2020. The Plant DNA C-values database (release 7.1): an
updated online repository of plant genome size data for comparative
studies. New Phytol. 226(2):301–305.
Przybycz LP, Galdabino T. 2016. Redundans: an assembly pipeline for highly
heterozygous genomes. Nucleic Acids Res. 44(12):e113.
Pucker B. 2019. Mapping-based genome size estimation. bioRxiv.
doi:10.1101/607390.
Quandt CA, Bushley KE, Spatafora JW. 2015. The genome of the truffle-
parasite Tolypocladium ophioglossoides and the evolution of antfungal
peptibiotics. BMC Genomics 16(1):553.
Ramos AP, et al. 2015. Flow cytometry reveals that the rust fungus,
Uromyces bidentis (Pucciniales), possesses the largest fungal genome
reported-2489 Mbp. Mol Plant Pathol. 16(9):1006–1010.
Rubin BER, Moreau CS. 2016. Comparative genomics reveals convergent
rates of evolution in ant-plant mutualisms. Nat Commun. 7(1):12679.
Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM.
2015. BUSCO: assessing genome assembly and annotation complete-
ness with single-copy orthologs. Bioinformatics 31(19):3210–3212.
Simpson JT, Durbin R. 2012. Efficient de novo assembly of large genomes
using compressed data structures. Genome Res. 22(3):549–556.
 Sims D, Sudbery I, Iott NE, Heger A, Ponting CP. 2014. Sequencing depth
and coverage: key considerations in genomic analyses. Nat Rev Genet.
15(2):121–132.
Stanke M, Steinkamp R, Waack S, Morgenstern B. 2004. AUGUSTUS: a
web server for gene finding in eukaryotes. Nucleic Acids Res. 32(Web
Server):W309–W312.
Stephens ZD, et al. 2015. Big data: astronomical or genomical? PLoS Biol.
13(7):e1002195–11.
Sun H, Ding J, Reddol M, Schneeberger K. 2018. findCSE: estimating
genome size variation within human and Arabidopsis using k-mer
frequencies. Bioinformatics 34(4):550–557.
Tavares S, et al. 2014. Genome size analyses of Pucciniales reveal the
largest fungal genomes. Front Plant Sci. 5:422.
Teixeira MM, et al. 2017. Exploring the genomic diversity of black yeasts
and relatives (Chaetothyriales, Ascomycota). Stud Mycol. 86:1–28.
Thompson JN. 2005. Coevolution: the geographic mosaic of coevolu-
ary arms races. Curr Biol. 15(24):R992–R994.
Wang Y-Y, et al. 2014. Genome characteristics reveal the impact of lichen-
ization on lichen-forming fungus Endocarpon pusillum Hedwig
(Verrucariales, Ascomycota). BMC Genomics 15(1):34.
Weber NA. 1979. Historical note on culturing attine-ant fungi. Mycologia
71(3):633–634.
Wetterstrand KA. 2019. DNA sequencing costs: data. Available from:
http://www.genome.gov/sequencingcostsdata. Accessed June 12,
2019.
Wheeler WM. 1907. The fungus-growing ants of North America. B Am
Mus Nat Hist. 23:669–U130.
Yang J, et al. 2011. Genomic and proteomic analyses of the fungus
Arthrobtys oligospora provide insights into nematode-trap forma-
tion. PLoS Pathog. 7(9):e1002179.

Associate editor: Laura Katz