

Fungal rDNA Analysis in the NGS Era: Progress, Insights and Challenges

Oliver Blechert, Ping Zhan*

Department of Dermatology, Jiangxi University of Chinese Medicine, Nanchang, China

ABSTRACT

Background: Ribosomal DNA (rDNA), consisting of the 18S, 5.8S, 25S and 5S rRNA genes and noncoding functional elements in eukaryotes, is a region in the genome with very high activity with respect to transcription and genomic rearrangement. Due to its highly repetitive and complex structure, detailed sequence analyses are challenging and often neglected. In this study, a systematic survey of rDNA organization across the fungal kingdom was performed.

Methods: We analysed all fungal genome assemblies available at NCBI, in total 6779, by bioinformatic methods. For this, we developed a batch of Python scripts to process the data. For de-novo assembling of NGS data, we wrote the Ansi C program 'YNGS'.

Results: Only 15% of the fungal genome assemblies at NCBI possessed the sequence of an entire rDNA unit. Most assemblies of the rDNA were incomplete and were interrupted in the internal spacer regions. Next, we analysed the rDNA organisation. Basidiomycota, with the exception of the Ustilaginomycetes, had 5S gene inserted in the rDNA units as well Blastocladiomycota, Chytridiomycota and Mucoromycota. Ascomycota, with the exception of Saccharomycetes, lacked 5S insertions in the rDNA units. Here, the 5S genes were scattered in the genome and tandem arrays of 5S genes were very uncommon. With the YNGS software, we assembled 17 entire rDNA unities and calculated an amount of 25 to 161 rDNA repeated units per genome.

Conclusion: In contrast to the rDNA gene region, the internal spacer regions are still challenging to analyse, even in the NGS era. Transcriptional regulation, replication machinery and genome rearrangement elements are meeting together. Sophisticated regulations are necessary for a smooth flow.

Keywords: Genome; rDNA; 5S insertion; Mycology; Bioinformatics

INTRODUCTION

Ribosomal RNAs (rRNAs) are essential macro-molecules and are present in all living cells, in Bacteria, Archaea and Eukaryota. Together with protein, they are forming together the ribosomes, the machinery translating mRNA into protein. For the correct translation, a matching tRNA docks to the mRNA template at the small sub-unit (SSU), whereas the large sub-unit (LSU) facilitates the process of protein synthesis. In Eukaryota, the SSU contains a single RNA molecule, the 18S, whereas the LSU has three different kinds of rRNA molecules, 5S, 5.8S and 25S (Table 1). For historical reasons, the different rRNAs are named according to their sedimentation coefficients measured in Svedberg units (S). Unfortunately, the use

of this unit leads to inconsistency in the nomenclature of the rRNA molecules, since the sedimentation coefficient of homologous rRNA differs among taxa. Following, we are using the nomenclature of *Saccharomyces cerevisiae*.

rDNA organization in the eukaryotic genome

In eukaryotic organisms, ribosomal genes are present in multiple repeat units to satisfy the high demand for ribosomes. They are organized in one to a few arrays (Figure 1A). In plants and animals, 5S rDNA genes are commonly separated from main rDNA arrays in Figure 1B, but are commonly organized in separate arrays of tandem repeats (Potapova and Gerton) [1].

Correspondence to: Ping Zhan, Department of Dermatology, Jiangxi University of Chinese Medicine, Nanchang, China, E-mail: zhanping1980@163.com

Received: 02-Nov-2022, Manuscript No. FGB-22-19117; **Editor assigned:** 04-Nov-2022, PreQC No. FGB-22-19117 (PQ); **Reviewed:** 18-Nov-2022, QC No. FGB-22-19117; **Revised:** 25-Nov-2022, Manuscript No. FGB-22-19117 (R); **Published:** 05-Dec-2022, DOI: 10.35841/2165-8056.22.12.202

Citation: Blechert O, Zhan P (2022) Fungal rDNA Analysis in the NGS Era: Progress, Insights and Challenges. Fungal Genom Biol. 12: 210.

Copyright: © 2022 Blechert O, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

rDNA copy number varies between 39 and 19 300 in animals and from 150 to 26 048 in plants, whereby the copy number and genome size of the organism are strongly correlated (Prokopowich et al.) [2]. However, large variations in gene numbers exist not only among closely related species but also among different strains of the same taxon (Long and Dawid et al.) [3].

Challenges in studying rDNA organization

Most bioinformatic studies focus on low-copy DNA regions. Analyses of repeated sequences are often neglected. In the final assemblies, the repeated sequences are omitted, are collapsed to a few repeated units or cause interruptions in the contigs (Biscotti et al.) [4]. From a computational perspective, highly repeated sequences can create ambiguities in alignments and can subsequently create errors in genome assemblies (Treangen and Salzberg) [5]. De novo assembly software frequently fails to resolve rDNA sequences. Even when combining NGS (Illumina and 454) data, Sanger sequences and BLAST searches against the National Center for Biotechnology Information (NCBI) database, only one of five programs worked efficiently (Agrawal and Ganley) [6].

rDNA within fungal model organisms

The genomes of the four most intensively studied fungal model species were resolved at the chromosome level. To accomplish this, the assemblies were based on various sequencing data, next-generation sequencing (NGS), bacterial artificial chromosome (BAC), Fosmid, Cosmid and Sanger sequencing data.

***Saccharomyces cerevisiae*:** Each rDNA unit, with a length of 9.1 kB, consists of one 5S, 5.8S, 15S and 25S gene (Figure 1C). The genes are separated by three kinds of spacers: nontranscribed spacers (NTS1 and NTS2), external transcribed spacers (ETS1 and 3' ETS2) and internal transcribed spacers (ITS1 and ITS2). The rDNA is organized into 100-200 tandem repeats on chromosome XII. In the reference genome assembly (R64-2-1), only two units of the array are displayed (www.yeastgenome.org, Christie et al.) [7].

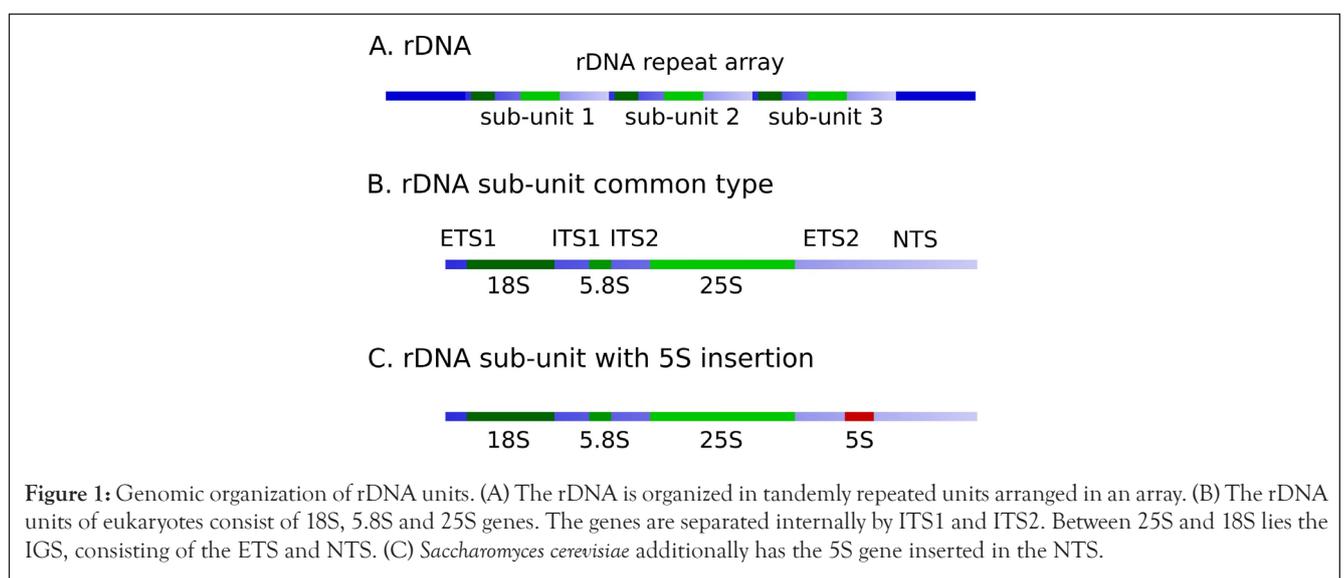
***Candida albicans*:** The arrangement of the rDNA genes in *C. albicans* is similar to that in *S. cerevisiae* but includes an additional low-complexity region of approximately 2 kb. Each of the approximately 55 units, estimated for a haploid set of chromosomes, is 12.8 kB long (Jones et al.) [8]. In the *C. albicans* 'SC5314' genome assembly, only one unit is displayed (www.candidagenome.org, Skrzypek et al.) [9].

***Schizosaccharomyces pombe*:** In *S. pombe*, the rDNA is arranged in two tandem arrays, estimated to be 1225 kb and 240 kb in size. In the genome assembly ASM294v2 of strain 972h, the two tandem arrays flank both ends of chromosome III and are only partially displayed. Distal flanking regions are missing in the assembly (www.pombase.org, Lock et al.) [10]. The 5S genes are, in contrast to those in *S. cerevisiae* and *C. albicans*, dispersed within the genome (Mao et al.) [11]. The six copies of 5S rDNA (SPRRNA.03, 04, 05, 06, 07 and 10) are not linked to each other.

Table 1: Common rDNA nomenclature in fungi, animals and bacteria.

Fungi*I	Human*II	Bacteria*III	Subunits	RS
80S	80S	70S	Ribosome	RS
60S	60S	50S	large subunit (LSU) of the ribosome	RS
40S	40S	30S	small subunit (SSU) of the ribosome	RS
35S	45S	30S	pre-rRNA molecule; before splicing	RS
25S	28S	23S	rRNA of LSU; transcribed by the RNA polymerase I	RS
18S	18S	16S	rRNA of the SSU; transcribed by the RNA polymerase I	RS
5.8S	5.8S	-	rRNA of LSU; transcribed by the RNA polymerase I	RS
5S	5S	5S	rRNA of LSU, RNA polymerase III	RS

Note: *I based on *Saccharomyces cerevisiae*, *II based on *Homo sapiens*, *III based on *Escherichia coli*.



***Aspergillus nidulans*:** The rDNA units of *A. nidulans* are relatively short, with a size of 7.7 kB. The intergenic spacer (IGS), consisting of the NTS and ETS, is just 1.7 kB long (Lockington et al.) [12]. In the genome assembly at *Aspergillus*-database, four of the rDNA units are displayed, whereas the 5S genes are not mapped (www.aspergillusgenome.org, Cerqueira et al.) [13]. In *A. nidulans*, the 5S rDNA genes are dispersed throughout the genome and are not organized in arrays. Six 5S rDNA fragments are present, four have complete 5S rRNA sequences and two are *pseudogenes* (Bartnik et al.) [14].

MATERIALS AND METHODS

All calculations were performed with a ThinkPad P52 equipped with an i7 processor and 64 GB RAM. The operating system was a Debian Release 10 Linux operating system with Python 3.7 and GCC version 8.3.

Fungal assemblies, rDNA annotation and analysis

Four datasets of fungal assemblies were downloaded from the NCBI (Table 2). To analyse the DNA sequences, we wrote six Python scripts in Table 3 and run the scripts successively (Figure 2A). First, we created a list of all genome assemblies, whereby duplicate entries were removed, by successively running Python scripts #1 and #2. In the next step, we annotated the rDNA genes with RNAmmer version 1.2 software (Lagesen et al.) [15]. For automatic analysis of the datasets, Python script #3 unzipped each assembly and consecutively called RNAmmer software. Next, we sorted the rDNA annotations with Python script #4. To correlate the rDNA organization in the genome with phylogenetic information, we downloaded lineage information from the NCBI Taxonomy Database for each assembly and created

a list with information about the kingdom, phylum, class, order, family and genus (Python script #5). Python script #6 analysed the data and combined the results with phylogenetic information.

rDNA sequences assembled from fastq files

For de novo assembly of rDNA repeat units, we downloaded NGS data in Table 4 from the Sequence Read Archive (SRA) of NCBI. For the de novo assembly of rDNA sequences, we wrote YNGS v1.0 software and released it as open source software at Bioinformatics.org. YNGS software assembles DNA sequences from a 50 bp seed by elongating the sequence with 24 bp per loop (Figure 2C). In cases of interruption of the elongation, due to parts of the sequence that are difficult to align, the fastq forwards and paired-end sequences were displayed in a separate window and the string was elongated manually. Afterwards, the quality of the de novo assemblies were verified by counting the fastq fragments matching each position of the assembly (Figure 2D) with the sliding window approach.

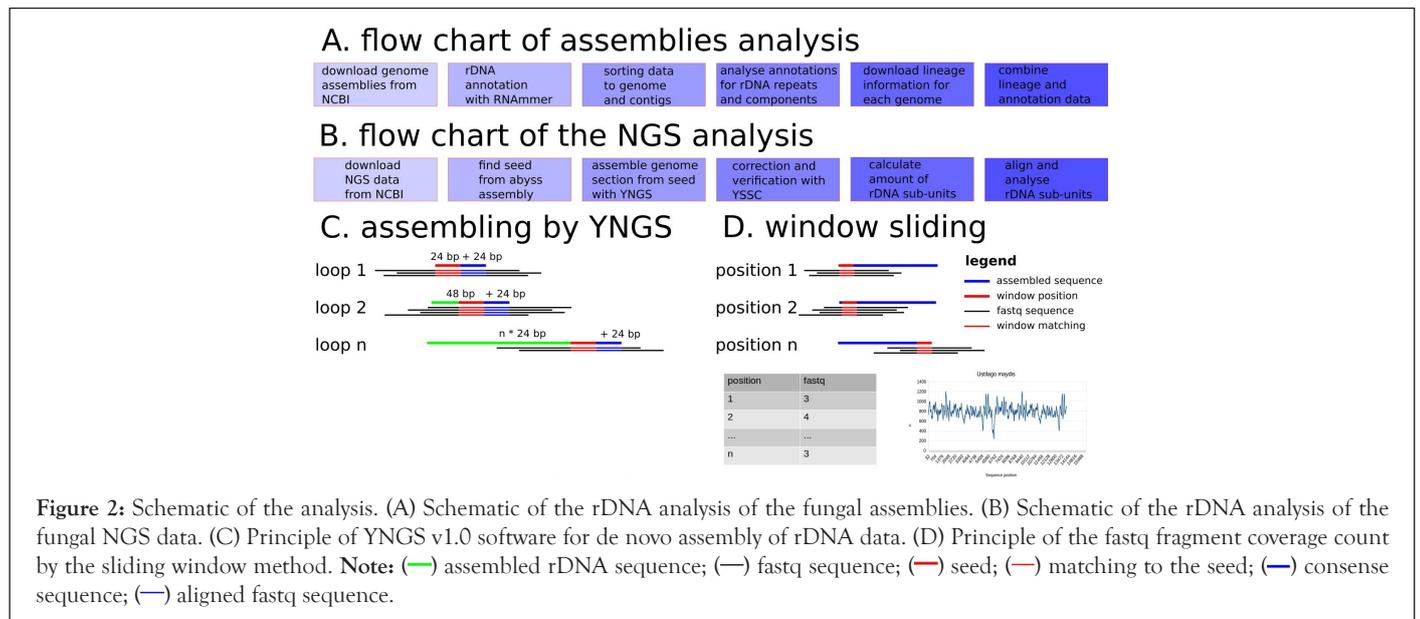
The rDNA sequences were aligned with Clustalx v2.1 software. For the annotation of the rDNA genes, data for the model organism *S. cerevisiae* were used as a template. The phylogenetic dendrogram was created with the sequence from the 35S gene with Phylip v3.697 software. The alignment was bootstrapped with 100 replications. For the calculation of repeats per genome, we identified the fastq fragments matching the rDNA sequence. For normalization, we used the four genes ACT1, TRP3, TUB4 and URA3. We calculated the number of fastq fragments matching the four sequences with the sliding window application in YNGS software. Finally, we divided the average amount of rDNA matching fastq fragments with the amount of the normalizing genes.

Table 2: Databases of fungal assemblies downloaded from the NCBI assembly database.

Database name	Release Type	Filter	Entries
GenBank_total	GenBank	fungi, latest	6453
GenBank_genome	GenBank	fungi, latest and complete genome	72
RefSeq_total	RefSeq	fungi, latest	326
RefSeq_genome	RefSeq	fungi, latest and complete genome	11

Table 3: Python scripts for annotating and analysing rDNA genes in assemblies downloaded from the NCBI database.

Python	Database
#1	Create a list of assembly name, fasta file comment of each contig and contig length of the gzipped fasta files
#2	Computes as list of assembly name, species name and database
#3	Calls the software RNAmmer on a batch of gzipped fasta files
#4	Python script for summarizing and preprocessing RNAmmer results
#5	Python script for getting the lineage, phylogeny, for all genome assemblies
#6	Create a list of the maximal 35S and 5S rDNA repeat amount for each assembly and add lineage information
#The rDNA genes that analysing in NCBI database.	

**Table 4:** Fastq files downloaded from the NCBI SRA database for de novo assembling.

Taxon	sra file downloaded from NCBI
<i>Phaeotremella foliacea</i>	DRR032527.sra
<i>Cerinomyces ceraceus</i>	SRR4151924.sra
<i>Boletus edulis</i>	SRR5764827.sra
<i>Moesziomyces sp.</i>	SRR6053896.sra
<i>Ustilago maydis</i>	DRR032503.sra
<i>Rhodotorula mucilaginosa</i>	SRR11859272.sra
<i>Verticillium dahliae</i>	SRR12769281.sra
<i>Sordaria macrospora</i>	SRR5749461.sra
<i>Morchella importuna</i>	SRR4136319.sra
<i>Candida auris</i>	SRR12784121.sra
<i>Saccharomyces cerevisiae</i>	SRR4244871.sra
<i>Candida parapsilosis</i>	SRR8984820.sra
<i>Schizosaccharomyces pombe</i>	SRR11003486.sra
<i>Taphrina betulina</i>	SRR10079519.sra
<i>Paraphysoderma sedebokerense</i>	SRR6447920.sra
<i>Allomyces javanicus</i>	SRR11025671.sra
<i>Geranomyces variabilis</i>	SRR12278674.sra
<i>Zygnema circumcarinatum</i>	SRR8552832.sra
<i>Chlorella vulgaris</i>	SRR8083359.sra
<i>Boletus edulis</i>	SRR5764827.sra
<i>Moesziomyces sp.</i>	SRR6053896.sra
<i>Ustilago maydis</i>	DRR032503.sra
<i>Rhodotorula mucilaginosa</i>	SRR11859272.sra
<i>Verticillium dahliae</i>	SRR12769281.sra
<i>Sordaria macrospora</i>	SRR5749461.sra
<i>Morchella importuna</i>	SRR4136319.sra
<i>Candida auris</i>	SRR12784121.sra
<i>Saccharomyces cerevisiae</i>	SRR4244871.sra
<i>Candida parapsilosis</i>	SRR8984820.sra
<i>Schizosaccharomyces pombe</i>	SRR11003486.sra
<i>Taphrina betulina</i>	SRR10079519.sra
<i>Paraphysoderma sedebokerense</i>	SRR6447920.sra
<i>Allomyces javanicus</i>	SRR11025671.sra
<i>Geranomyces variabilis</i>	SRR12278674.sra
<i>Zygnema circumcarinatum</i>	SRR8552832.sra
<i>Chlorella vulgaris</i>	SRR8083359.sra

RESULTS

To examine the genomic organization of the rDNA in fungal species, we analysed two different data sources. First, we analysed fungal genomic assemblies, which were available in the NCBI database. The high number of available datasets allowed a systematic study of rDNA organization. Second, we focused on a few manually selected species and assembled the rDNA units.

Fungal assemblies

A total of 6779 fungal assemblies were available in the NCBI databank. Of these, 6453 were in the GenBank dataset, and 326 were in the RefSeq dataset. In the GenBank dataset, 72 were assembled to genome quality, and in the RefSeq dataset, 11. With RNAmmer software, 5S genes were annotated in 6050 of the total 6779 assemblies, 18S genes in 4776 of the assemblies and 25S genes in 4556 of the assemblies (Table 5). In 1032 of the assemblies, at least one entire repeat unit was present. In 603 assemblies, at least three repeating units were present (Figure 3). Among all assemblies, less than 9% possessed at least three complete units, among those in GenBank, less than 9%, and among those in RefSeq, less than 14%. In the assemblies with genome quality, 61% of the GenBank and 36% of the RefSeq assemblies had at least three entire units in a row.

Organization of the rDNA array in the genome

Based on the annotation of the rDNA genes, we identified the rDNA arrays having 5S insertions in the intergenic spacer region. Therefore, we selected the largest rDNA array from each of the 603 assemblies. The 5S insertions were present in 298 of the arrays and were absent in 305.

To correlate the presence of 5S genes with phylogenetic data, we downloaded the lineage information for each taxon from the NCBI database. Most of the Basidiomycota had 5S insertions in the rDNA repeat units, whereas most of the Ascomycota lacked 5S insertions. For the phyla Blastocladiomycota, Chytridiomycota and Mucoromycota, five assemblies could be evaluated, and in all cases, 5S insertions were present in the rDNA arrays (Figure 4). We sub-classified the data to the class level. In the phylum Ascomycota, nearly all assemblies, with the exception of Saccharomycetes, lacked 5S insertions in the rDNA repeats. In more than 98% of the analysed Ascomycota in the classes Dothideomycetes, Eurotiomycetes, Leotiomycetes, Orbiliomycetes, Pezizomycetes, Schizosaccharomycetes and Sordariomycetes, 5S insertions were absent (Table 6). In contrast

to the other classes of Ascomycota, most taxa of the Saccharomycetes had 5S insertions. In 46 taxa, 5S insertions were present (Table 7). The 11 aberrant assemblies belonged to the genera *Hypophopichia*, *Komagataella*, *Metschnikowia*, *Sugiyamaella* and *Yarrowia* in Table 6, accounting for ~6% of the total Saccharomycetes assemblies.

In Basidiomycota, 5S insertions in the rDNA arrays were very common (Table 8), in Agaricomycetes, Malasseziomycetes, Microbotryomycetes, Moniliellomycetes, Pucciniomycetes and Tremellomycetes. In contrast, in the class Ustilaginomycetes, all 7 assemblies lacked 5S insertions in the rDNA arrays (Table 9). In the Mucoromycota in Table 10, Blastocladiomycota in Table 11 and Chytridiomycota in Table 12, 5S insertions were present.

Organization of 5S rDNA

In plants and animals, which commonly lack 5S insertions in rDNA arrays, 5S genes are frequently clustered in the genome and arranged in separated 5S gene arrays. To explore the distribution and size of 5S genes in the 603 fungal assemblies, we calculated the number of neighboring 5S genes (Table 13). In nearly all of the fungal assemblies, 5S gene arrays were not present. Most commonly, the 5S genes were single, separated and scattered in the genome (~79%). Only 8 assemblies (belonging to only four species) had arrays exceeding 5S gene repeats (Table 14). The largest 5S gene array was detected in assemblies of the powdery mildew *Blumeria graminis*, with 20 tandemly repeated 5S genes. *Anthracoctystis flocculosa*, a member of Ustilaginales, had 18 repeats, giving it the second largest 5S gene array. The 5S gene arrays in *S. cerevisiae* were striking. The dataset contained 124 *S. cerevisiae* assemblies and nine of them possessed neighbouring 5S genes. Whereas extended 5S gene rDNA arrays were rare, shorter arrays of 2 or 3 repeated 5S genes were more frequent. In some species, we observed an enrichment of two neighbouring 5S genes. In three fungal species, *Candida sake*, *Coccidioides immitis* and *Trichoderma virens*, more than 30% of the 5S genes were arranged in arrays of two neighbouring 5S genes, without the presence of larger arrays.

rDNA sequences assembled from NGS fastq files

We assembled the complete rDNA units of seventeen fungal species from NGS sequencing data (Table 15). To root the phylogenetic dendrogram in Figure 5, we additionally assembled the rDNA units of two algal species. The alga *Zygnema circumcarinatum* belongs to the Zygnematophyceae, whereas *Chlorella vulgaris* belongs to the Trebouxiophyceae.

Table 5: Annotation of rDNA genes and repeat length of the genome assemblies.

Assemblies	total	GenBank	GenBank genome	RefSeq	RefSeq genome
entries	6779	6453	72	326	11
5S*I	6050	5747	66	303	11
18S*II	4776	4574	67	202	11
25S*III	4556	4359	66	197	11
at least one repeat*IV	1032	946	56	86	7
at least three repeats*V	603	559	44	44	4

Note:*I amount of assemblies with at least one 5S rDNA gene annotation.

*II amount of assemblies with at least one 18S rDNA gene annotation.

*III amount of assemblies with at least one 25S rDNA gene annotation.

*IV amount of assemblies with at least one entire repeat unit of rDNA.

*V amount of assemblies with at least three repeat units of rDNA in a continuous row.

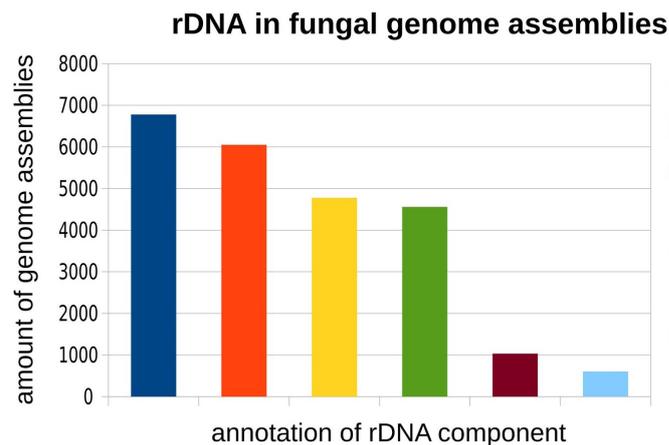


Figure 3: Annotation of rDNA genes in the fungal assemblies. The rDNA genes of 6779 assemblies were annotated. A total of 6050 assemblies had 5S genes, 4776 had 18S, and 4556 had 25S annotations. In 1409 of the assemblies, at least one complete rDNA unit in an array was annotated, and in 247 assemblies, at least three units in an array were annotated. **Note:** (■) assemblies; (■) with 5S; (■) with 18S; (■) with 25S; (■) with at least one repeat sub-unit; (■) with at least three repeat sub-unit

Table 6: Ascomycota taxa without 5S insertions in the rDNA arrays.

<i>Dothideomycetes</i>
<i>Acidomyces richmondensis</i> , <i>Alternaria alternata</i> , <i>Alternaria brassicae</i> , <i>Alternaria brassicicola</i> , <i>Alternaria gaisen</i> , <i>Alternaria solani</i> , <i>Amniculicola lignicola</i> , <i>Ascochyta lentis</i> , <i>Ascochyta rabiei</i> , <i>Ascochyta versabilis</i> , <i>Bipolaris sorokiniana</i> , <i>Botryosphaeria dothidea</i> , <i>Cucurbitaria berberidis</i> , <i>Didymella segeticola</i> , <i>Dothidotthia symphoricarpi</i> , <i>Hortaea werneckii</i> , <i>Karstenula rhodostoma</i> , <i>Leptosphaeria biglobosa</i> , <i>Leptosphaeria maculans</i> , <i>Lizonia empirigonia</i> , <i>Macrophomina phaseolina</i> , <i>Paraphaeosphaeria minitans</i> , <i>Parastagonospora nodorum</i> , <i>Peltaster fructicola</i> , <i>Phoma sp.</i> , <i>Pseudopyrenochaeta lycopersici</i> , <i>Pyrenophora teres</i> , <i>Pyrenophora tritici-repentis</i> , <i>Venturia inaequalis</i> , <i>Zymoseptoria tritici</i>
<i>Eurotiomycetes</i>
<i>Ajellomyces capsulatus</i> , <i>Arthroderma uncinatum</i> , <i>Aspergillus awamori</i> , <i>Aspergillus campestris</i> , <i>Aspergillus cejpai</i> , <i>Aspergillus cristatus</i> , <i>Aspergillus flavus</i> , <i>Aspergillus fumigatus</i> , <i>Aspergillus lentulus</i> , <i>Aspergillus nidulans</i> , <i>Aspergillus niger</i> , <i>Aspergillus ochraceus</i> , <i>Aspergillus oryzae</i> , <i>Aspergillus persii</i> , <i>Aspergillus steynii</i> , <i>Aspergillus terreus</i> , <i>Aspergillus tubingensis</i> , <i>Blastomyces dermatitidis</i> , <i>Blastomyces gilchristii</i> , <i>Coccidioides immitis</i> , <i>Coccidioides posadasii</i> , <i>Emergomyces orientalis</i> , <i>Exophiala lecanii-corni</i> , <i>Exophiala spinifera</i> , <i>Microsporium canis</i> , <i>Monascus purpureus</i> , <i>Paracoccidioides brasiliensis</i> , <i>Paracoccidioides lutzii</i> , <i>Paracoccidioides sp.</i> , <i>Penicillium capsulatum</i> , <i>Penicillium expansum</i> , <i>Penicillium oxalicum</i> , <i>Penicillium solitum</i> , <i>Penicillium sp.</i> , <i>Talaromyces funiculosus</i> , <i>Talaromyces marneffeii</i> , <i>Talaromyces pinophilus</i> , <i>Uncinocarpus reesii</i>
<i>Leotiomycetes</i>
<i>Articulospora tetracladia</i> , <i>Blumeria graminis</i> , <i>Ciboria shiraiana</i> , <i>Helotiales sp.</i> , <i>Podospaera xanthii</i> , <i>Pseudogymnoascus destructans</i> , <i>Sclerotinia sclerotiorum</i>
<i>Orbiliomycetes</i>
<i>Arthrotrichum flagrans</i> , <i>Arthrotrichum oligospora</i>
<i>Pezizomycetes</i>
<i>Morchella sextelata</i> , <i>Tuber indicum</i>
<i>Schizosaccharomycetes</i>
<i>Schizosaccharomyces japonicus</i> , <i>Schizosaccharomyces octosporus</i> , <i>Schizosaccharomyces pombe</i>
<i>Sordariomycetes</i>
<i>Amesia nigricolor</i> , <i>Amphirosellinia nigrospora</i> , <i>Aquanectria penicillioidea</i> , <i>Beauveria bassiana</i> , <i>Ceratocystis albifundus</i> , <i>Cladobotryum protrusum</i> , <i>Colletotrichum acutatum</i> , <i>Colletotrichum destructivum</i> , <i>Colletotrichum fructicola</i> , <i>Colletotrichum gloeosporioides</i> , <i>Colletotrichum higginsianum</i> , <i>Colletotrichum siamense</i> , <i>Colletotrichum sp.</i> , <i>Colletotrichum sublineola</i> , <i>Colletotrichum tofieldiae</i> , <i>Colletotrichum trifolii</i> , <i>Colletotrichum viniferum</i> , <i>Coniochaeta sp.</i> , <i>Cordyceps militaris</i> , <i>Cryphonectria parasitica</i> , <i>Daldinia childiae</i> , <i>Daldinia concentrica</i> , <i>Dicyma pulvinata</i> , <i>Entonaema liquescens</i> , <i>Epichloe festucae</i> , <i>Esteya vermicola</i> , <i>Fragosphaeria purpurea</i> , <i>Fusarium culmorum</i> , <i>Fusarium graminearum</i> , <i>Fusarium oxysporum</i> , <i>Fusarium venenatum</i> , <i>Fusarium verticillioides</i> , <i>Gaeumannomyces sp.</i> , <i>Grossmannia penicillata</i> , <i>Hypomontagnella monticulosa</i> , <i>Hypomyces perniciosus</i> , <i>Hypomyces rosellus</i> , <i>Hypoxylon fragiforme</i> , <i>Hypoxylon lienhwacheense</i> , <i>Hypoxylon pulicidum</i> , <i>Magnaportheales sp.</i> , <i>Magnaporthe oryzae</i> , <i>Neurospora crassa</i> , <i>Neurospora sp.</i> , <i>Ophiocordyceps camponoti-floridani</i> , <i>Ophiocordyceps sinensis</i> , <i>Pyrenopeziza hunteri</i> , <i>Pyricularia oryzae</i> , <i>Raffaella albimanens</i> , <i>Raffaella quercivora</i> , <i>Reticulascus tulasneorum</i> , <i>Sarocladium brachiariae</i> , <i>Tolypocladium inflatum</i> , <i>Tolypocladium sp.</i> , <i>Trichoderma harzianum</i> , <i>Trichoderma reesei</i> , <i>Trichoderma sp.</i> , <i>Trichoderma virens</i> , <i>Verticillium albo-atrum</i> , <i>Verticillium alfalfae</i> , <i>Verticillium dahliae</i> , <i>Xylaria grammica</i> , <i>Xylaria hypoxylon</i> , <i>Xylaria sp.</i>
<i>Saccharomycetes</i>
<i>Hypophichia burtonii</i> , <i>H. pseudoburtonii</i> , <i>Komagataella pastoris</i> , <i>K. phaffii</i> , <i>M. citriensis</i> , <i>M. reukaufii</i> , <i>M. reukaufii</i> , <i>M. sp.</i> , <i>Sugiyamaella lignohabitans</i> , <i>Yarrowia lipolytica</i>

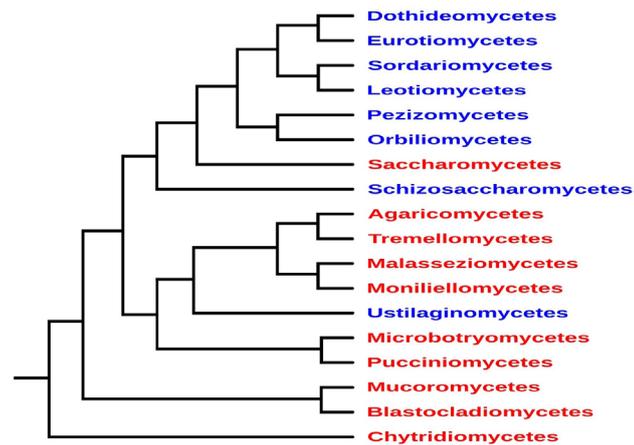


Figure 4: Phylogenetic groups with 5S insertions in the rDNA arrays. Basidiomycota, with the exception of the Ustilaginomycetes, have 5S insertions in the rDNA arrays. Ascomycota, with the exception of the Saccharomycetes, lack 5S insertions in the rDNA arrays.

Table 7: Ascomycota with 5S insertions in the rDNA arrays

<i>Saccharomycetes</i>
<i>Ashbya gossypii</i> , <i>Brettanomyces custersianus</i> , <i>Brettanomyces nanus</i> , <i>Candida albicans</i> , <i>Candida auris</i> , <i>Candida duobushaemulonis</i> , <i>Candida glabrata</i> , <i>Candida metapsilosis</i> , <i>Candida parapsilosis</i> , <i>Candida sake</i> , <i>Candida sojae</i> , <i>Candida tropicalis</i> , <i>Candida vartiovaarae</i> , <i>Candida viswanathii</i> , <i>Clavispora lusitaniae</i> , <i>Eremothecium gossypii</i> , <i>Eremothecium sincaudum</i> , <i>Hanseniaspora uvarum</i> , <i>Kazachstania servazzii</i> , <i>Kluyveromyces lactis</i> , <i>Kluyveromyces marxianus</i> , <i>Kodamaea ohmeri</i> , <i>Lodderomyces elongisporus</i> , <i>Magnusiomyces capitatus</i> , <i>Magnusiomyces ingens</i> , <i>Meyerozyma guilliermondii</i> , <i>Millerozyma farinosa</i> , <i>Ogataea polymorpha</i> , <i>Pichia kluyveri</i> , <i>Pichia kudriavzevii</i> , <i>Pichia manshurica</i> , <i>Saccharomyces cerevisiae</i> , <i>Saccharomyces eubayanus</i> , <i>Saccharomyces paradoxus</i> , <i>Saccharomyces pastorianus</i> , <i>Saccharomyces sp.</i> , <i>Saccharomycopsis fibuligera</i> , <i>Saccharomycopsis malanga</i> , <i>Saprochaete fungicola</i> , <i>Saprochaete ingens</i> , <i>Saprochaete suaveolens</i> , <i>Scheffersomyces stipitis</i> , <i>Torulaspora delbrueckii</i> , <i>Zygosaccharomyces mellis</i> , <i>Zygosaccharomyces rouxii</i> , <i>Zygosaccharomyces sapae</i>

Table 8: Basidiomycota with 5S insertions in the rDNA arrays.

<i>Agaricomycetes</i>
<i>Agaricus bisporus</i> , <i>Agrocybe pediades</i> , <i>Amylostereum areolatum</i> , <i>Armillaria cepistipes</i> , <i>Armillaria gallica</i> , <i>Auricularia comea</i> , <i>Auricularia heimuer</i> , <i>Corioloopsis trogii</i> , <i>Flammulina velutipes</i> , <i>Floccularia luteovirens</i> , <i>Ganoderma australe</i> , <i>Grammothele lineata</i> , <i>Hericium coralloides</i> , <i>Irpex lacteus</i> , <i>Lentinula edodes</i> , <i>Macrocybe gigantea</i> , <i>Phellinus noxius</i> , <i>Pleurotus eryngii</i> , <i>Pleurotus tuoliensis</i> , <i>Polyporus brumalis</i> , <i>Sanghuangporus sanghuang</i> , <i>Sarcomyxa edulis</i> , <i>Serpula lacrymans</i> , <i>Trametes hirsuta</i>
<i>Malasseziomycetes</i>
<i>Malassezia furfur</i> , <i>Malassezia globosa</i> , <i>Malassezia restricta</i> , <i>Malassezia sympodialis</i>
<i>Microbotryomycetes</i>
<i>Microbotryum intermedium</i> , <i>Microbotryum lychnidis-dioicae</i> , <i>Microbotryum saponariae</i> , <i>Microbotryum silenes- acaulis</i> , <i>Microbotryum violaceum</i> , <i>Rhodosporidium toruloides</i> , <i>Rhodotorula graminis</i> , <i>Rhodotorula kratochwilovae</i> , <i>Rhodotorula toruloides</i>
<i>Moniliellomycetes</i>
<i>Moniliella sp.</i>
<i>Pucciniomycetes</i>
<i>Hemileia vastatrix</i> , <i>Melampsora larici-populina</i> , <i>Puccinia graminis</i> , <i>Puccinia hordei</i> , <i>Austropuccinia psidii</i>
<i>Tremellomycetes</i>
<i>Apiotrichum porosum</i> , <i>Cryptococcus gattii</i> , <i>Cryptococcus neoformans</i> , <i>Cryptococcus sp.</i> , <i>Saitozyma podzolica</i>

Table 9: Basidiomycota without 5S insertions in the rDNA arrays.

<i>Ustilaginomycetes</i>
<i>Anthracoystis flocculosa</i> , <i>Moesziomyces antarcticus</i> , <i>Sporisorium scitamineum</i> , <i>Thecaphora thlaspeos</i> , <i>Ustilago bromivora</i> , <i>Ustilago trichophora</i>

Table 10: Mucoromycota with 5S insertions in the rDNA arrays.

<i>Mucoromycetes</i>
<i>Absidia repens</i> , <i>Mucor lusitanicus</i>

Table 11: Blastocladiomycota with 5S insertions in the rDNA arrays.

<i>Blastocladiomycetes</i>
<i>Allomyces macrogynus</i> , <i>Catenaria anguillulae</i>

Table 12: Chytridiomycota with 5S insertions in the rDNA arrays.

<i>Chytridiomycetes</i>
<i>Rhizoclostium globosum</i>

Table 13: 5S gene arrays in the fungal assemblies

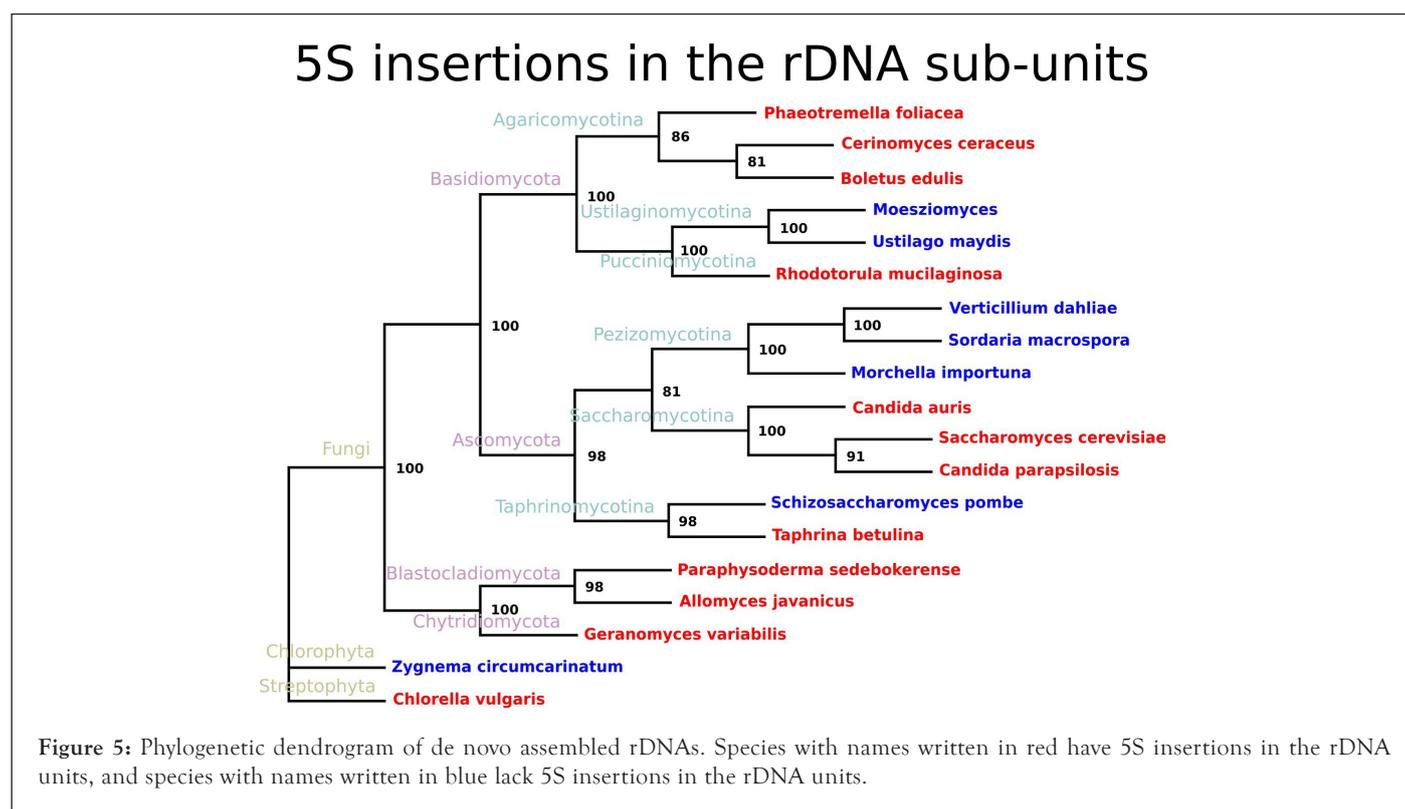
Amount neighboring 5S genes	Amount assemblies
20	1
18	1
15	1
9	2
7	1
6	2
5	2
4	5
3	20
2	88
1	477
0	3

Table 14: Fungal taxa with extended 5S gene arrays.

Species	Phylum	Order	Assembly	size of 5S gene array
<i>Blumeria graminis</i>	Ascomycota	Erysiphales	GCA_900237765	20
<i>Anthracoystis flocculosa</i>	Basidiomycota	Ustilaginales	GCA_900322975	18
<i>Blumeria graminis</i>	Ascomycota	Erysiphales	GCA_900519115	15
<i>Saccharomyces cerevisiae</i>	Ascomycota	Saccharomycetales	GCA_001305865	9
<i>Blumeria graminis</i>	Ascomycota	Erysiphales	GCA_900239735	9
<i>Saccharomyces cerevisiae</i>	Ascomycota	Saccharomycetales	GCA_000773925	7
<i>Morchella sextelata</i>	Ascomycota	Pezizales	GCA_009741755	6
<i>Saccharomyces cerevisiae</i>	Ascomycota	Saccharomycetales	GCA_902192305	6

Table 15: 5S insertions, orientation of the 5S insertions and rDNA copy number per haploid genome of de novo assembled rDNA units.

Taxon	Phylum	Subphylum	Insertion	Repeats
<i>Phaeotremella foliacea</i>	Basidiomycota	Agaricomycotina	forw	60
<i>Cerinomyces ceraceus</i>	Basidiomycota	Agaricomycotina	rev	93
<i>Boletus edulis</i>	Basidiomycota	Agaricomycotina	forw	64
<i>Moesziomyces sp.</i>	Basidiomycota	Ustilaginomycotina	without	51
<i>Ustilago maydis</i>	Basidiomycota	Ustilaginomycotina	without	68
<i>Rhodotorula mucilaginosa</i>	Basidiomycota	Pucciniomycotina	forw	55
<i>Verticillium dahliae</i>	Ascomycota	Pezizomycotina	without	74
<i>Sordaria macrospora</i>	Ascomycota	Pezizomycotina	without	130
<i>Morchella importuna</i>	Ascomycota	Pezizomycotina	without	117
<i>Candida auris</i>	Ascomycota	Saccharomycotina	rev	25
<i>Saccharomyces cerevisiae</i>	Ascomycota	Saccharomycotina	rev	147
<i>Candida parapsilosis</i>	Ascomycota	Saccharomycotina	rev	43
<i>Schizosaccharomyces pombe</i>	Ascomycota	Taphrinomycotina	without	83
<i>Taphrina betulina</i>	Ascomycota	Taphrinomycotina	forw	47
<i>Paraphysoderma sedebokerense</i>	Blastocladiomycota	incertae sedis	forw	66
<i>Allomyces javanicus</i>	Blastocladiomycota	incertae sedis	rev	161
<i>Geranomyces variabilis</i>	Chytridiomycota	incertae sedis	forw	68
<i>Zygnema circumcarinatum</i>	Streptophyta	incertae sedis	without	321
<i>Chlorella vulgaris</i>	Chlorophyta	incertae sedis	rev	38



Congruent with the results of the first part of the analysis, the three Basidiomycota belonging to the Agaricomycotina and the species belonging to the Pucciniomycotina had 5S gene insertions in the rDNA unit, whereas the species from the Ustilaginomycotina did not. Also congruent results were obtained for the Ascomycota (with the exception of *Taphrina betulina*), Blastocladiomycota and Chytridiomycota.

The alga *Z. circumcarinatum*, a member of the higher-plant group Streptophyta, had no 5S insertions. The alga *C. vulgaris* is a member of the Chlorophyta, which is a sister group to the Streptophyta, and has 5S genes in the rDNA units.

The orientation of 5S insertions is not homogeneous. In three of the analysed Basidiomycota, 5S had the same orientation as the 35S gene, but in *Cerinomyces ceraceus*, the orientation was reversed. In taxa of the Saccharomycotina, the 5S gene had reversed orientation. Importantly, the orientation of the 5S gene in the two Blastocladiomycota was reversed. The number of repeats ranged from 25 copies in *Candida auris* to 161 in *Allomyces javanicus*.

Analysing the rDNA assemblies with the sliding window approach revealed that the rDNA unities within a genome were not necessarily identical. In most of the analysed species, polymorphism were present downstream of the 25S region. The polymorphism was distinct in the yeast species *Candida auris*, *C. parapsilosis* and *Saccharomyces cerevisiae*, but missing in *Morchella importuna* and *Taphrina betulina*.

DISCUSSION

rDNA assembly

Assembling fungal genomes to chromosomal quality is still challenging. Whereas the assembly of greater parts of the genome from NGS data into contigs is now routine, a small portion of the genome is often difficult to assemble, which leads to a higher contig number than number of chromosomes. The rDNA segment is an intricate part of the genome with highly repetitive sequences. The number of rDNA repeated unit's ranges from 14 to 1442 in fungal species (Lofgren et al.) [16], but only approximately 15% of the analysed assemblies include a complete rDNA unit, and less than 9% include an rDNA array with at least three repeats. In addition to the known obstacles in assembling highly repetitive sequences, rDNA units can have repetitive structures within themselves, regions that are variable between them (James et al. 2009) and intergenic spacers with low complexity. To overcome these challenges, we wrote a program called YNGS allowing the assembly of paired-end fastq data. The assembly starts from a seed and increases the string loop by loop. In the case of an interruption, the fastq fragments, including the paired-end fragments, of the region can be displayed, and the string can be elongated manually. After finishing the string, the sequence is verified by the sliding window approach.

rDNA structure throughout the living organism

Most living organisms have multiple copies of rDNA operons. The copy number in bacterial species commonly ranges between 1 and 15, where more than 80% of the species have more than one copy. The 5S, 16S, and 23S rDNA genes, homologous to 5S, 18S and 25S rDNA in eukaryotes, are organized in gene clusters (Espejo and Plaza) [17]. In archaea, the copy number is generally lower and ranges from 1-4 copies, and the 5S gene is not always part of the operon and can be transcribed separately [18]. As in eukaryotes, both structural variations, in which 5S rDNA is positioned within or outside the

rDNA repeat unit, are observed in archaea. However, the rDNA structure of archaea shares more similarity with the rDNA structure of eukaryotes.

rDNA repeats without 5S insertions are common in eukaryotic organisms. Beyond the numerous species in the fungal kingdom, only a few species have been reported to have 5S insertions in their rDNA arrays (Table 16).

Table 16: Records of 5S insertions in the rDNA arrays of eukaryotic taxa outside the fungal kingdom.

Taxa	Systematic	Reference
<i>Toxoplasma gondii</i>	Alveolata; Apicomplexa;	Guay15 et al. [42]
<i>Dictyostelium discoideum</i>	Amoebozoa; Dictyostelia;	Hofmann et al. 1993; Maizels et al. 1976;
<i>Meloidogyne arenaria</i>	Metazoa; Nematoda;	Vahidi et al. 1988; Vahidi et al. 1991;
<i>Calanus finmarchicus</i>	Metazoa; Arthropoda; Crustacea;	Drouin et al. [43]
<i>Temora longicornis</i>	Metazoa; Arthropoda; Crustacea;	Drouin et al. [43]
<i>Semibalanus balanoides</i>	Metazoa; Arthropoda; Crustacea;	Drouin et al. [43]
<i>Thysanoessa raschii</i>	Metazoa; Arthropoda; Crustacea;	Drouin et al. [43]
<i>Araneus</i>	Metazoa; Arthropoda; Arachnida;	Drouin et al. [43]
<i>Chlorella vulgaris</i>	Chlorophyta; Trebouxiophyceae;	this study;
<i>Funaria hygrometrica</i>	Streptophyta; Bryophyta;	Capesius 1997;
<i>Ginkgo biloba</i>	Streptophyta; Spermatophyta;	Galián 2012;
<i>several cryptogams</i>	Streptophyta and Chlorophyta	Wicke et al. 2011
<i>Asteraceae</i>	Streptophyta; Spermatophyta;	García et al. 2010;

Strikingly, the rearrangement of 5S rDNA occurred several times in the evolution of eukaryotes. By comparing phylogenetic data with rearrangement data of rDNA genes, it is possible to reveal the evolution of rDNA. In land plants, 5S rDNA popped out of the rDNA array at least twice (Wicke et al.) [19]. In early-diverging land plants, rRNA genes were co-localized within the repeated units, whereas most basal seed plants and the distinct group of water ferns had separated 5S genes. Additionally, the opposite rearrangement occurred in the evolution of plants. In the modern family Asteraceae, nearly 25% of the species had a linked arrangement of rRNA genes. These species belong to the tribe Anthemideae, tribe Gnaphalieae and "Heliantheae alliance", whereas in the other five tribes, only 5S genes outside the rDNA array were detected (García et al.) [20]. The evolution in animals is similar, and 5S rDNA linkages were repeatedly established and lost during the evolution of eukaryotic genomes (Drouin and de Sá) [21].

rDNA structure throughout the fungal kingdom

In fungi, the 18S, 5.8S and 25S rDNA genes are organized in tandem repeats. Tandem repeat arrays can be organized in the genome in one string or separated in a few arrays. Two kinds of arrays can be discriminated. In the first case, each 5S gene is integrated into rDNA units in the intergenic spacer region. In the second case, these insertions are missing, and the 5S genes are scattered within the genome. Then, the 5S genes are typically separated from each other and not arranged in 5S tandem arrays, as is the case in most genomes of higher plants and animals. Most Ascomycota lack 5S insertions in the rDNA repeats, with the exception of Saccharomycetes. Most Basidiomycota have 5S insertions, with the exception of Ustilaginomycetes. For Mucoromycota, Blastocladiomycota and Chytridiomycota, data availability is limited, but it seems that 5S insertions are common in these three phyla. Our analysis and that of Bergeron and Drouin [22] revealed only a very few exceptions to the rules. Some of the exceptions are supposedly artefacts. In particular, when all closely related species are homogeneous and fit the rules, aberrant data should be verified by other methods. In a few groups of Saccharomycetes, aberrant data occur in clusters, which indicate that they are not caused by artefacts.

Ascomycota: The 5S rDNA distribution in Ascomycota fungi can be divided into two groups. The first group included all Ascomycota fungi, without Saccharomycetes, and the second group included Saccharomycetes. For the first group, Ascomycota fungi without Saccharomycetes, Bergeron and Drouin reported two exceptions, *Fusarium solani* and *Pyrenophora graminea*, in addition to *Cercospora sojina* and *Taphrina betulina* in this study. The few exceptions should be treated with caution, since the results are not stringent and were not verified by a second method. Additionally, we could not confirm the insertion of 5S in the genera *Fusarium* and *Pyrenophora*, and a total of 21 *Fusarium* and 11 *Pyrenophora* assemblies were analysed in our study.

In non-Saccharomycetes taxa, the 5S rDNA genes were scattered in the genome and are mostly separated from each other. Neighbouring 5S genes were detected very seldomly, whereas the largest array was detected in the genome of *B. graminis*. In our study, four assemblies of *B. graminis* were integrated, and in each assembly, 5S gene arrays were detected. *B. graminis* has an ~120 Mb genome, which is unusually large, e.g., nearly 10 times larger than that of *S. cerevisiae*. The genome consists of 64% transposable elements, and the dysfunctionality of the RIP pathway has probably contributed to genome-size inflation and to the increase in highly repeated sequences (Spanu et al.) [22,23].

For the Saccharomycetes, 5S insertions in the rDNA units are typical, whereas for the other taxa, 5S insertions are missing. *Yarrowia lipolytica* within Saccharomycetes is an exception because it lacks a 5S insertion (Bergeron and Drouin) [4,22]. In our study, we detected seven other Saccharomycetes species without 5S insertions: *Hyphopichia burtonii*, *Hyphopichia pseudoburtonii*, *Komagataella pastoris*, *Komagataella phaffii*, *Metschnikowia citriensis*, *Metschnikowia reukauffii* and *Sugiyamaella lignohabitans*.

Basidiomycota: In the Basidiomycota group, 5S insertions in the rDNA repeats were typical, whereas in the Ustilaginomycetes group, the insertions were missing. Only eight assemblies in the non-Ustilaginomycetes were aberrant, without 5S insertion. Three assemblies in the family Sparassidaceae were aberrant. Bergeron and Drouin reported the loss of 5S insertions in one of two *Filobasidiella*

species. Whereas all analysed Ustilaginomycetes lacked 5S insertions, in other groups of Ustilaginomycotina, 5S insertions were present. Additionally, three *Tilletia* spp., belonging to the Exobasidiomycetes within the Ustilaginomycotina, were reported to have 5S insertions in rDNA arrays (Bergeron and Drouin) [22].

Blastocladiomycota: In the group of Blastocladiomycota, all analysed taxa had 5S insertions. We analysed the data of two

genome assemblies and two additional assembled rDNA repeats from NGS data.

Chytridiomycota: We analysed two taxa within the Chytridiomycota, one genome assembly and one de novo assembly, and in both cases, 5S insertions were present. In contrast, Bergeron and Drouin reported that *Batrachochytrium dendrobatidis* lacks 5S insertions [22].

Microsporidia: The rDNA organization of Microsporidia deviates from that of other fungi. The 18S and 25S genes are unusually short. The 5.8S rDNA sequences are missing in the ITS between the 18S and 25S genes; instead, in *Nosema bombycis*, a sequence homologous to the 5.8S gene was detected in the 25S gene. Stringently, in *N. bombycis*, 25S is located on the reverse DNA strand of the 18S gene (Huang et al.) [24]. Insertions of 5S genes in the rDNA repeat units were detected in three *Nosema* spp. (Dong et al.) [25,26]. The rDNA genes were repeated in all the genomes analysed so far, and the existence of multiple rDNA arrays was documented for *Nosema apis*. In *Nosema ceranae*, significant variations between the rDNA units within a genome were reported, which leads to the conclusion that concerted evolution and homogenization of the rDNA is only rudimentary (Sagastume et al.) [27].

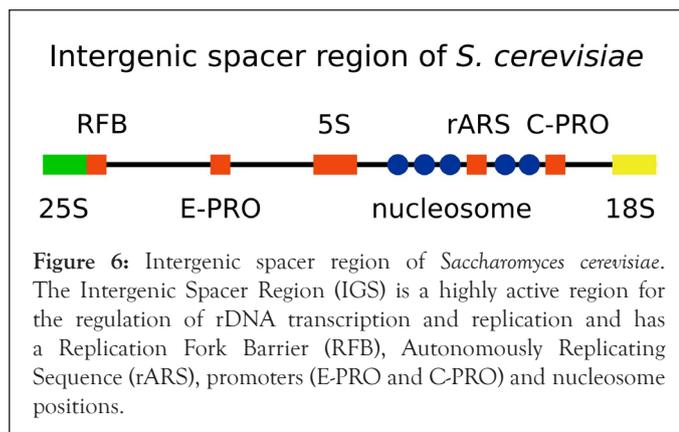
Mucoromycota: For the Mucoromycota group, 5S insertions were identified. In our study, two assemblies were integrated, and in the study of Bergeron and Drouin [22,28], four taxa were studied.

Evolution of rDNA

Whereas the molecular mechanism of 5S gene rearrangement is still unknown, the knowledge about rearrangement of complete rDNA units and copy number variations is much more advanced. The rDNA copy number is strikingly variable in eukaryotes, with 39–19 300 copies in higher animals and 150–26 048 copies in plants. The rDNA copy number in fungi, calculated among 91 fungal taxa, ranged from 14 to 1442. Interestingly, the copy number within a species can be highly variable. For example, among 12 different isolates of *Suillus brevipes*, the copy number ranged from 72 to 156 (Lofgren et al.). One factor that influences the actual copy number is the environment. The chemical lithium acetate, which is routinely used in the gene transformation of yeast, can change the copy number of rDNA in *S. cerevisiae* (Kwan et al.) [29]. Salim et al. reported that yeast cells with a reduced number of rDNA repeats grew better than their counterparts with normal copy numbers under conditions of DNA replication stress (Salim et al.) [30]. Another important factor influencing the copy number is genetics. A mutation in the *orc4* gene, resulting in a delay of replication initiation and consequently in a prolonged S-phase, severely reduced rDNA copy number. Although reducing rDNA copy number may help ensure complete chromosome replication, *orc* mutant cells struggle to meet the high demand for ribosomal RNA synthesis (Sanchez et al.) [31].

Despite the fact that the number of rDNA repeats varies, cells strive to keep the number of rDNA repeats in a favorable range. For *S. cerevisiae*, the average copy number is approximately 150 (Kobayashi et al.). Homologous recombination between the repeated rDNA

units led to a reduction in the copy number (Iida and Kobayashi) [32]. The loss of rDNA copies in *S. cerevisiae* was counteracted by gene amplification, which increased the copy number (Iida and Kobayashi) [33]. The replication fork barrier (RFB) sequence downstream of the 25S gene is important for the reduction and the increase in the copy number (Figure 6). Knocking out the Fob1 protein, which binds to the rDNA replication fork barrier site, disabled a change in rDNA copy number (Kobayashi et al.) [34]. Although knowledge about the mechanism underlying changes in the rDNA copy number is already substantial and sophisticated models exist (Kobayashi, Egidio et al.) [35,36] the mechanism of long-term evolution is still unclear. In the long term, the entire rDNA array could move from one position in the genome to another and even jump to other chromosomes, as detected within members of the family *Saccharomycetaceae* (Proux-Wéra et al.) [37].



5S rDNA is a mobile element that has changed position and orientation several times during evolution. As a result, no continuous lineage from the early evolved eukaryote to modern species exists. On the other hand, we see large groups where the 5S position is constant. The molecular mechanisms of insertion, deletion and inversion of 5S rDNA remain unclear, as do their evolutionary consequences. With the deletion of 5S from the main rDNA array, a reduction in the 5S copy number is normally detected. In species with 5S linked to the rDNA array, the numbers of 5S insertions and rDNA units are equal. In species with separated 5S insertions, the copy number is normally lower, e.g., in *S. pombe*, with 6 copies. The copies are separated and variable, allowing independent evolution of the 5S sequences. Among animals, including some amphibians, bony fishes and sea urchins, 5S rRNA genes from different loci are expressed in oocytes and somatic cells (Komiya and Bellavia et al.) [38,39].

Molecular mechanism

The exact molecular mechanism underlying the switch from 5S rDNA insertion to deletion and vice versa is unknown. In principle, after insertion or deletion of the 5S gene in one rDNA unit, the new rDNA variation can spread by rDNA rearrangement and homogenization. From the model *S. cerevisiae*, we learned about the molecular mechanism of rDNA rearrangement, maintaining the rDNA copy number and rDNA homogenization. The rDNA region is highly active, and the intergenic spacer region. Figure 6 performs the main regulatory role in the processes. All three RNA polymerases and the replication machinery are recruited here, and collision of the processes needs to be avoided [40,41].

The replication of the rDNA starts at the autonomously replicating sequence (ARS), and in the beginning, it is bidirectional. In one direction, towards the 5S gene, replication stops at the next RFB, and in the other direction, towards the 35S precursor, it continues until it meets a stalled RFB [42,43].

CONCLUSION

A stalled fork is frequently converted into a double-strand break (DSB) and repaired by homologous recombination. Due to the repetitive nature of the rDNA sequence, homologous recombination between different rDNA units could occur and could lead to copy number variations and consecutive homogenization of the rDNA array. In *S. cerevisiae*, as well as in *C. auris* and *C. parapsilosis*, the rDNA is only partially homogenized. A small segment in the intergenic spacer region displays high variability within the array. By comparison of the IGS between different *S. cerevisiae* strains, high variations were found not only among different strains but also within a strain (James et al.). By cloning and subsequent sequencing of the intergenic spacer, variations within a genotype were also detected in the basidiomycete *Puccinia striiformis*. The phenomenon in which nearly all of the rDNA gene but not the intergenic spacer region is more or less homogeneous within a strain seems to be common in many fungal species.

REFERENCES

1. Potapova TA, Gerton JL. Ribosomal DNA and the nucleolus in the context of genome organization. *Chromosome Res.* 2019;27(1):109-127.
2. Prokopowich CD, Gregory TR, Crease TJ. The correlation between rDNA copy number and genome size in eukaryotes. *Genome.* 2003;46(1):48-50.
3. Long EO, Dawid IB. Repeated genes in eukaryotes. *Annu Rev Biochem.* 1980;49(1):727-764.
4. Biscotti MA, Olmo E, Heslop-Harrison JS. Repetitive DNA in eukaryotic genomes. *Chromosome Res.* 2015;23(3):415-420.
5. Treangen TJ, Salzberg SL. Repetitive DNA and next-generation sequencing: computational challenges and solutions. *Nat Rev Genet.* 2012;13(1):36-46.
6. Agrawal S, Ganley AR. Complete sequence construction of the highly repetitive ribosomal RNA gene repeats in eukaryotes using whole genome sequence data. *Methods Mol Biol.* 2016;161-181
7. Christie KR, Weng S, Balakrishnan R, Costanzo MC, Dolinski K, Dwight SS, et al. *Saccharomyces Genome Database (SGD)* provides tools to identify and analyze sequences from *Saccharomyces cerevisiae* and related sequences from other organisms. *Nucleic acids res.* 2004;32:311-314.
8. Jones T, Federspiel NA, Chibana H, Dungan J, Kalman S, Magee BB, et al. The diploid genome sequence of *Candida albicans*. *Proc Natl Acad Sci U S A.* 2004;101(19):7329-7334.
9. Skrzypek MS, Binkley J, Binkley G, Miyasato SR, Simison M, Sherlock G. The *Candida Genome Database (CGD)*: incorporation of Assembly 22, systematic identifiers and visualization of high throughput sequencing data. *Nucleic Acids Res.* 2016.
10. Lock A, Rutherford K, Harris MA, Hayles J, Oliver SG, Bähler J, Wood V. *PomBase 2018: user-driven reimplementations of the fission yeast database provides rapid and intuitive access to diverse, interconnected information.* *Nucleic Acids Res.* 2019;821-827.
11. Mao J, Appel B, Schaack J, Sharp S, Yamada H, Söll D. The 5S rRNA genes of *Schizosaccharomyces pombe*. *Nucleic Acids Res.* 1982;10(2):487-500.

12. Lockington RA, Taylor GG, Winther M, Sczzocchio C, Davies RW. A physical map of the ribosomal DNA repeat unit of *Aspergillus nidulans*. *Gene*. 1982;20(2):135-137.
13. Cerqueira GC, Arnaud MB, Inglis DO, Skrzypek MS, Binkley G, Simison M, et al. The *Aspergillus* Genome Database: multispecies curation and incorporation of RNA-Seq data to improve structural gene annotations. *Nucleic Acids Res*. 2014;705-710.
14. Bartnik E, Bartoszewski S, Borsuk P, Empel J. *Aspergillus nidulans* 5S rRNA genes and pseudogenes. *Curr Genet*. 1986;10(6):453-457.
15. Lagesen K, Hallin P, Rødland EA, Stærfeldt HH, Rognes T, Ussery DW. RNAMmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic acids res*. 2007;35(9):3100-108.
16. Lofgren LA, Uehling JK, Branco S, Bruns TD, Martin F, Kennedy PG. Genome-based estimates of fungal rDNA copy number variation across phylogenetic scales and ecological lifestyles. *Mol Ecol*. 2019;28(4):721-730.
17. Espejo RT, Plaza N. Multiple ribosomal RNA operons in bacteria; their concerted evolution and potential consequences on the rate of evolution of their 16S rRNA. *Front Microbiol*. 2018;9:1232.
18. Garrett RA, Daigaard J, Larsen N, Kjems J, Mankin AS. Archaeal rRNA operons. *Trends Biochem Sci*. 1991;16:22-6.
19. Wicke S, Costa A, Muñoz J, Quandt D. Restless 5S: the rearrangement (s) and evolution of the nuclear ribosomal DNA in land plants. *Mol Phylogenet Evol*. 2011;61(2):321-332.
20. Garcia S, Panero JL, Siroky J, Kovarik A. Repeated reunions and splits feature the highly dynamic evolution of 5S and 35S ribosomal RNA genes (rDNA) in the Asteraceae family. *BMC Plant Biol*. 2010;10(1):1-18.
21. Drouin G, De Sa MM. The concerted evolution of 5S ribosomal genes linked to the repeat units of other multigene families. *Mol Biol Evol*. 1995;12(3):481-493.
22. Bergeron J, Drouin G. The evolution of 5S ribosomal RNA genes linked to the rDNA units of fungal species. *Curr Genet*. 2008;54(3):123-131.
23. Spanu PD, Abbott JC, Amselem J, Burgis TA, Soanes DM, Stüber K, et al. Genome expansion and gene loss in powdery mildew fungi reveal tradeoffs in extreme parasitism. *Science*. 2010;330(6010):1543-1546.
24. Huang WF, Tsai SJ, Lo CF, Soichi Y, Wang CH. The novel organization and complete sequence of the ribosomal RNA gene of *Nosema bombycis*. *Fungal Genet Biol*. 2004;41(5):473-481.
25. Dong S, Shen Z, Zhu F, Tang X, Xu L. Complete sequence and gene organization of the *Nosema heliothidis* ribosomal RNA gene region. *J Eukaryot Microbiol*. 2011;58(6):539-541.
26. Huang WF, Tsai SJ, Lo CF, Soichi Y, Wang CH. The novel organization and complete sequence of the ribosomal RNA gene of *Nosema bombycis*. *Fungal Genet Biol*. 2004;41(5):473-481.
27. Sagastume S, Del Águila C, Martín-Hernández R, Higes M, Henriques-Gil N. Polymorphism and recombination for rDNA in the putatively asexual microsporidian *Nosema ceranae*, a pathogen of honeybees. *Environ Microbiol*. 2011;13(1):84-95.
28. Drouin G, De Sa MM. The concerted evolution of 5S ribosomal genes linked to the repeat units of other multigene families. *Mol Biol Evol*. 1995;12(3):481-493.
29. Kwan EX, Wang XS, Amemiya HM, Brewer BJ, Raghuraman MK. rDNA copy number variants are frequent passenger mutations in *Saccharomyces cerevisiae* deletion collections and de novo transformants. *G3: Genes, Genomes, Genetics*. 2016;6(9):2829-38.
30. Salim D, Bradford WD, Freeland A, Cady G, Wang J, Pruitt SC, et al. DNA replication stress restricts ribosomal DNA copy number. *PLoS Genet*. 2017.
31. Sanchez JC, Kwan EX, Pohl TJ, Amemiya HM, Raghuraman MK, Brewer BJ. Defective replication initiation results in locus specific chromosome breakage and a ribosomal RNA deficiency in yeast. *PLoS Genet*. 2014.
32. Iida T, Kobayashi T. RNA polymerase I activators count and adjust ribosomal RNA gene copy number. *Mol Cell*. 2019;73(4):645-654.
33. Iida T, Kobayashi T. How do cells count multi-copy genes?: “Musical Chair” model for preserving the number of rDNA copies. *Curr Genet*. 2019;65(4):883-885.
34. Kobayashi T, Heck DJ, Nomura M, Horiuchi T. Expansion and contraction of ribosomal DNA repeats in *Saccharomyces cerevisiae*: requirement of replication fork blocking (Fob1) protein and the role of RNA polymerase I. *Genes Dev*. 1998;12(24):3821-3830.
35. Kobayashi T. Regulation of ribosomal RNA gene copy number and its role in modulating genome integrity and evolutionary adaptability in yeast. *Cell Mol Life Sci*. 2011;68(8):1395-1403.
36. Egidi A, Di Felice F, Camilloni G. *Saccharomyces cerevisiae* rDNA as super-hub: the region where replication, transcription and recombination meet. *Cell Mol Life Sci*. 2020;77(23):4787-498.
37. Proux-Wera E, Byrne KP, Wolfe KH. Evolutionary mobility of the ribosomal DNA array in yeasts. *Genome Biol Evol*. 2013;5(3):525-531.
38. Komiya H, Hasegawa M, Takemura S. Differentiation of oocyte- and somatic-type 5S rRNAs in animals. *J Biochem*. 1986;100(2):369-374.
39. Bellavia D, Dimarco E, Caradonna F. Characterization of three different clusters of 18S-26S ribosomal DNA genes in the sea urchin *P. lividus*: Genetic and epigenetic regulation synchronous to 5S rDNA. *Gene*. 2016;580(2):118-124.
40. James SA, O’Kelly MJ, Carter DM, Davey RP, van Oudenaarden A, Roberts IN. Repetitive sequence variation and dynamics in the ribosomal DNA array of *Saccharomyces cerevisiae* as revealed by whole-genome resequencing. *Genome Res*. 2009;19(4):626-635.
41. Wang Y, Hao B, Zhang Q, Tuo E, Sun G, Zhang R, et al. Discovery of multiple IGS haplotypes within genotypes of *Puccinia striiformis*. *Fun Biol*. 2012;116(4):522-528.
42. Guay JM, Huot A, Gagnon S, Tremblay A, Levesque RC. Physical and genetic mapping of cloned ribosomal DNA from *Toxoplasma gondii*: primary and secondary structure of the 5S gene. *Gene*. 1992;114(2):165-171.
43. Drouin G, De Sa MM. The concerted evolution of 5S ribosomal genes linked to the repeat units of other multigene families. *Mol Biol Evol*. 1995;12(3):481-493.