

The Ribosomal Operon Database (ROD): A full-length rDNA operon reference database

Anders K. Krabberød, Embla E. Stokke, Ella Thoen, Inger Skrede & Håvard Kauserud

a.k.krabberod@ibv.uio.no

Department of Biosciences & Department of Informatics

Oslo Mycology Group, Norwegian Sequencing Centre

University of Oslo, Norway



UNIVERSITY
OF OSLO

MOLECULAR ECOLOGY RESOURCES

RESOURCE ARTICLE |  Open Access | 

The Ribosomal Operon Database: A Full-Length rDNA Operon Database Derived From Genome Assemblies

Anders K. Krabberød , Embla Stokke, Ella Thoen, Inger Skrede, Håvard Kauserud

First published: 21 October 2024 | <https://doi.org/10.1111/1755-0998.14031>


Handling Editor: Kin-Ming (Clement) Tsui

Funding: The authors received no specific funding for this work.

 SECTIONS

 PDF

 TOOLS

 SHARE

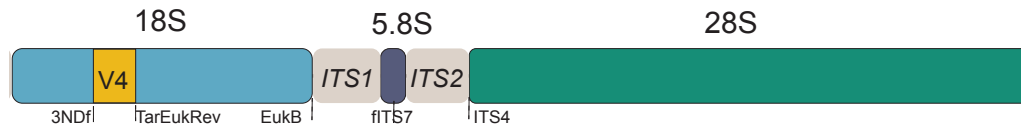


UNIVERSITY
OF OSLO

<https://github.com/krabberod/ROD>

Background and motivation

- **Why a database of the entire rDNA operon?**
- Recent advancements in sequencing technologies enable the sequencing of amplicons of several thousand base pairs.
 - A move from short reads (4-500bp) to long reads (5-10000 bp).
 - From V4 (or V9,ITS...) to full operon



- Longer reads provide better phylogenetic and taxonomic resolution.



The family of ONT



PacBio Revio



UNIVERSITY
OF OSLO

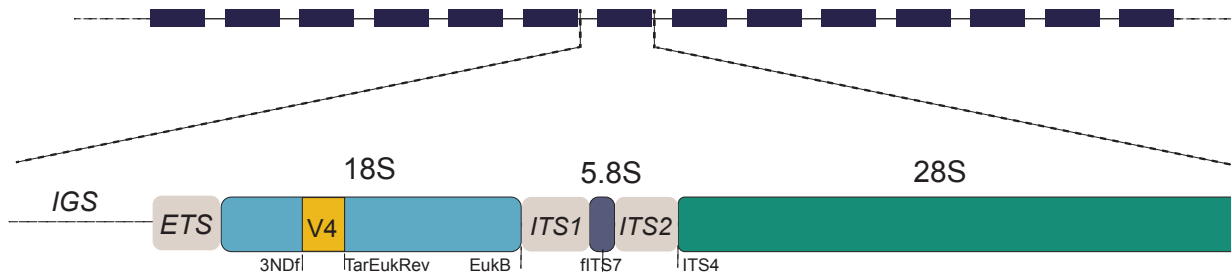
Background and motivation

- **BUT:** Most databases are tailored for short-read technology (e.g., Illumina)
- Limitations of Existing Databases
 - **GenBank (NCBI):** full of errors and incomplete sequences.
 - **BOLD:** Focuses on COI
 - **PR2:** Primarily covers 18S, very good for protists.
 - **Silva:** Contains 18S and 28S separately, not as a continuous segment.
 - First attempt was to concatenate 18S and 28S from Silva, with very poor result
 - No ITS



The Ribosomal Operon Database

- Enter ROD!
 - The Ribosomal Operon Database.
 - Full-length ribosomal operon from all available eukaryotic genomes in NCBI.



Eukaryotic Genomes from NCBI
(35,000)

Scanned with HMMER3

Confirmed 18S and 28S
(continuously) with Barnap

Removed contaminants

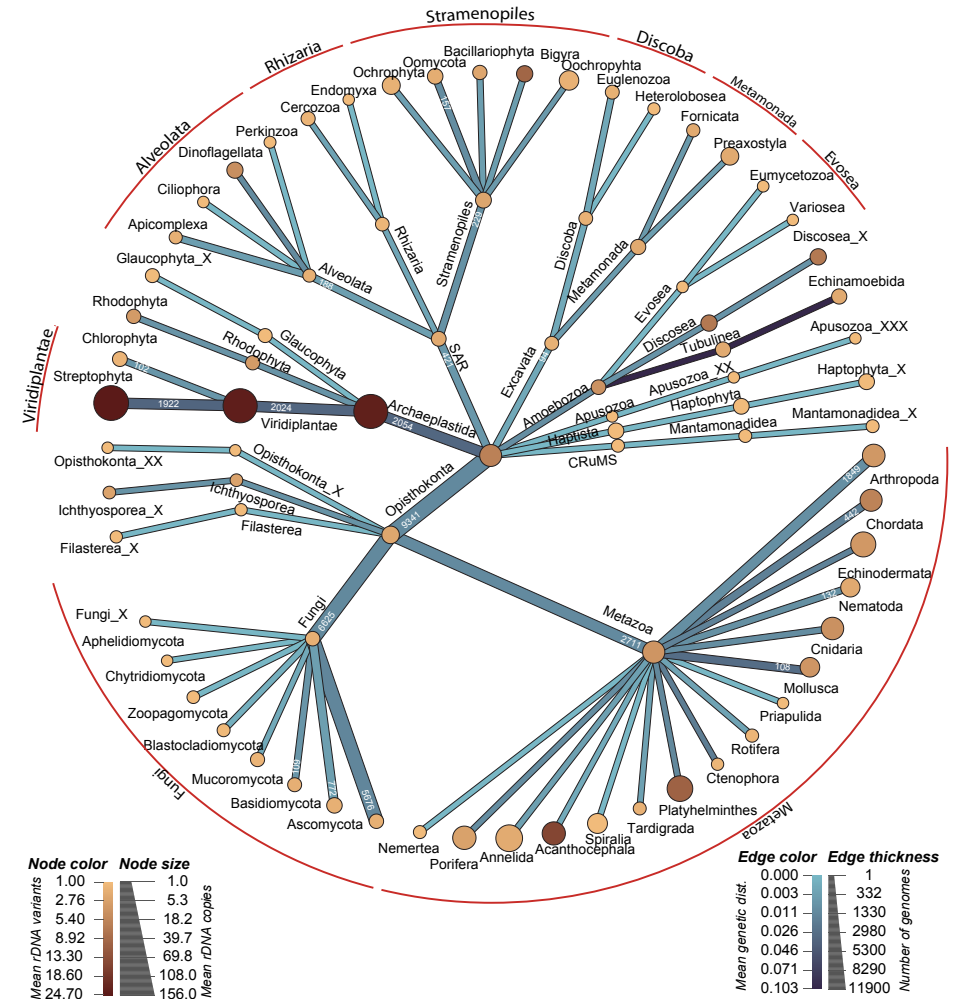
Clustered Intragenomically at
99%

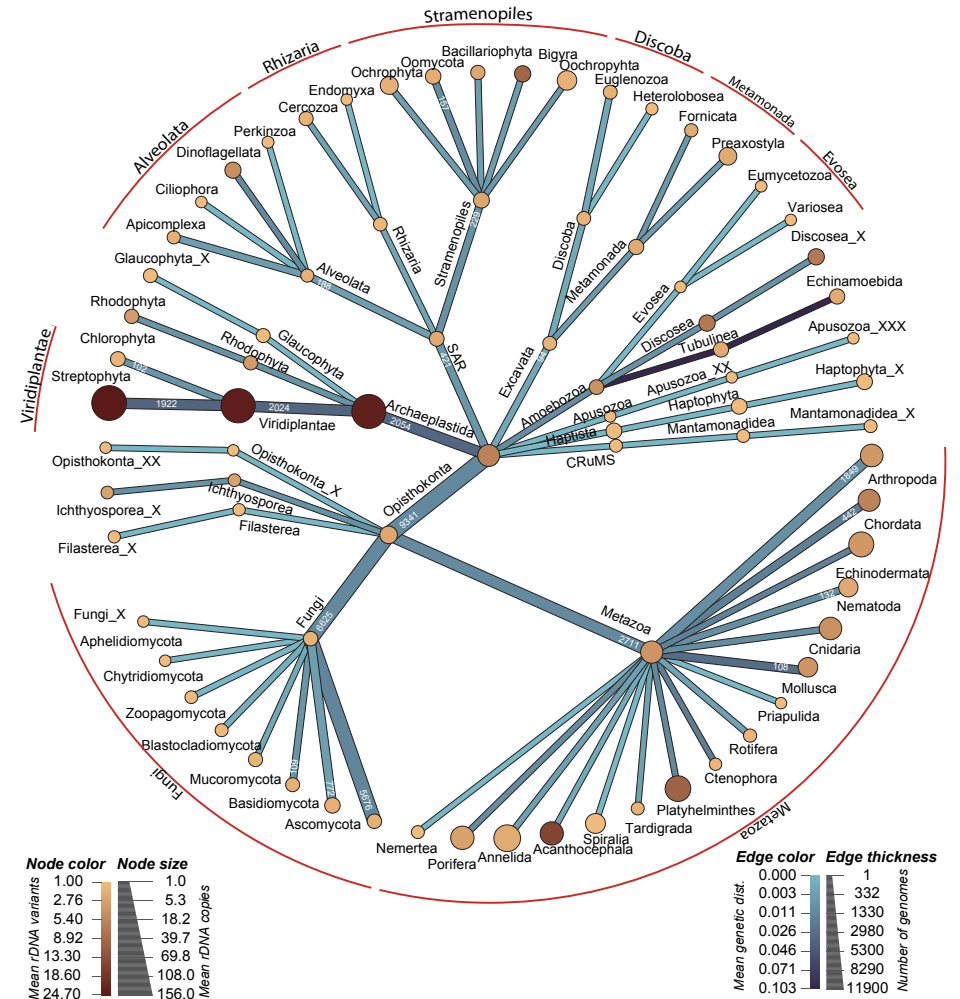
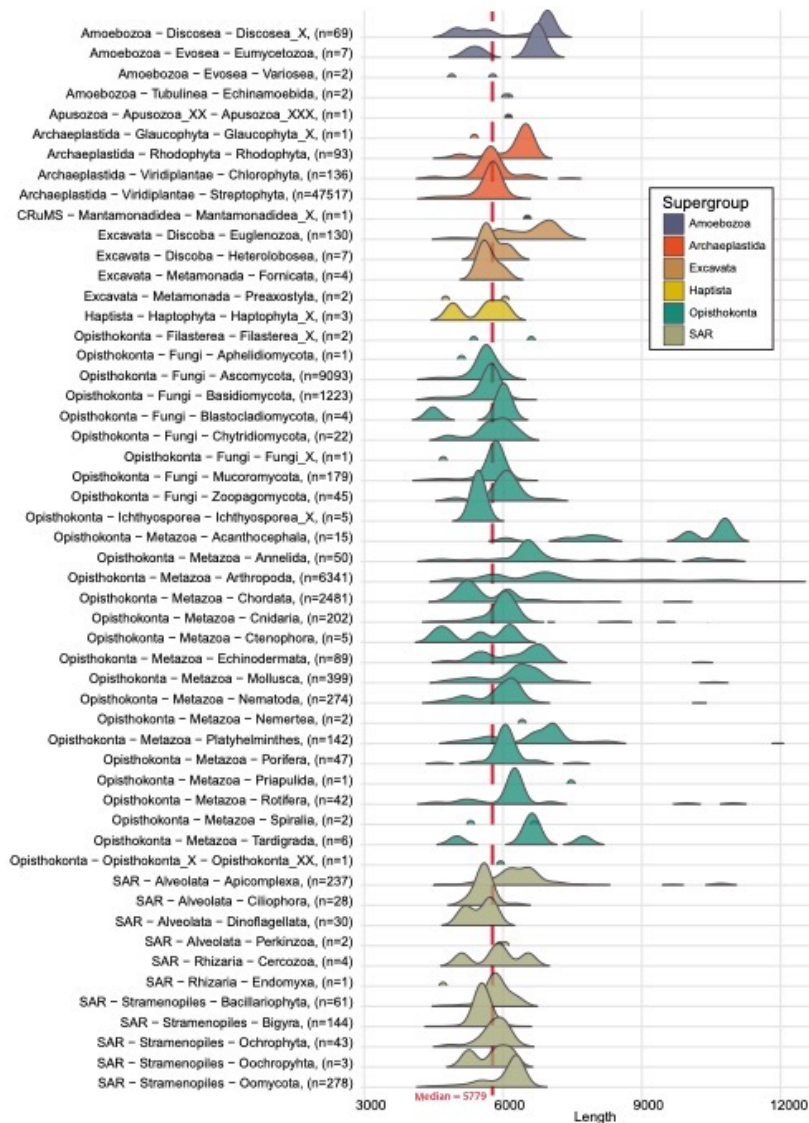


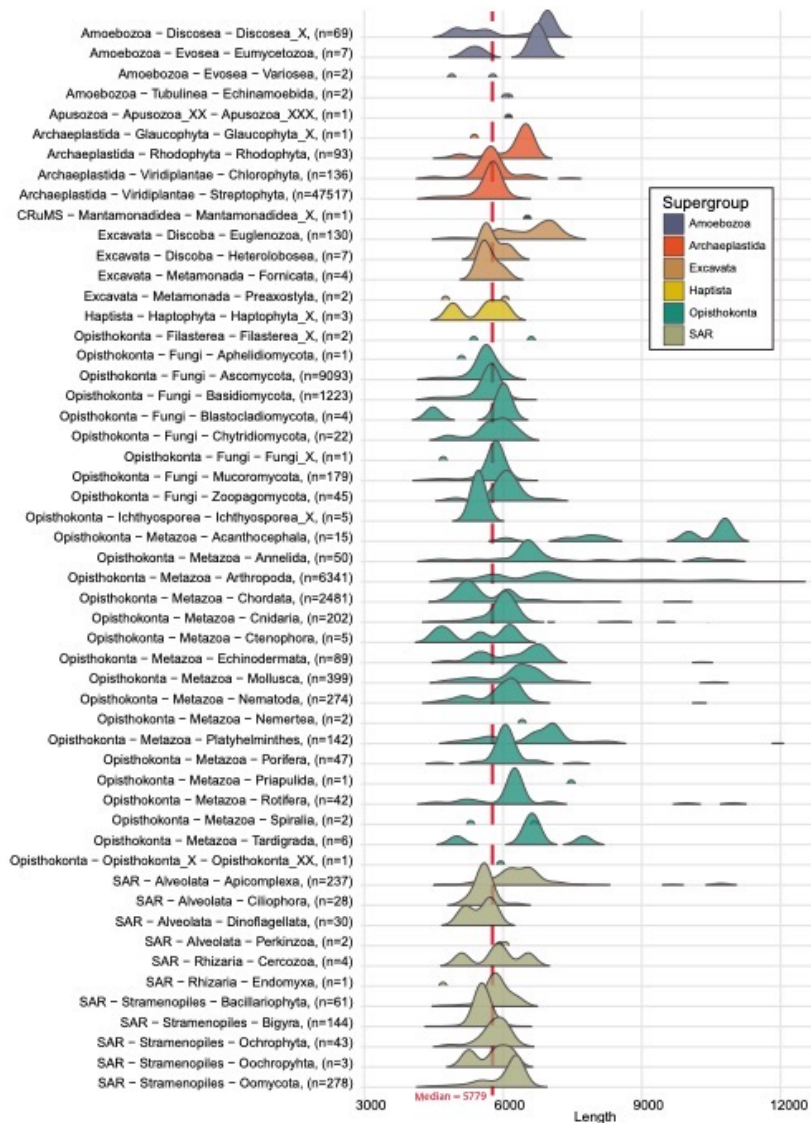
UNIVERSITY
OF OSLO

ROD v1.2 – Golden ROD

- 11,935 genomes
- 453,697 non-clustered operons
- 69,480 clustered operons
- Taxonomic
 - Supergroup : 8
 - Divisions: 19
 - Subdivisions: 54
 - Class: 151
 - Order: 422
 - Genus: 2446
 - Species: 5280
- Genomes with only 1 copy: 5,601
- Genomes > 100 copies: 708
- Max copy number: 5,947
 - Streptophyta; Zea mais







- Median length: 5,779
- Max length: 16,463
 - Insecta; Bombylius major



The Ribosomal Operon Database

- Plans for the future:
- Integration in PR2 (the work has started)
- Update with more data as more genomes are sequenced.
- Look for genomes currently not in NCBI



UNIVERSITY
OF OSLO

Acknowledgment



Håvard Kausrud



Inger Skrede



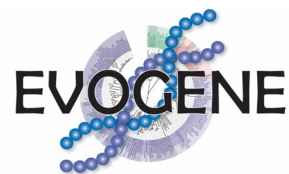
Oslo Mycology group



Ella Thoen



Embla Stokke



UNIVERSITY
OF OSLO

MOLECULAR ECOLOGY RESOURCES

RESOURCE ARTICLE | [Open Access](#) |

The Ribosomal Operon Database: A Full-Length rDNA Operon Database Derived From Genome Assemblies

Anders K. Krabberød , Embla Stokke, Ella Thoen, Inger Skrede, Håvard Kausrud

First published: 21 October 2024 | <https://doi.org/10.1111/1755-0998.14031>

Handling Editor: Kin-Ming (Clement) Tsui

Funding: The authors received no specific funding for this work.

SECTIONS

PDF TOOLS SHARE

<https://github.com/krabberod/ROD>

a.k.krabberod@ibv.uio.no