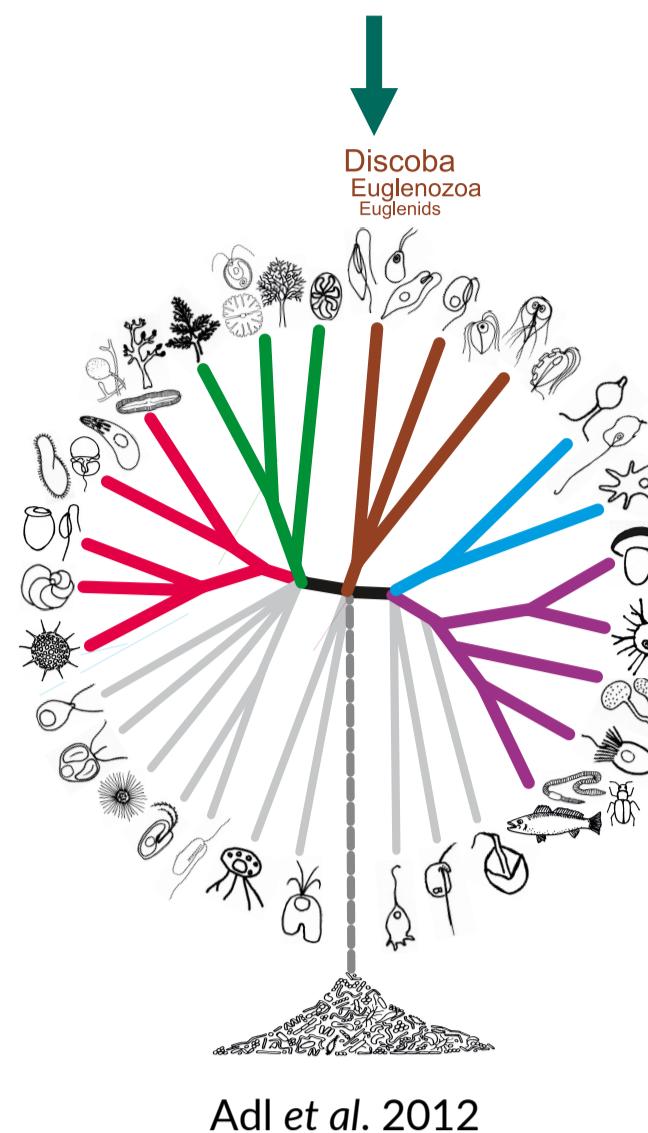


Exception within exceptions

- unusual organization of rRNA genes in *Euglena longa*

Paweł Hałakuc, Natalia Gumińska, Anna Karnkowska, Rafał Milanowski

Department of Molecular Phylogenetics and Evolution, Faculty of Biology, Biological and Chemical Research Centre, University of Warsaw, Żwirki i Wigury 101, 02-089 Warszawa
pawelhalakuc@biol.uw.edu.pl



INTRODUCTION

In almost all eukaryotes three ribosomal RNAs (18S, 5.8S and 28S) are encoded within single operon – rDNA. In Euglenids only the model species *Euglena gracilis* was thoroughly investigated. All detected rDNA copies were localized on 11.5 kb extrachromosomal circular molecules, each of them containing single rDNA operon.

Our analyses of short read whole genome sequencing data support this picture in majority of euglenids. However, in secondary osmotrophic *Euglena longa*, sister species of *E. gracilis*, structure of rDNA could not be resolved using short read data.

Analysis of assembly graphs allowed to propose two hypothetical circular structures (Figure 1):

- one large circle analogical to that of *E. gracilis* with repeated region in ITS-1 and IGS,
- two separate circles, containing SSU and LSU respectively (both with repeated region).

OBJECTIVE

Verification of the hypothesis assuming the presence of three different molecules encoding rDNA in the *E. longa* genome.

RESULTS

From 10 052 267 analyzed PacBio reads, 22 050 contained fragments corresponding to used queries. Specific considered variants and number of corresponding reads is shown in Table 1. In all cases specific order was required, e.g. in the SSU-repeat-LSU variant following hits needed to be present in order and on concordant strand: SSU-end, repeat and LSU-beginning. If only single type of hit was found in read, it was classified as singular fragment. In case of unexpected arrangement (e.g. SSU-end, repeat, LSU-end) or discordant strand, such read was classified as “other, possibly erroneous configuration”. Lower number of reads in the (X-)repeat-SSU variants are most probably artifact, caused by longer sequence between repeated region and the beginning of the SSU than in other cases.

Consensus sequences of all four informative groups contained only short indels compared to SPAdes results and indicated that in all cases additional small repeat was present within repeated region.

PCR products were acquired in all four variants (Figure 1). Their lengths are in accordance with acquired PacBio consensus results.

Table 1

PacBio reads and length of predicted PCR products.

Number of PacBio reads classified to analysed variants. Reads corresponding to both hypotheses are bolded. For those variants reads were aligned, consensus sequences acquired and length of expected PCR products predicted.

Variant	# of reads	predicted length
SSU – repeat – SSU	5	2 240
SSU – repeat – LSU	614	1 150
LSU – repeat – SSU	155	2 280
LSU – repeat – LSU	11	1 300
SSU – repeat	1 272	-
LSU – repeat	1 483	-
repeat – SSU	561	-
repeat – LSU	949	-
singular fragments	16 850	-
other, possibly erroneous	150	-
total	22 050	-

DISCUSSION

PacBio results show that in fact 670 bp repeated region is located in both the ITS-1 and the IGS. The PCR results confirm that at least small fraction of the SSU and the LSU genes are organized differently than in model species *E. gracilis*. Just a few PacBio reads correspond to those PCR products, however this may be the result of fractioning of the DNA size (>20 000 bp) before sequencing. It is more probable that bigger fraction of larger circles with both SSU and LSU (11 800 bp long) was included in analysis, than the smaller ones (4 400 bp for the SSU circle, 7 400 bp for the LSU circle).

It is also possible that other organizations are present in *E. longa* cells, e.g. small linear molecules, chromosomally located rDNA operons or mix of several types. Furthermore, the structure may change in time, depending on the environmental conditions (e.g. in older culture).

FURTHER READING:

1. Greenwood, S. J., Schnare, M. N., Cook, J. R., & Gray, M. W. (2001). Analysis of intergenic spacer transcripts suggests “read-around” transcription of the extrachromosomal circular rDNA in *Euglena gracilis*. *Nucleic Acids Research*, 29(10), 2191–2198. 10.1093/nar/29.10.2191.
2. Gumińska, N., Plecha, M., Walkiewicz, H., Hałakuc, P., Zakryś, B., & Milanowski, R. (2018). Culture purification and DNA extraction procedures suitable for next-generation sequencing of euglenids. *Journal of Applied Phycology*, 1–9. 10.1007/s10811-018-1496-0.
3. Ravel-Chapuis, P., Nicolas, P., Nigon, V., Neyret, O., & Freyssinet, G. (1985). Extrachromosomal circular nuclear rDNA in *Euglena gracilis*. *Nucleic Acids Research*, 13(20), 7529–7537. 10.1093/nar/13.20.7529.
4. Schnare, M. N., Cook, J. R., & Gray, M. W. (1990). Fourteen internal transcribed spacers in the circular ribosomal DNA of *Euglena gracilis*. *Journal of Molecular Biology*, 215(1), 85–91. 10.1016/S0022-2836(05)80097-X.
5. Torres-Machorro, A. L., Hernández, R., Cevallos, A. M., & López-Villaseñor, I. (2010). Ribosomal RNA genes in eukaryotic microorganisms: Witnesses of phylogeny? *FEMS Microbiology Reviews*, 34(1), 59–86. 10.1111/j.1574-6976.2009.00196.x.

METHODS

1. Long reads genomic sequencing (PacBio)

Total DNA of *E. longa* was isolated using modified CTAB protocol (Gumińska et al. 2018) and fractionated by size (>20 000 bp). Acquired PacBio reads were corrected and searched using local blastn. Fragments of repeated region, beginning and end of 18S rRNA, 5.8S rRNA and end of 28S rRNA genes were used as queries (Figure 1). Results were parsed using in-house script, reads corresponding to four arrangements (SSU-SSU, SSU-LSU, LSU-SSU and LSU-LSU) extracted and aligned in four separate groups using MAFFT. Consensus sequences were used to resolve ambiguous regions.

2. Inverse PCR

To confirm bioinformatic results inverse PCR was performed. Primers were corresponding to queries from blast search (beginning and end of 18S rRNA, 5.8S rRNA, and end of 28S rRNA genes). Four possible products were amplified: SSU-F/SSU-R, SSU-F/LSU-R, LSU-F/SSU-R and LSU-F/LSU-R (Figure 1). Products were sized on 2% agarose gel and visualized.

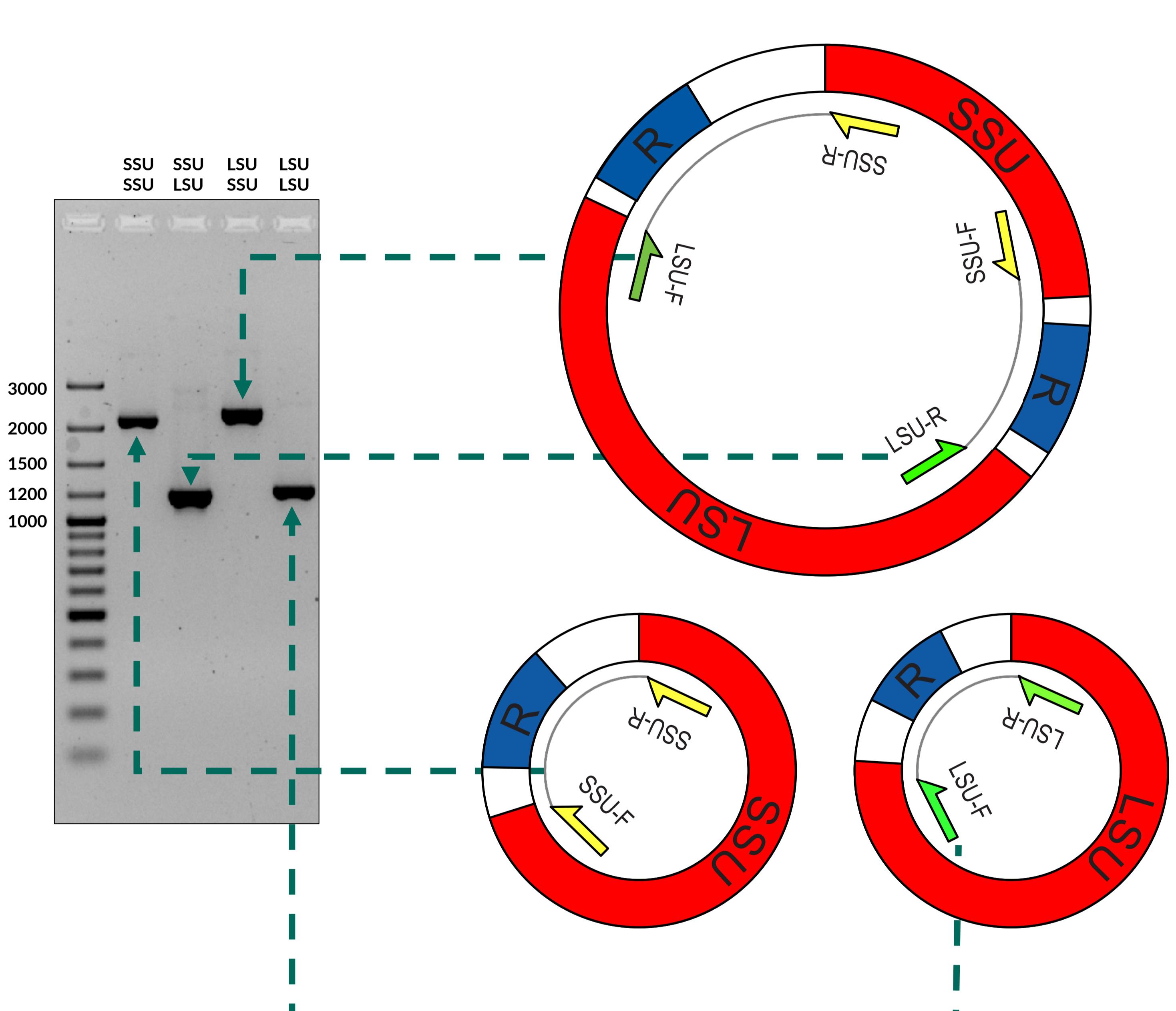


Figure 1

Schemes of hypothetical rDNA molecules in *E. longa* and localization of primers used in inverse PCR experiment.

Schematic representation of hypothetical circular rDNA molecules in *E. longa*. The LSU 1-14 were concatenated for clarity. On scheme: the SSU and the LSU genes (red), repeated region (R, blue) inverse PCR primers (SSU-F and SSU-R in yellow; LSU-F and LSU-R in green), acquired PCR products (gray arcs). PCR products were sized on 2% agarose gel.

PERSPECTIVES

To confirm the type of rDNA molecules in *E. longa* we will conduct Southern blot analysis with restriction enzyme digestion. It will allow us to unambiguously distinguish all possible variants (linear, circular, chromosomal) and ascertain ratio between different different forms. Additionally we plan to analyze ratios between the forms in various growth conditions (stress vs no-stress). Ratio will be estimated using HiSeq sequencing of circular DNA fraction, without size-dependent fractioning.



Take me home!



This work was supported by grant 2015/19/B/NZ8/00166 from National Science Centre, Poland.