Target Journal: Molecular Phylogenetics and Evolution – Short communication - 3000 words

LBNL-59012

Rolling Circle Amplification of Metazoan Mitochondrial Genomes

W. Brian Simison[a,b*], D. R. Lindberg[a,c] J. L. Boore[a,b],

[a] *Department of Integrative Biology, University of California, Berkeley, CA 94720*

[b] *DOE Joint Genome Institute and Lawrence Berkeley National Laboratory, 2800 Mitchell Dr., Walnut Creek, CA 94598*

[c] *Museum of Paleontology, University of California, Berkeley, CA 94720*

1. Introduction

Ever since Darwin (1859) envisioned his grandeur (BRIAN – I RECOGNIZE THE REFERENCE, BUT "GRANDEUR" IS A NOUN, NOT AN ADJECTIVE) view of life, evolutionary biologists have been resolute in their determination to reconstruct the Tree of Life. This is a daunting, yet essential task, as many medical, economic, environmental, conservation, and other social issues depend on the accurate placement of an organism or lineage on the tree (Cracraft and Donoghue, 2004). Comparisons of fragments of nuclear and organellar genomes have been prominent in reconstructing evolutionary relationships (Wilson et al., 1977), but larger scale genome sequencing provides additional phylogenetic information above the nucleotide level, including gene order, gene content and organizaton, RNA and protein folding patterns, etc. While

* Corresponding Author. *E-mail address: simey@Berkeley.Edu*

recent successes in sequencing complete eukaryotic genomes (*Homo sapiens*, *Drosophila melangoster*, *Caenorhabditis elegans*, *Arabidopsis thaliana*, *Plasmodium falciparum*, and others) (CBS Genome Atlas Database; Hallin and Ussery, 2004) have rightly garnered substantial attention, complete nuclear genome sequences for even an infinitesimal fraction of the Tree of Life remains as remote now as it was in Darwin's day. However, the smaller size of the mitochondrial genome has resulted in the publication of more than 600 complete mt genomes and that number is growing rapidly (CBS Genome Atlas Database; Hallin and Ussery, 2004).

The use of organelle DNA has been a mainstay in the reconstruction of phylogenetic relations (Simon et al., 2004) and has substantial advantages over nuclear genes (e.g., small compact genome, haploid with no recombination, typically maternally inherited, conserved amino acid sequences, etc.). Moreover, complete organellar genomes provide new sources of phylogenetic information, such as gene order and secondary and tertiary structures of proteins or RNAs (Boore, 1999).

With the exception of the Vertebrata and Arthropoda, existing techniques to sequence complete metazoan mt genomes (such as long PCR) require unique solutions for each taxon under investigation – often in spite of close relationships (e.g., Rawlings et al., 2003). Transpositions, nucleotide substitutions, insertions, or deletions, and duplications of portions of the mtDNA all contribute to the need for extensive troubleshooting and optimization for the production of complete mt genomes using long PCR approaches.

Here we report the successful use of rolling circle amplification (RCA) for the amplification of complete metazoan mt genomes to make a product that is amenable to high-throughput genome sequencing techniques. The benefits of RCA over PCR are many and with further development and

refinement of RCA, the sequencing of organellar genomics will require far less time and effort than current long PCR approaches.

2. Strand displacement rolling circle amplification (RCA)

RCA has been used in the amplification of microbial genomes, for the amplification of signal from probes, and in plasmid amplification for sequencing (Dean et al. 2001, Hawkins et al. 2002, Nelson et al., 2002;Lizardi et al. 1998). The benefits of RCA lie in its universal priming conditions; RCA employs numerous random short primers, which anneal at many sites to a DNA template. Once the primers have annealed to the DNA, the phi29 DNA polymerase extends the annealed primers and, because of its strong strand displacement activity and high processivity, generates long concatamers of the circular template DNA in a single isothermal reaction.. The phi29 polymerase also has an inherent 3´– 5´ proofreading exonculease activity, accounting for its high fidelity (Esteban et al. 1993). Despite the name, this reaction does not work only on circular DNA; linear DNA also will undergo repeated rounds of simple strand displacement amplification (Lizardi et al. 1998), so it is critical that the DNA preparation be highly enriched for mtDNA, although the amount can be miniscule.

Here we describe the extraction protocols used for both RCA and long PCR on the same two gastropod mollusks, followed by their respective amplification protocols, a description of the shotgun sequencing protocol used, and the sequence annotation procedure to generate complete mtDNA sequences.

Because we had repeated difficulty, we tried several methods of purifying DNA, always trying to minimize the amount of nuclear DNA present. This protocol was only successful when we used fresh, unfertilized eggs dissected from aquarium-raised specimens. RCA never succeeded when applied to DNA we'd extracted from somatic tissues despite multiple attempts at optimizing this.

Approximately 0.1 g of eggs were mixed with 10 ml of homogenization buffer in a 50 ml Oakridge tube. The homogenization buffer was a 1:5 mixture of the following pH 7.5 buffers: (1) 1.5 M sucrose/10 mM TRIS/100 mM EDTA and (2) 10 mM TRIS/100 mM EDTA/10 mM NaCl. This mixture was homogenized with a Tekmar T25 Tissuemizer at a speed setting of one using 5-10 strokes of 10-15 seconds each. The homogenized sample was centrifuged for 5 minutes at 2,300 x g and 4º C to pellet and remove nuclei. The supernatant was transferred to a fresh Oakridge tube and centrifuged for 25 minutes at 23,000 x g and 4º C. The mitochondrial pellet was resuspended in 3 ml of room temperature (RT) TE, then 750 µl of 10% SDS was gently mixed into the suspension and incubated at RT for 10 minutes to lyse mitochondrial membranes. The sample was centrifuged for 10 minutes at 17,000 x g and 4º C to pellet and remove membranous debris. The resulting supernatant was extracted first with phenol and then with chloroform, then a half-volume of 7.5 M ammonium acetate was added and the solution gently mixed. Two volumes of 100% ethanol were added and mixed by inversion. The mixture was incubated at -20º C for 1 hour and precipitated by centrifugation at 20,000 x g for 45 minutes at 4º C. The pellet was air-dried and resuspended in 100 µl ddh_20.

For whole genomic extractions that were effective for PCR, we extracted DNA from a 2-6 mg sample of foot tissue using a Qiagen DNeasy Tissue kit (Valencia, CA). Long PCR of the *Lottia digitalis* and *Ilyanassa obsoleta* mt genomes were completed using five and four sets of primers respectively (Table 1). The primers were designed from *rrnL* and *cox1* sequence fragments, which

were generated using the 16Sa-L2510 and 16Sb-H3080 primers described by Palumbi (1996) and the HCO-2193 and LCO-1490 primers described by Folmer et al. (1994), respectively. After sequencing these short fragments, primers were designed using Oligo 4.0 (Molecular Biology Insights, Inc., Cascade, CO) for creating long PCR products that would span the regions between *rrnL* and *cox1*. Long PCR cycling parameters started with a 94Cº step for 2 minutes followed by 10 cycles of 94Cº for 10 seconds, 55±10Cº for 30 seconds and 72Cº for 10 minutes. This was followed by 20 cycles of 94Cº for 10 seconds, 55±10Cº for 30 seconds and 72Cº for 10 minutes plus 10 seconds per cycle.

Rolling circle amplification was performed with the Amersham Biosciences TempliPhi DNA amplification kit (Piscataway, NJ). The reaction required two steps. The first step was a denaturation step, in which 2 µl of the product from the mitochondrial isolation protocol above was mixed with 2 µl of the TempliPhi denaturation buffer and 6 µl H_20. The sample was then denatured at 95ºC for 5 minutes then brought to RT. 10 µl of TempliPhi amplification buffer (includes enzyme) was added to the sample and incubated at 34º C for 18 hours. The reaction enzyme was then inactivated at 65º C for 10 minutes. The reaction product was purified using an Amicon Ultrafree-MC minicolumn kit (Fisher Scientific, Hampton, NH). The purified product was run on a 1% electrophoresis agarose gel to visualize the RCA reaction result. The product was also cut with Invotrogen TaqI restriction enzyme (Carlsbad, CA) and run on a 1% electrophoresis agarose gel, with the expectation that only mtDNA would yield a banding pattern, with amplified nuclear DNA, in contrast, expected to generate a smear, since there would be a very large number of randomly located TaqI sites.

The DNA from successful amplification reactions was sheared into small fragments by repeatedly passing it through a narrow aperture using a Gene Machines HydroShear . The DNA stretches as it passes through under high pressure, and breaks if it is longer than a size specified by the pressure, typically between 1.0 and 1.5 kb for mtDNA. Repairs to make the ends blunt were made using T4 DNA polymerase and Klenow fragment enzymes, then the products was electrophoresed in 1% agarose gels for size selection. The portion between 1.0 and 1.5 kb was excised from the gel and extracted using Qiagen Gel Extraction kits (Valencia, CA). These were ligated into pUC18 plasmid vectors using Fast-Link ligation kits (Epicentre Technologies, Madison, WI) and the resulting plasmids were used to transform *E. coli* DH10b via electroporation. Colonies were grown overnight on LB/Amp/X-gal plates and then an automated Genetix colony picker was used to select colonies into 384-well plates of LB with 10% glycerol. These were incubated overnight in a static incubator, without shaking or enhanced aeration, and then a small aliquot was processed robotically through plasmid amplification using RCA. Sequencing was carried out using an ABI 3730xl automated capillary DNA sequencer with ABI BigDye chemistry (Foster City, CA), and SPRI reaction clean up (Elkin et al., 2002). Detailed protocols are available at http://www.jgi.doe.gov/Internal/prots_index.html. We determined at least 10 times the amount of sequence in the target template, and then assembled the sequencing reads using the assembler program PHRAP (Green, 1996). Quality of sequence and assembly were manually verified using the program CONSED (Green, 1996). Genes were annotated using DOGMA (Wyman et al., 2004), then imported into MacVector (Accelrys, San Diego, CA) for subsequent analyses.

3. The mitochondrial genomes of *Ilyanassa obsoleta* and *Lottia digitalis* (Fig. 1)

The complete mitochondrial (mt) genomes of the gastropod mollusks *Ilyanassa obsoleta* and *Lottia digitalis* were obtained through RCA. The sequences obtained through RCA were independently confirmed by comparing them to mt genome sequences obtained using long PCR techniques. Both genomes contain the typical 13 protein genes, 2 ribosomal RNA genes, and 22 tRNA genes (Fig. 1).

The 26,400 bp mt genome of *Lottia digitalis* contains several repeating units distributed across two regions of the genome (Fig. 1A). The first region spans approx.7,000 bp between *rrnS* and *cob* and contains a series of overlapping tandem repeats . The second region spans 1,500 bp and lies between *nad1* and *nad6*. We suspect length polymorphism for the former repeat region because long PCR amplification across this region generated from eggs pooled from multiple individuals produced several bands of similar sizes while the same PCR reactions of single individuals produced single bands of different sizes. Significant variation (0-6.2 kb) has been reported in tandem repeat regions in the mt genome of other organisms, including lizards (Moritz and Brown 1986), rabbits (Pfeuty et al. 2001), sheep (Hiendleder et al. 1998) and the blood fluke (Bieberich & Minchella), often involving the control region (D-loop), but sometimes including genes as well.

We found numerous single nucleotide polymorphisms (SNPs) in the *Ilyanassa obsoleta* library created from the RCA product. Of the SNPs found in coding regions, 91.3% (105/115) were synonymous. This suggests to us that they are not due to any polymerase or sequencing errors, but rather, represent actual variation in mtDNA sequences. The *Lottia digitalis* assembly, on the other hand, revealed very few ambiguous positions.

4. Extra-chromosomal replicas of the nuclear ribosomal RNAs?

In addition to the complete mt genome, we found that libraries generated from RCA of *L. digitalis* consistently recovered the sequences of the nuclear 18S, 5.8S, and 28S ribosomal complex as a continuous unit including ETS 1 and 2 and ITS 1 and 2 (Fig 2). The ends of this assembly match one another and therefore suggest that this complex exists as a tandem repeat or as a circle. The amount of this rDNA sequence product is much greater than one would expect from random sequences or background nuclear DNA amplification. Ribosome production is particularly critical during early embryonic development and Brown and Dawid (1968) have demonstrated that *Xenopus* oocyte nuclei may contain up to a million extra-chromosomal replicas of these genes for the production of ribosomal RNA during early development. They also reported extra copies of the genes for 18S and 28S in eggs of the bivalve mollusk *Spisula solidisima* and in the annelid worm *Urechis caupo*. Therefore, it might not be too surprising to find that eggs of a gastropod would contain such extrachromosomal DNA, or that it would amplify during RCA (especially if it is circular) and be found in the sequenced product. Unfortunately, sequence assemblies cannot differentiate between tandem repeats and circular entities. Based on the results of Brown and Dewid (1968) it is more likely that the nuclear ribosomal complex exists in high copy number of tandem repeats.

In contrast to the presence of all three ribosomes in *L. digitalis*, a BLAST search of the *I. obsoleta* RCA sequence library revealed only a small number of matches with the *I. obsoleta* 28S gene.

The apparent difference between copy numbers of different components of the nuclear ribosome complex in *L. digitalis* and *I. obsoleta* may have resulted from their respective eggs being at different stages of oogenesis. However, these two species also have substantially different

modes of larval development – *L. digitalis* has lecitrophic development, while *I. obsoleta* has planktotrophic development, and this marked difference in resource provisioning for the developing larvae may also affect production of ribosomes as well. Clearly, the possible role, implications, and mechanics associated with the presence of high copy numbers of nuclear and/or extranuclear ribosome genes in the Mollusca require further study.

5. Discussion

The benefits of RCA over long PCR are numerous and include: (1) the elimination of time-consuming primer design and testing. (2) The elimination of multiple PCR products to complete a genome. (3) A substantial reduction of the cost of PCR primer walking. (4) The elimination of thermal cycling. (5) The reduction of the number of shotgun libraries from between 4 and 12 to 1. (6) The dramatic reduction of overall time investment in obtaining a complete mt genome by as much as 24 fold (2 weeks from RCA to complete mt genome sequence compared to 6 months to a year using long PCR and primer walking techniques. (7) Reduced equipment needs because RCA reactions require only standard bench top heat blocks.

Currently, there are only about 600 complete metazoan mt genomes published in Genbank and the vast majority of these are for vertebrates (>70%) (Helfenbein et al., 2004). Because of this limited sampling what is now considered to be “typical” mt genomes may someday be considered unique to their respective lineages. This is especially apparent when we consider the complexity of gene rearrangements so far documented by the limited number of complete mt genomes in the mollusks and other lophotrochozoan taxa (e.g., Hoffmann et al., 1992; Boore and Brown, 1995;

Kurabyashi and Ueshima, 2000). Additional data will undoubtedly provide a better understanding of the variation present in metazoan mt genomes, especially as continued development of RCA accelerates our progress in sequencing complete organellar genomes. Moreover, as the patterns of variation become better understood, it will also expand our knowledge of the mechanics of organellar evolution and reveal unexpected structures like the possibility of circular phases of nuclear ribosomal genes noted above.

Our demonstration of a successful application of RCA to metazoan mtDNAs may make possible the widespread use of organellar genomics in phylogenetic reconstruction as well as additional insights into the evolution of the mt genome and other circular DNA entities. Additional challenges remain, especially advancements in mitochondrial purification and continued development and refinement of RCA techniques. However, we predict that with a modest ramping up of investigator interest, RCA will make the retrieval and use of metazoan mitochondrial genomes commonplace in evolutionary and molecular studies, and Darwin's grand view may no longer be quite as distant.

Acknowledgments

We thank S. Moshel-Lynch for providing specimens of *Ilyanassa obsoleta*, M. Medina for her assistance with primer design for long PCR analysis and Matt Fourcade for his troubleshooting assistance with the mt DNA extraction protocol. This work was support by National Science Foundation Grant DEB-0089624 and performed partly under the auspices of the US Department of Energy's Office of Science, Biological and Environmental Research Program, and by the University of California, Lawrence Berkeley National Laboratory under Contract No. DE-AC02-05CH11231.

References

Bieberich AA, Minchella DJ. 2001. Mitochondrial inheritance in Schistosoma mansoni: mitochondrial variable number tandem repeat mutation produces noise on top of the signal. J Parasitol. 2001 Oct;87(5):1011-5.

*Boore, J. L., 1999 Animal mitochondrial genomes Nucleic Acids Res 27, 1761-1780.

*Boore, J. L. Brown, W.M., 1995. Complete DNA sequence of the mitochondrial genome of the annelid worm, *Lumbricus terrestris*. Genetics 141, 305-319.

*Brown, D.D. Dawid, I. B. (1968). Specific gene amplification in oocytes. Oocyte nuclei contain extrachromosomal replicas of the genes for rRNA. Science 160, 272-280.

*Cracraft, J., Donoghue, M. J., 2004. Assembling the Tree of Life. Oxford University Press, NY.

*Darwin, C., 1859. On the Origin of Species by Means of Natural Selection. John Murray, London.

*Dean, F. B., Nelson, J. R., Giesler, T. L., Lasken, R. S., (2001). Rapid amplification of plasmid and phage DNA using phi29 DNA polymerase and multiply-primed rolling circle amplification. Genome Res. 11, 1095-1099.

*Elkin, C., Kapur, H., Smith, T., Humphries, D., Pollard, M., Hammon, N., Hawkins, T., 2002. Magnetic bead purification of labeled DNA fragments for high-throughput capillary electrophoresis sequencing. Biotechniques 32**,** 1296-1302.

Esteban, J.A., Salas, M., and Blanco, L. (1993) Fidelity of phi29 DNA polymerase. Comparison between protein-primed initiation and DNA polymerization. *J Biol Chem* **268:**2719-2726.

*Folmer, O., Black, M., Hoeh, W., Lutz, R., Vrijenhoek, R., 1994. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. Mol. Mar. Biol. Biotechnol. 3, 294–299.

*Green, P., 1996. PHRAP. http://bozeman.mbt.washington.edu/phrap.docs/phrap.html.

*Hallin, P.F., Ussery, D., 2004. CBS Genome Atlas Database: A dynamic storage for bioinformatic results and sequence data. Bioinformatics 20, 3682-3686.

*Hawkins, T. L., Detter, J. C., Richardson, P. M., 2002. Whole genome amplification applications and advances. Curr. Opin. Biotech. 13**,** 65-67.

*Helfenbein, K.G., Fourcade, H.M., Vanjani, R.G., Boore, J.L., 2004. The mitochondrial genome of *Paraspadella gotoi* is highly reduced and reveals that chaetognaths are a sister group to protostomes. Proc. Nat. Acad. Sci. 101, 10639-10643.

*Hoffmann, R.J., Boore, J.L., Brown, W. M., 1992. A novel mitochondrial genome organization for the blue mussel, Mytilus edulis. Genetics 131, 397-412.

Hiendleder S, Lewalski H, Wassmuth R, Janke A. 1998. The complete mitochondrial DNA sequence of the domestic sheep (Ovis aries) and comparison with the other major ovine haplotype. J Mol Evol. 1998 Oct;47(4):441-8.

*Kurabayashi, A., Ueshima, R., 2000. Complete sequence of the mitochondrial DNA of the primitive opisthobranch gastropod *Pupa strigosa*: systematic implication of the genome organization. Mol. Biol. Evol. 17, 266-277.

Lizardi, P.M., Huang, X., Zhu, Z., Bray-Ward, P., Thomas, D.C., and Ward, D.C. (1998) Mutation detection and single-molecule counting using isothermal rolling-circle amplification. *Nat Genet* **19:**225-232.

*Mortiz, C., Brown, W.M., 1986. Tandem duplication of D-loop anda ribosomal RNA sequences in lizard mitochondrial DNA. Science 233, 1425-1427.

*Nelson, J.R., Cai, Y.C., Giesler, T.L., Farchaus, J.W., Sundaram, S.T., Ortiz-Rivera, M., Hosta, L.P., Hewitt, P.L., Mamone, J.A., Palaniappan, C., Fuller, C.W., 2002. TempliPhi, phi29 DNA polymerase based rolling circle amplification of templates for DNA sequencing. Biotech. Suppl., 44–47.

*Palumbi, S. R., 1996. Nucleic acids II: the polymerase chain reaction. In: Hillis, D.M., Moritz, C., Mable, B.K., (Eds.), Molecular Systematics, second ed. Sinauer, Sunderland, MA, pp. 205-246.

Pfeuty A, Gueride M, Lecellier G. 2001. Expansion/contraction of mammalian mitochondrial DNA repeats in Escherichia coli mimics the mitochondrial heteroplasmy. J Mol Biol. 2001 Dec 7;314(4):709-16.

*Rawlings, T.A., Collins, T.M., Bieler, R. 2003. Changing identities: tRNA duplication and remolding within animal mitochondrial genomes. Proc. Nat. Acad. Sci. 100, 15700-15705.

*Simon, C., Frati, F., Beckenbach, A.T., Crespi, B., Liu, H., Flook, P., 1994. Evolution, weighting and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. Ann. Entomol. Soc. Amer. 87, 651-701.

*Wilson, A.C., Carlson, S.S., White, T.J., 1977. Biochemical evolution. Ann. Rev. Biochem. 46, 573-639.

*Wyman, S. K., Jansen, R. K., Boore, J. L., 2004. Automatic annotation of organellar genomes with DOGMA. Bioinformatics 20, 3252-3255.

Tables

Table 1.

Long PCR primers used to successfully amplify large portions of the *Lottia digitalis* and *Ilyanassa obsoleta* mitochondrial genomes.Long PCR combinations for *L digitalis* were 1+2, 3+4, 5+6, 7+8 and 9+10. *I. Obsoleta* 11+12, 11+13, 13+14 and 14+15.

Taxon	No.	Primer	Sequence (5'-3')
Lottia digitalis	1	LdigCOI456U20	ACCTTTATCGGCTCTACCTT
	2	Ldig16S169L25	AAATTAATGCAACTGAGACAGCGTC
	3	WBS307Urc	TAGCCAACTCCTATCTATGC
	4	WBS3240Lrc	TCAGATGAGTTAACCCGAAG
	5	L5444digit	ATTTTCGTTAGGTTGGCTTTTCTTTGTA
	6	U2326digit	AATTTCTGCTGTGGCGGTGGGTCTCGTC
	7	YLAYKZ14206L25	AGAAAAGAGATAAACCAAGTCAGGA
	8	YLAYKZ785U29	ACACGGGTACAGGATTTTTAGTCTCAGGT
	9	YLAYKZ17697L28	TATTACTTAGCCCCTGTTCTTGTCATTG
	10	YLAYKZ5210U25	GGTTCCCCCAAAAAGGCACATTAGT
Ilyanassa obsoleta	11	16SCOII366U28i	AAAGGAATTAGTTACCGTAGGGATAACA
	12	16SCOII815L24i	ATAAACAGTTCACCCAGTCCCAAC
	13	COI16SI902L28i	TTAAAGCTCGATAGGGTCTTCTTGTCTT
	14	COI16SI307U27i	TTGGGACTGGGTGAACTGTTTATCCAC
	15	COI16SI809L28i	TTCATGTCAAACCATTCATACTAGCCTT

Figure Captions

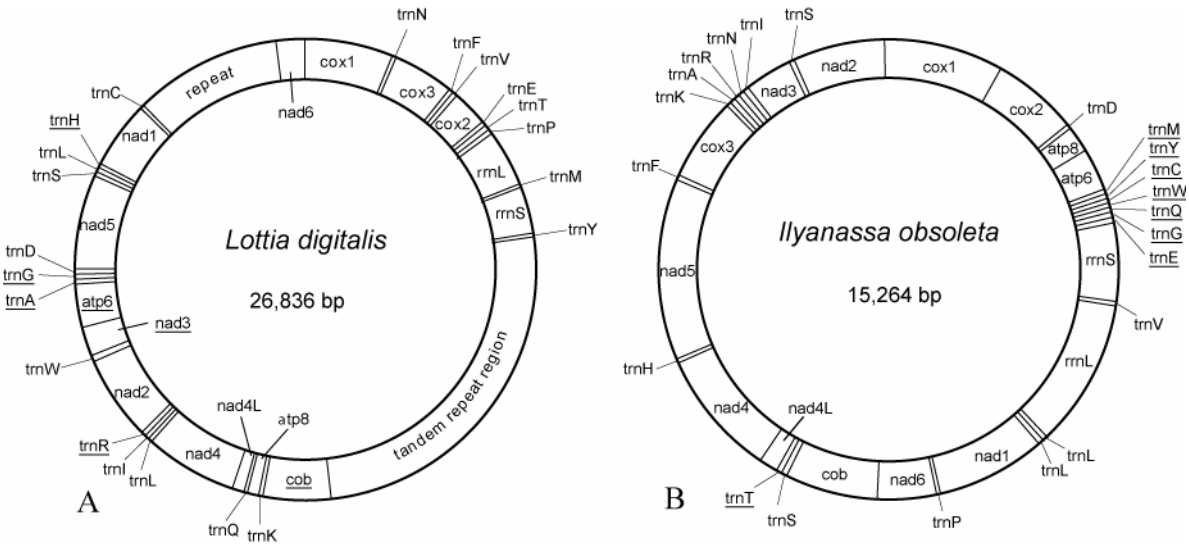

Fig. 1. Complete mt genomes obtained through Rolling Circle Amplification for the gastropod molluscs: A. *Lottia digtalis* (GenBank Acc. # XXXXX), and B. *Ilynassa obsoleta* (GenBank Acc. # XXXXX).

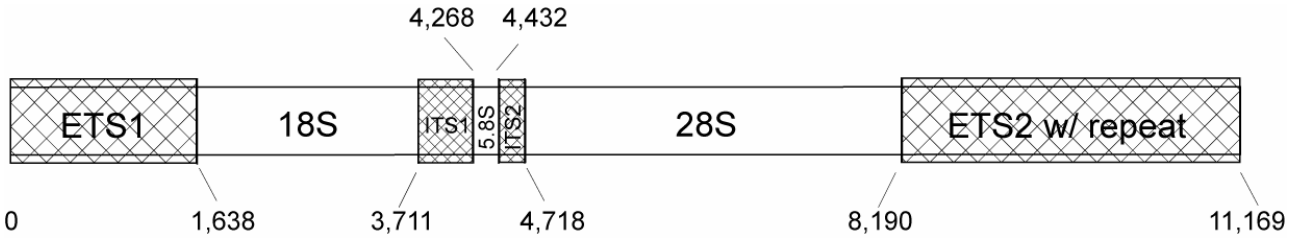

Fig. 2. Nuclear ribosomal complex for the gastropod mollusc: *Lottia digtalis* (GenBank Acc. # XXXXX),