Final Report

Brilliant Marine Research Idea 2019

1. General information

Title of the idea	Small worms and their problematic mitochondrial genomes
Name PhD student	Marlies Monnens
Name supervisor	Prof. Dr. Tom Artois (co-supervisor: Prof. Dr. Maarten Vanhove)
Flemish University or Flemish University College	Hasselt University

2. Brilliant Marine Research Idea – Report about the activities

Abstract

With about 1800 species formally described, Rhabdocoela MEIXNER, 1925 readily surpasses all other turbellarian taxa (i.e. non-neodermatan flatworms) in terms of species richness. Also in terms of ecology, these animals are highly diversified, occupying numerous niches in marine, freshwater, and even (limno-)terrestrial environments. At least four different lineages have independently acquired an obligate symbiotic lifestyle. The sister group-relationships of these 'shifted' lineages are in most cases firmly established. As such, symbiotic rhabdocoels form an important model to assess the molecular-genetic effects of such large evolutionary shifts at the genomic level in a comparative framework.

However, accessing these sequences has proven difficult. Considering the small size of most rhabdocoels (< 2 mm), acquiring adequate concentrations of target DNA makes up a first challenge – even in the current NGS era. The few molecular data available today also show a high degree of nucleotide diversity in the (mito)genomes of these animals, further complicating otherwise-standard procedures such as *cox*1-barcoding or even simple PCR protocols. The fact that no reference genome is available in this modern age speaks volumes – and only a single mitochondrial genome has been published to date. In this report, the first steps towards a quick and streamlined pipeline for sequencing complete mitochondrial genomes of rhabdocoels are presented. This approach can readily be customised for other meiofaunal groups. The costs and benefits of this process are considered and the resulting assemblies are discussed.

Intro

In the course of animal evolution, few taxa have become as ecologically diverse as Rhabdocoela Ehrenberg, 1831[1] (Platyhelminthes). Easily overlooked due to their small size and with very few people even aware of their existence, these flatworms make up a highly diverse assemblage encompassing ±1800 formally-described species[2]. While anciently marine, several independent lineages have successfully invaded freshwater, terrestrial, and brackish habitats. Moreover, at least four different transitions towards an obligate symbiotic lifestyle are known[3]. With the sister groups of these "derived" taxa well-characterised, rhabdocoels make up a promising, yet completely unexplored model to assess the effects of such evolutionary steps in a comparative framework. For instance, one might expect significant physiological, morphological, and/or molecular-genetic changes when such a large, evolutionary shift occurs. Here, we focus on the genomic impact of the transition towards endosymbiosis, the specific goal of this project being the sequencing, assembly, and annotation of complete mitochondrial genomes starting from single specimens.

Material & Methods

Representative specimens of all target species included in this project were collected. Free-living species were extracted from sandy sediments using the $MgCl_2$ decantation method[4]. Hosts of endosymbionts were collected by free-diving and after dissection scanned under a stereomicroscope for the presence of rhabdocoels. Obtained worms were photo-vouchered and stored in ethanol (99%) or RNAlater™ (Fisher Scientific) for downstream molecular work.
Pilot work had already shown that standard extraction methods yield inconsistent DNA quantities to proceed with library prep and subsequent shotgun sequencing. As such, a pre-amplification step was developed. Primary steps to acquire raw sequences always included a phenol-chloroform-based extraction. For a number of specimens, an alkaline lysis step was introduced in the protocol to exclude linear (hence genomic) DNA from the samples. The initial set-up included a targeted, long-range polymerase chain reaction (LR-PCR), followed by sequencing through primer walking. As so-called universal (*cox*1) primers generally do not perform in platyhelminths, primers were designed *in silico*, using the four rhabdocoel mitochondrial genomes available today as a template[5,6]. Annotated sequences were mined from Genbank[7] and imported in Geneious v11.1.5[8]. A translation-based multiple alignment was built using the Muscle v3.8.425[9] executable (transl_table 9) and visually inspected for conserved regions. Primers were chosen in these regions based on criteria put forward by Qiagen[10] and commercially synthesised (Biolegio). In addition, previously-designed 'rhabdocoel-specific' *cox*1[11] primers were used. Standard PCR was prepared using Illustra PuReTaq Ready-To-Go beads (GE Healthcare), or when optimisation steps were mandatory, a custom mix of platinum Taq DNA polymerase (Fisher Scientific), dNTPs (Fisher Scientific), and Bovine Serum Albumin (Fisher Scientific). Reactions were carried out on a Bio-Rad PCR T-100 Touch machine. Amplification was verified on a 2.5% agarose gel stained with GelRed® (Biotium). Successfully-amplified products were sent for commercial sequencing on an ABI3730XL machine (Macrogen Europe, Amsterdam). Acquired reads were assembled and trimmed (95% confidence) in Geneious and checked for signs of contamination using a BLAST search[12] on the NCBI server (ncbi.nlm.nih.gov). Resulting sequences were then used

as a template for reverse primer design in Geneious, following the LR-PCR specifications offered by Qiagen[13]. LR-PCR was carried out using a Qiagen LongRange PCR kit, following manufacturer's instructions and assuming minimal concentrations of starting template. Products were checked on a 0.5% agarose gel stained with GelRed®. Product purification was carried out using ExoSAP-IT™ (Thermo Fisher). Purified products were sent to Macrogen for a first 'primer walk' in commercial Sanger sequencing. In addition, we were offered an opportunity to include these amplicons on an Illumina MiSeq run (Biomed, UHasselt). Shotgun (WGS) Illumina-libraries were prepared following manufacturers' instructions for the Nextera DNA Flex Library Prep kit (Illumina). Libraries were quantified on a Qubit™ Fluorometer (Qubit dsDNA HS Assay Kit, Thermo Fisher) and size-checked on an Agilent 2100 Bioanalyzer. Sequencing reactions were prepared with the Illumina NextSeq V3 kit. In addition, Dr. Jonathan Van Damme (Thompson Rivers University) kindly offered to include a number of samples in one of his runs on the Ion Torrent platform (400 bp Ion S5).

As amplification success turned out low (see Results/Conclusions section), an alternative approach was explored to acquire sufficient mitogenomic template, including a so-called whole-genome amplification (WGA) step. WGA comprises addition of random hexamer primers and Phi29 DNA polymerase to the reaction mix, followed by a single PCR-free amplification step. An initial run was performed using the illustra GenomiPhi V2 DNA amplification kit (GE Healthcare), following manufacturers' recommendations. Later, illustra TempliPhi™ amplification kits (GE Healthcare) were employed instead (see Results/Conclusions section). Amplicons were loaded on a GelRed®-stained 0.5% agarose gel and quantified on a Qubit™ Fluorometer (Qubit dsDNA HS Assay Kit, Thermo Fisher). As sequencing by primer walking is impossible without prior knowledge of sequence reads, we here opted for a shotgun-sequencing approach. Libraries were prepared in-house in a manner identical to what has been described above and sent for commercial sequencing on the HiSeq X platform (Macrogen, South Korea). Acquired reads were trimmed in Trimmomatic v0.39[14], quality-checked in FastQC[15], and *de novo* assembled in SPAdes v3.13.0[16] or NOVOplasty v2.7.2[17], using available rhabdocoel *cox*1 sequences as seed. K-mer values were step-wise adjusted for the respective read pools to maximise contig lengths. Assemblies were annotated on the MITOS webserver[18] and manually curated from open-reading frames (as identified in Geneious, transl_table 9) and in case of protein-coding genes, alignment to other platyhelminth mitogenomes.

Results/Conclusions
LR-PCR proved challenging for our specimens. First of all, yield of standard PCR was comparatively low, with many reactions yielding either no amplicons of expected size or no UV-positive signal at all. Moreover, most PCR products acquired in this way resulted in 'mixed template' errors when processed at Macrogen. Such problems can occur when primers are not specific enough and either anneal at multiple positions throughout the genome (e.g. pseudogenes), or attach to contaminating nucleotides in the sample (e.g. host DNA or other organisms present during fixation). Species-specific or even individuum-specific primers might provide an outcome in those cases. However, since no hints of double products were present on gels, this might not be the answer. Notably, instances of mitochondrial heteroplasmy have anecdotally been reported in platyhelminths (though never formally confirmed in rhabdocoels). While we are hesitant to make preliminary assumptions, this would explain the paradoxical observation of single bands in combination with mixed template issues. Products that did result in Sanger sequences were included in downstream LR-PCR steps. Initially, amplification yield for these products was again highly limited, the vast majority of samples not resulting in any band. After several optimisation runs, a number of amplicons were obtained. However, these did not migrate out of their respective wells when loaded onto gel, even when working at very low agarose concentrations (0.5%) and applying low voltages over extended time periods. Attempts to enzymatically cut the mitochondrial genomes with universal restriction enzymes (e.g. EcoRI) to decrease amplicon sizes did not work for any of the (cleaned) samples. Notably, a Sanger trial of these products resulted in another 'mixed template' notification and no usable reads, further hinting at possible heteroplasmy in our samples. The WGA pipeline is more promising and successfully increased nucleotide concentrations from below Qubit reading threshold to >600 ng/ml. While there is no unambiguous way to assess amplicon quality ahead of shotgun sequencing, running the products on gel did result in clear bands (among others) around the 14 Kb mark. This is approximately the expected size of a flatworm mitochondrial genome. Assembly of acquired WGS read pools was straightforward in most instances, making this an interesting, yet imperfect route to acquire rhabdocoel mitochondrial genomes. Indeed, apart from the unavoidable downside of amplification bias, a relatively large portion of reads represent genomic sequences or are derived from contaminating species present during fixation. Of course, this is to be expected since WGA indiscriminately amplifies everything present in any particular sample. To prevent such waste of lane space, an alkaline lysis step (see Material & Methods) was incorporated in the extraction protocol. This technique was originally developed for acquisition of plasmids from bacterial colonies and should theoretically only extract circular molecules from samples. However, a MiSeq pilot run already showed that this technique does not consistently eliminate genomic DNA in our case, while no mitochondrial genome could be assembled from these pools. An alternative, promising new route is found in TempliPhi, a type of WGA with a chemical preference for amplifying circular template. Using this approach for our samples also successfully increased starting template to workable inputs for library prep – theoretically containing high concentrations mitochondrial DNA. These samples have been sent to Macrogen for HiSeq sequencing and we are currently waiting for the resulting read pools.

3. Overview of the expenditures

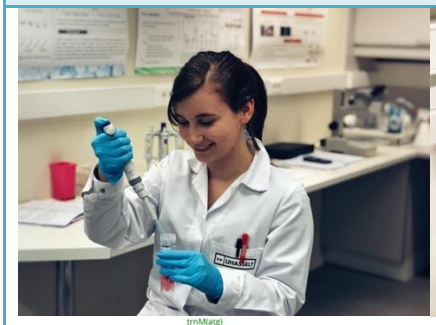
Describe in detail how the requested fund was spent within the implementation period (1 March 2019 and 28 February 2020). Be as specific as possible.

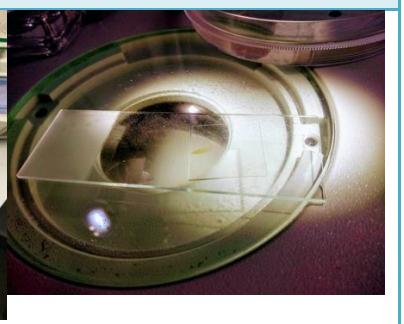
The total expenses of products bought in the framework of this project largely exceeds the requested BMRI fund. This is due to the fact that most of these products are only purchasable in larger quantities than was strictly needed for the test specimens included here. Moreover, we were for the largest part collaborating with colleagues working on similar projects and the total costs were hence allocated to their respective budgets.

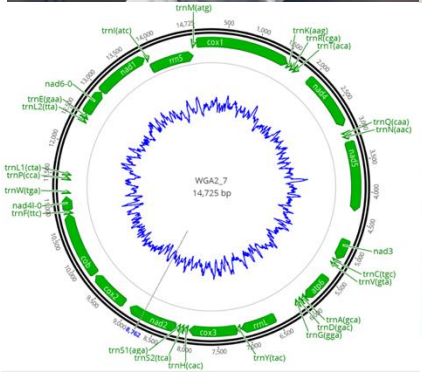
product	pricing (+VAT)	purpose
-600 rxn, platinum TAQ DNA polymerase	€564.72	PCR optimisation
-Biolegio Oligo 10 nmol synthesis	€13.20	Customised primer design
-Biolegio Oligo 10 nmol synthesis	€15.80	Customised primer design
-Biolegio Oligo 10 nmol synthesis	€54.20	Customised primer design
-Biolegio Oligo 10 nmol synthesis	€76.90	Customised primer design
-Bovine Serum Albumin, 20 mg/ml, 5 mg	€21.49	PCR optimisation
-dNTP mix 10 mM	€77.33	PCR optimisation
-ExoSAP-IT	€454.96	PCR product clean-up
-IDT for Illumina Nextera DNA Unique Dual Indexes Set A (96)	€377.52	Library prep
-Illumina Nextera DNA Flex Library (96)	€2265.12	Library prep
-Illustra GenomiPhi V2 (25 rxn)	€245.39	Whole genome amplification
-Illustra Puretaq PCR beads 0.2ml (96 rxn)	€170.25	PCR
-Macrogen HiSeq X single lane	€1400.00	Shotgun sequencing
-Macrogen PCR product sequencing	€135.00	Sanger sequencing
-Macrogen PCR product sequencing	€191.68	Sanger sequencing
-Macrogen PCR product sequencing	€216.00	Sanger sequencing
-Macrogen PCR product sequencing	€33.00	Sanger sequencing
-Macrogen PCR product sequencing	€615.12	Sanger sequencing
-Proteinase K (20 mg/ml) 2.5 ml	€246.84	DNA extraction (digestion)
-QIAGEN LongRange PCR Kit (100)	€461.37	LR-PCR
-Rack PCR 96-well natural (10)	€62.07	PCR
-Tween 20 (10%) 50 ml	€173.03	DNA extraction (elution)
-Ultrapure dnase/rnase-free water	€17.44	PCR
-UltraPure Phenol:Chloroform:Isoamyl Alcohol (100 ml)	€205.76	DNA extraction (isolation)
-X250 Strips of 8 Thermo-Tubes (250)	€212.90	PCR optimisation
-Yeast tRNA 500 ul (10 mg/ml)	€162.14	DNA extraction (pellet visualisation)

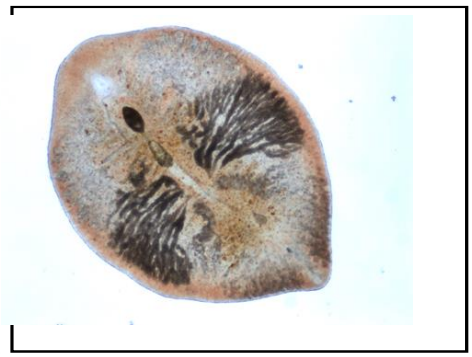
4. Pictures

A set of five pictures (low resolution in this document).









References

1 Ehrenberg, C. G. in *Hemprich und Ehrenberg "Symbolae physicae." Animalia evertebrata exclusis insectis. Series prima cum tabularum decade prima.* (ed C G Ehrenberg) folia a-d, T4-5 (Berolini, Fol. Phytozoa Turbellaria, 1831).

2 WoRMS. *Rhabdocoela*, <www.marinespecies.org/aphia.php?p=taxdetails&id=16236> (2020).

3 Van Steenkiste, N. *et al.* A comprehensive molecular phylogeny of Dalytyphloplanida (Platyhelminthes: Rhabdocoela) reveals multiple escapes from the marine environment and origins of symbiotic relationships. *PLoS One* **8**, 1-13 (2013).

4 Schockaert, E. R. in *Methods for the examination of organismal diversity in soils and sediments* (ed Geoffrey S Hall) 211-225 (CAB International in association with United Nations Educational, Scientific and Cultural Organization (UNESCO) and the International Union of Biological Sciences (IUBS), 1996).

5 Monnens, M. *et al.* The first mitochondrial genomes of endosymbiotic rhabdocoels illustrate evolutionary relaxation of *atp*8 and genome plasticity in flatworms. *submitted* (2020).

6 Kenny, N. J., Noreña, C., Damborenea, C. & Grande, C. Probing recalcitrant problems in polyclad evolution and systematics with novel mitochondrial genome resources. *Genomics* **111**, 343-355 (2019).

7 Benson, D. A. *et al.* GenBank. *Nucleic Acids Res* **40**, D48-D53 (2012).

8 Kearse, M. *et al.* Geneious Basic: an integrated and extendable dekstop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**, 1647-1649 (2012).

9 Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**, 1792-1797 (2004).

10 Qiagen. *A comprehensive guide to PCR, including how to maximize your results*, <www.qiagen.com/be/service-and-support/learning-hub/molecular-biology-methods/pcr/#PCR%20primer%20design> (2020).

11 Stephenson, I., Van Steenkiste, N. W. L. & Leander, B. S. Molecular phylogeny of neodalyellid flatworms (Rhabdocoela), including three new species from British Columbia. *J Zool Syst Evol Res* **57**, 41-56 (2018).
12 Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**, 3389-3402 (1997).
13 Qiagen. *QIAGEN LongRange PCR Handbook*, <www.qiagen.com/nl/resources/resourcedetail?id=e90c8e4b-f3c4-4876-a320-9890aaec1e17&lang=en> (2020).
14 Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: A flexible trimmer for Illumina Sequence Data. *Bioinformatics* **btu170** (2014).
15 FASTQC: a quality control tool for high throughput sequence data v. 0.11.5 (2010).
16 Bankevich, A. *et al.* SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **19**, 455-477 (2012).
17 Dierckxsens, N., Mardulyn, P. & Smits, G. NOVOPlasty: *de novo* assembly of organelle genomes from whole genome data. *Nucleic Acids Res* **45**, e18 (2017).
18 Bernt, M. *et al.* MITOS: improved de novo Metazoan Mitochondrial Genome Annotation. *Mol Phylogenet Evol* **69**, 313-319 (2013).