A simple centrifugation protocol leads to a 55-fold mitochondrial DNA enrichment and paves the way for future mitogenomic research

Jan Niklas Macher[a,b], Vera Zizka[a,b], Alexander Martin Weigand[a,b], Florian Leese[a,b]*

a) Aquatic Ecosystem Research, Faculty of Biology, University of Duisburg-Essen, Universitätsstraße 5, 45141 Essen, Germany

b) Centre for Water and Environmental Research, University of Duisburg-Essen, Universitätsstraße 2, 45141 Essen, Germany

JNM: jan.macher@uni-due.de, +49.201.183-6710

VZ: vera.zizka@uni-due.de, +49.201.183-4053

AMW: alexander.weigand@uni-due.de, +49.201.183-6710

FL: florian.leese@uni-due.de, +49.201.183-4053

* corresponding author: *jan.macher@uni-due.de*

Keywords: Metagenomics, PCR-free, freshwater mock community, cost-effective

Abstract

1. DNA (meta)barcoding is increasingly used to study and monitor biodiversity and available for standardised assessments. However, it suffers from PCR bias which can lead to the loss of taxonomic groups. PCR-free techniques such as metagenomics are therefore thought to be more suited for biodiversity assessments but are currently limited by incomplete reference libraries.

2. The technique of 'mitogenome-skimming' or 'mitogenomics', where complete mitochondrial genomes are sequenced, is ideal to bridge the techniques of (meta)barcoding and metagenomics. However, without the enrichment of mitochondria prior to DNA extraction and sequencing, roughly 99 % of reads are of non-mitochondrial origin and therefore mostly useless for species identification.

3. Here, we present a simple centrifugation protocol that leads to a 55-fold enrichment of mitochondrial DNA. By sequencing six 'mock'- communities – comprising the freshwater taxa *Corbicula fluminea*, *Gammarus roeselii* and *Hydropsyche exocellata* each – we recovered whole mitochondrial genomes with a minimum coverage of 2417-fold. In addition, we 'skimmed' the whole mitochondrial genome of the acanthocephalan endoparasite *Pomphorhynchus laevis* (444-fold coverage) obviously being present in *G. roeselii*.

4. Correlation between body mass of specimens in each community and generated sequence number was moderate in samples enriched for mitochondria. Potential problems arise with species demonstrating a high weight, but produce less than expected mitochondrial reads (e.g. shelled organisms). The developed protocol will greatly speed up building reference libraries for whole mitochondrial genomes, as dozens of species can be sequenced on a single MiSeq run. Subsequently, it will also allow biodiversity assessments using mitogenomics at greatly reduced costs in comparison to mitogenomic approaches without enrichment for mitochondria.

Introduction

Biodiversity is highly important for intact ecosystems and inevitable for human well being (Rockström *et al.* 2009). Molecular techniques such as DNA barcoding (Hebert *et al.* 2003) and metabarcoding (Hajibabaei *et al.* 2011) are increasingly used for biodiversity research but suffer from PCR stochasticity and primer bias (Elbrecht & Leese 2015). The same bias can be introduced by the use of baits or probes (e.g. Liu *et al.* 2015; Mayer *et al.* 2016). Therefore, PCR and primer/probe-free techniques harbor the potential for future biodiversity assessments (Zhou *et al.* 2013; Tang *et al.* 2014; Elbrecht & Leese 2015; Crampton-Platt *et al.* 2016; Coissac *et al.* 2016), i.e. by circumventing taxon-dependent PCR amplification bias and presenting the possibility to correlate read numbers with biomasses. Since reference libraries are still largely incomplete for nuclear genomic information, but relatively comprehensive for mitochondrial genes, such as the cytochrome c oxidase subunit I (COI) gene for animals, the consequent step towards a PCR-free analysis of biodiversity samples could be seen in 'mitochondrial metagenomics', 'mitogenomics' or 'mitogenome-skimming' (e.g. Tang *et al.* 2014; Crampton-Platt *et al.* 2015). This technique enables the comparison of newly generated mitogenomes or mitogenome fragments with reference databases and thereby links genomic information to taxonomic knowledge. However, the currently applied approaches are relatively ineffective in terms of sequencing capacity, with most 'PCR-free' mitogenomic libraries comprising less than 1 % of sequence reads of mitochondrial origin (Crampton-Platt *et al.* 2016). The major methodological disadvantage is therefore the great sequencing depth needed, rendering those approaches very cost-intensive.

Enrichment of mitochondria prior to DNA extraction and sequencing is a potential solution, shifting the initial mitochondrial to nuclear DNA ratio towards a higher mitochondrial DNA proportion. It is known that ultracentrifugation in CsCl-gradients can enrich for the typically AT-rich mitochondrial genomes (e.g. Garber & Yoder 1983). However, this approach grounds on cost- and labor-intensive ultracentrifugation and mitochondria can have highly variable AT-contents, rendering extractions from bulk biodiversity samples in CsCl-gradients less straight-forward. The enrichment of mitochondria is most promising when organelles are intact, i.e. when living tissue is used (Tamura & Aotsuka 1988). Until now, this approach has not been tested for bulk samples in a biodiversity context, mainly because most

specimens used for biodiversity assessments and genome sequencing are commonly stored in preservation fluids, which damage or destroy mitochondria. Here, we use a 'mock'- community of three freshwater species to test a simple centrifugation protocol for mitochondrial enrichment. We demonstrate that our protocol strongly enriches mitochondrial DNA and therefore can greatly reduce costs of future mitogenomic approaches, e.g. when a) constructing mitochondrial reference libraries and b) assessing biodiversity by an approach which omits biases introduced by primers, probes and PCR reactions.

Material and Methods

Sampling and laboratory protocols

Sampling was conducted at two locations (51°00'52.6"N 6°41'04.5"E; 51°05'23.4"N 6°41'17.0"E) of the Gillbach (Germany) in December 2016. Twenty individuals of each of the three macrozoobenthic freshwater species *Corbicula fluminea*, *Gammarus roeselii* and *Hydropsyche exocellata* were sampled with a dip net or collected from stones. Specimens were transferred into water (500 mL) and transported to the laboratory for immediate processing. Specimens were weighed (Mettler Toledo XS105, table S1) and assembled to six 'mock'-communities, each containing three individuals of *G. roeselii* and *H. exocellata* and a single *C. fluminea* specimen. 'Mock'- communities were separately transferred into 3 mL 5° C cold homogenization buffer (0.25 M sucrose, 10 mM EDTA, 30 mM Tris-HCl, pH 7.5; Tamura & Aotsuka 1988) in a mortar and crushed with a pestle until tissue was homogenized (70 strokes each). Two millilitres of homogenate was pipetted into a 2 mL Eppendorf tube and samples were treated after the following centrifugation protocols (4° C, centrifuge Eppendorf 5427 R) (see Supplementary material 1 for short protocol).

(1) Samples 1-3 ('Complete'- no enrichment of mitochondria): samples were centrifuged for 1 minute at 1,000 g. This step was repeated four times. Final centrifugation was conducted for 10 minutes at 14,000 g. Supernatant was discarded and 600 µL TNES (50 mM Tris Base, 400 mM NaCl, 20 mM EDTA, 0.5 % SDS) buffer was added to the pelleted material. Samples were then homogenized by vortexing.

(2) Samples 4-6 ('Mito'- enrichment of mitochondria): samples were centrifuged for 1 minute at 1,000 g. Pelletized material was discarded, the supernatant transferred to a new tube and again centrifuged for 1 minute at 1,000 g. This step was repeated three times. Final centrifugation was conducted for 10 minutes at 14,000 g. Supernatant was discarded, 600 µL TNES buffer was added to the pelleted material and samples were homogenized by vortexing.

A total volume of 40 µL Proteinase K (300 U/ml, 7Bioscience, Hartheim, Germany) was added to each sample, which were than vortexed and incubated at 37° C for 12 hours (Eppendorf Thermomixer C). DNA was extracted using a salt precipitation protocol as in Weiss & Leese (2016). For RNA digestion, 1.5 µL RNAse (1.5 µg, Thermo Fisher Scientific, Oberhausen, Germany) was added to each reaction and incubated at 34° C for 30 min on a Thermomixer, followed by a clean up using the MinElute Reaction CleanUp Kit (Qiagen, Hilden, Germany). For DNA fragmentation, samples were placed in an ultrasonic bath (Bandelin SONOREX, RK 510 Hz) for 8 hours. Library preparation was performed with a TruSeq Nano DNA LT Library Prep Kit (Set A, step (2) 'Repair Ends and Select Library Size' – (5) 'Enrich DNA Fragments'). After each step, fragment lengths and concentrations were quantified on a Fragment Analyzer (Advanced Analytical, Automated CE Systems). Samples were equimolar pooled and sent for sequencing on a MiSeq sequencer (v2 chemistry, 250 bp paired-end) at GATC-Biotech (Konstanz, Germany).

Sequence analysis

Raw sequences were checked for remaining adapters and trimmed with BBDuk as implemented in Geneious v.10.0.9 (Kearse *et al.* 2012). The complete mitochondrial genomes of *Corbicula fluminea*, *Gammarus roeselii*, *Hydropsyche exocellata* and *Pomphorhynchus laevis* (an acanthocephalan endoparasite) were assembled using MIRA 4.0.2 (Chevreux, 2014) as implemented in Geneious (Settings: Genome, accurate). Annotations were performed with the MITOS (Bernt *et al.* 2013) web server and adjusted manually. Average genomic coverage was determined by mapping the 6,516,172 raw reads from the samples Mito 1-3 against reference genomes without repetitive regions with Bowtie2

(Langmead & Salzberg 2012; settings: Local alignment, seed length 22, max mismatches 1, best match only, min insert size 100, max insert size 251) as implemented in Geneious.

Analyses of mitochondrial enrichment were conducted in four ways:

1) Reads (pooled Mito 1-3, pooled Complete 1-3) were mapped against reference genomes including and excluding repetitive regions (i.e. the mitochondrial control region) using Bowtie2 (settings as described above), as such regions may frequently produce false-positive mitochondrial reads by attracting nuclear reads with similar repeat-motifs.
2) Raw reads were quality filtered using usearch (Edgar 2010, v9.0.2132_i86linux32) and reads with bases with a Phred Score <30 (-fastq_truncqual 30 \) and a length <200 were discarded. Of each sample (Mito 1-3, Complete 1-3), in order to compare identical read numbers per sample, 100,000 sequences were randomly selected with usearch and reads mapped against the assembled reference genomes without repetitive regions using Bowtie2 (settings as above). Random selection and mapping were repeated five times. The average percentage of mitochondrial reads was calculated for each species and for all samples enriched (Mito 1-3) and not enriched (Complete 1-3) for mitochondria.
3) BLASTn searches were conducted with raw reads against the four reference genomes (Supplementary material 2) as in Gómez-Rodríguez *et al.* (2015) with a threshold of $<10e^{-5}$. This BLAST-procedure was conducted in the following ways: (1) With reads of samples enriched for mitochondria against reference genomes including repetitive regions, (2) with reads of samples enriched for mitochondria against reference genomes excluding repetitive regions,
4) BLASTn searches were conducted with raw reads and $<10e^{-5}$ against a reference library of all 11 available trichopteran mitochondrial genomes, all 67 available amphipod mitochondrial genomes and 99 mitochondrial genomes of Heterodonta (Supplementary material 3), i.e. taxa related to the target species.

A correlation analysis between specimen body weight and read coverage per mitochondrial genome was conducted by correlating reads from the second Bowtie2 analysis (average of hq reads from five replicates that mapped against reference genomes without repetitive regions) with species weights per

sample. Since the weight of *Pomphorhynchus laevis* was not measured, it was excluded from the correlation analyses. Pearson's R^2 and Spearman's rho correlation coefficients were calculated and tested for significance.

Results

A total of 7,707,640 reads were obtained, with 1,191,468 reads for samples 'Complete 1-3' and 6,516,172 reads for samples 'Mito 1-3'. Quality filtered reads were assembled to the complete mitochondrial genomes of *Corbicula fluminea* (17,575 bp with and 16,306 bp without repetitive elements), *Gammarus roeselii* (15,017; 14,095), *Hydropsyche exocellata* (15,789; 14,909) and *Pomphorhynchus laevis* (13,886; 13,671). Average genomic coverage without repetitive regions was 2417.3 (± 295.8) for *Gammarus roeselii*, 3164.8 (± 452.6) for *Corbicula fluminea*, 3307.5 (±410) for *Hydropsyche exocellata* and 443.8 (±85) for *Pomphorhynchus laevis* when all reads from the mitochondrial enrichment samples were mapped against the reference genomes (Table 1).

1) We achieved a 54.84-fold enrichment for mitochondrial reads when complete raw reads were mapped against reference genomes without repetitive regions, respectively, a 54.82-fold enrichment when mapped against reference genomes with repetitive regions (Table S2).
2) Mapping of subsets (5x 100,000 reads) of high quality reads showed that samples not enriched for mitochondria (Complete 1-3) contained averagely 0.17 % mitochondrial reads (Figure 1 A, Table 2). In comparison, samples enriched for mitochondria included 10.31 % mitochondrial reads on average (Figure 1 B). Hence, an average 60.65-fold enrichment was achieved by the applied centrifugation protocol (Figure 1, Table 2). Mapping against genomes including repetitive elements showed a 57.9-fold enrichment of mitochondrial reads (Table S3).
3) Analyses with BLASTn against a reference library of the four assembled genomes showed that 12.46 % of all raw reads in samples enriched for mitochondria were of putative mitochondrial origin when BLAST was performed with $<10e^{-5}$ against reference genomes including repetitive

regions, and 8.996 % when repetitive regions were excluded. The potential overestimation due to the inclusion of repetitive regions was 38.46 % (Tables 3, S4).

4) Mapping of raw reads with BLASTn against a reference library of 11 trichopteran mitochondrial genomes, 67 amphipod mitochondrial genomes and 99 mitochondrial genomes of Heterodonta (i.e. relatives of the target species) showed that putatively 20.05 % (Mito samples) and 0.35 % (Complete samples) of reads were of mitochondrial origin, respectively. The potential overestimation when compared to reads that mapped against the four reference genomes excluding repetitive regions was 120.88 % (Mito samples) and 102.31 % (Complete samples) (Table 4).

Coverage per mitochondrial genome and species weight were uncorrelated for samples not enriched for mitochondria (Pearson's $R^2 = 0.071$; $p = 0.4872$; Spearman's rho = 0.017; $p = 0.9816$, Figure 2 A), and for samples enriched for mitochondria ($R^2 = 0.323$; $p = 0.1106$; rho = 0.633; $p = 0.0760$, Figure 2 B) (Table S1). When *Corbicula fluminea* was excluded, coverage and weight were not significantly correlated in samples not enriched for mitochondria ($R^2 = 0.618$; $p = 0.0639$; rho = 0.143; $p = 0.8028$, Figure 2 A), but highly and significantly correlated in samples enriched for mitochondria ($R^2 = 0.850$; $p = 0.0089$; rho = 0.886; $p = 0.0333$, Figure 2 B) (Table S5).

Discussion

We developed and tested a simple centrifugation protocol for the enrichment of mitochondrial DNA from 'mock'-communities. By using this technique and sequencing 6,516,172 raw reads for samples enriched for mitochondria, the full mitochondrial genomes of the amphipod *Gammarus roeselii*, the caddisfly *Hydropsyche exocellata* and the clam *Corbicula fluminea* were sequenced with a very high average coverage of minimum 2417-fold. In addition, we recovered the mitogenome of the acanthocephalan parasite *Pomphorhynchus laevis* with an average coverage of 444-fold. This highlights the great potential of the applied technique for fast sequencing of whole mitochondrial genomes. Subsequently, those genomes can be used for fast reference library development (Coissac *et al.* 2016) and

bioassessments based on bulk samples. By using our protocol, bulk samples can be expected to be sequenced with a high enough coverage that allows the detection of mitochondrial sequences for many hundreds to thousands of specimens in parallel. However, taxonomic assignments of these reads require the completion of reliable, well-curated full mitochondrial reference libraries. Our technique for mitochondrial enrichment can greatly speed up this process, expecting to obtain full mitochondrial genomes of at least several dozens of species with a single MiSeq lane if specimens are carefully selected and biomass is accounted for (Elbrecht *et al.* 2017). This has been shown to work well with samples not enriched for mitochondrial DNA (Gillett *et al.* 2014), and efficiency is expected to greatly increase by mitochondrial enrichment. The latter is important as specimens with a high biomass can potentially prevent smaller specimens from being sequenced with a high enough coverage.

We demonstrated that samples enriched for mitochondria contained on average 10.31 % high quality mitochondrial reads, while samples not enriched for mitochondria contained on average 0.17 % only, corresponding to a 60-fold enrichment. The enrichment factor was still 55-fold when based on raw reads. BLAST searches of raw reads against reference genomes of the four sequenced species revealed significant rates of overestimation of mitochondrial reads (38.46 % for Mito 1-3 and 45.08 % for Complete 1-3) when performing BLAST searches against reference genomes including repetitive regions such as the control regions of mitogenomes. This rate was even higher when performing BLAST searches against reference genomes of 177 taxa related to the three target taxa. Here, 20.05 % of all reads from samples enriched for mitochondria and 0.35 % reads from samples not enriched for mitochondria were reported of mitochondrial origin, suggesting an overestimation of at least 120.88 % and 102.31 % (information on *Pomphorhynchus laevis* not included in the reference library). Our results suggest that estimates of mitochondrial reads or of informative mitochondrial reads might have been overestimated in previous studies when reads were mapped against full mitochondrial genomes of related taxa. We suggest to report the amount of reads covering the mitochondrial genome without the mostly uninformative control region in future mitogenomic studies. However, we acknowledge that correct estimates of mitochondrial read number are difficult to obtain, as it is known that different assembly and mapping methods will generate different results (Crampton-Platt et al. 2016).

One potential advantage of meta- and mitogenomic approaches is to imply biomass estimates via a correlation of biomass and the number of obtained sequences for each species. As PCR bias are

omitted, evenly homogenised tissue should result in roughly correlated amounts of mitochondrial reads and sequenced specimens' body mass. As classic biodiversity assessments often rely on abundance data of the detected species, this potential feature of metagenomic approaches seems very promising. A positive correlation between biomass and sequences has been already found in previous studies (Gómez-Rodríguez *et al.* 2015), but it is yet unclear how mitochondrial enrichment influences the biomass to sequence correlation. Our analyses, albeit based on a limited amount of species and data, demonstrate that a positive correlation between biomass and number of mitochondrial reads exists in samples enriched for mitochondria when the clam *C. fluminea* was excluded. However, this was not observed for samples enriched or not enriched for mitochondria when all taxa are included. This observation might be explained by the unequal homogenisation of *C. fluminea* and the fact that specimen shells contributed to weight estimates but not to biomass, whereas only the latter compulsively generates mitochondrial reads. This observation also highlights the limitations of the technique, as proportion of reads and biomass might not correlate in certain taxa. Furthermore, and because some specimens could not be completely dried, water might have artificially increased weight estimates. The acanthocephalan endoparasite *Pomphorhynchus laevis* accounted for 3.72 % of all mitochondrial reads in samples enriched for mitochondria. Although the number and weight of *Pomphorhynchus laevis* specimens that were unwittingly included in the study is not known, it seems unlikely that those small parasites made up around 4 % of the initial tissue weight. We assume that the cystacanth was not evenly homogenised during extraction of mitochondria and thus a larger amount of its tissue and mitochondria was transferred to the mitochondrial pellet.

In order to make the procedure of extracting mitochondria from tissue more reliable and standardised, we propose an automated homogenisation technique using machines instead of manual homogenisation with mortar and pestle. This is expected to lead to an even higher enrichment of mitochondrial DNA – a desirable goal since still around 90 % of all produced reads are of putative nuclear origin. Despite this, our study shows that a simple centrifugation protocol can enrich mitochondrial DNA 55-fold from samples containing several species. The achieved coverage of complete mitochondrial genomes of minimum 2417-fold for our target species makes it obvious that even with 10 % of resulting mitochondrial reads, hundreds of specimens could be sequenced and their mitogenomes assembled in a single MiSeq run. The discovery of the acanthocephalan parasite *Pomphorhynchus laevis* further

strengthens the approach, as PCR primers and probes often do not capture unexpected taxa for which primers have not been designed. Finally, our protocol can easily be used in the field if a cooling centrifuge can be transported to a nearby location, and thus allows to process the fresh tissue material needed for high(er) rates of mitochondrial enrichments. The ease of application in combination with a) a minimized laboratory workload, b) greatly reduced costs compared to mitogenomic approaches without mitochondrial enrichment and c) the high sequencing coverage per recovered mitogenome renders our mitochondrial enrichment protocol ideal for the fast generation of reference libraries ('mitogenome skimming') and subsequently also for biodiversity assessments.

Acknowledgments

We thank Cristina Hartmann-Fatu for help in the lab and the Herzog-Sellenberg Foundation for financial support. This article is based upon work from COST Action DNAqua-Net (CA15219), supported by the COST (European Cooperation in Science and Technology) program.

Author contributions

JNM, VZ, AMW and FL designed the study. JNM and VZ sampled the specimens and performed laboratory work. JNM, VZ, AMW analysed the data. JNM, VZ, AMW, FL wrote the manuscript. All authors read and approved the final version of the manuscript.

References

Bernt, M., Donath, A., Jühling, F., Externbrink, F., Florentz, C., Fritzsch, G., Pütz, J., Middendorf, M. & Stadler, P.F. (2013). MITOS: Improved de novo metazoan mitochondrial genome annotation. *Molecular phylogenetics and evolution*, **69**, 313–319.

Coissac, E., Hollingsworth, P.M., Lavergne, S. & Taberlet, P. (2016). From barcodes to

genomes: extending the concept of DNA barcoding. *Molecular ecology*, **25**, 1423–1428.

Crampton-Platt, A., Timmermans, M.J.T.N., Gimmel, M.L., Kutty, S.N., Cockerill, T.D., Vun Khen, C. & Vogler, A.P. (2015). Soup to Tree: The Phylogeny of Beetles Inferred by Mitochondrial Metagenomics of a Bornean Rainforest Sample. *Molecular biology and evolution*, **32**, 2302–2316.

Crampton-Platt, A., Yu, D.W., Zhou, X. & Vogler, A.P. (2016). Mitochondrial metagenomics: letting the genes out of the bottle. *GigaScience*, **5**.

Edgar, R.C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* , **26**, 2460–2461.

Elbrecht, V. & Leese, F. (2015). Can DNA-Based Ecosystem Assessments Quantify Species Abundance? Testing Primer Bias and Biomass--Sequence Relationships with an Innovative Metabarcoding Protocol. *PloS one*, **10**, e0130324.

Elbrecht, V., Peinert, B. & Leese, F. (2017). Sorting things out-assessing effects of unequal specimen biomass on DNA metabarcoding. *PeerJ Preprints*, e2561v1

Garber, R.C. & Yoder, O.C. (1983). Isolation of DNA from filamentous fungi and separation into nuclear, mitochondrial, ribosomal, and plasmid components. *Analytical biochemistry*, **135**, 416–422.

Gillett, C.P.D.T., Crampton-Platt, A., Timmermans, M.J.T.N., Jordal, B.H., Emerson, B.C. & Vogler, A.P. (2014). Bulk de novo mitogenome assembly from pooled total DNA elucidates the phylogeny of weevils (Coleoptera: Curculionoidea). *Molecular biology and evolution*, **31**, 2223–2237.

Gómez-Rodríguez, C., Crampton-Platt, A., Martijn J T, Baselga, A. & Vogler, A.P. (2015). Validating the power of mitochondrial metagenomics for community ecology and

phylogenetics of complex assemblages. *Methods in ecology and evolution / British Ecological Society*, **6**, 883–894.

Hajibabaei, M., Shokralla, S., Zhou, X., Singer, G.A.C. & Baird, D.J. (2011). Environmental barcoding: a next-generation sequencing approach for biomonitoring applications using river benthos. *PloS one*, **6**, e17497.

Hebert, P.D.N., Ratnasingham, S. & deWaard, J.R. (2003). Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proceedings. Biological sciences / The Royal Society*, **270 Suppl 1**, S96–9.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P. & Drummond, A. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* , **28**, 1647–1649.

Langmead, B. & Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature methods*, **9**, 357–359.

Liu, S., Wang, X., Xie, L., Tan, M., Li, Z., Su, X., Zhang, H., Misof, B., Kjer, K.M., Tang, M., Niehuis, O., Jiang, H. & Zhou, X. (2015). Mitochondrial capture enriches mito-DNA 100 fold, enabling PCR-free mitogenomics biodiversity analysis. *Molecular ecology resources*, **16**, 470–479.

Mayer, C., Sann, M., Donath, A., Meixner, M., Podsiadlowski, L., Peters, R.S., Petersen, M., Meusemann, K., Liere, K., Wägele, J.-W., Misof, B., Bleidorn, C., Ohl, M. & Niehuis, O. (2016). BaitFisher: A Software Package for Multispecies Target DNA Enrichment Probe Design. *Molecular biology and evolution*, **33**, 1875–1886.

Rockström, J., Steffen, W., Noone, K., Persson, A., Chapin, F.S., 3rd, Lambin, E.F., Lenton,

T.M., Scheffer, M., Folke, C., Schellnhuber, H.J., Nykvist, B., de Wit, C.A., Hughes, T., van der Leeuw, S., Rodhe, H., Sörlin, S., Snyder, P.K., Costanza, R., Svedin, U., Falkenmark, M., Karlberg, L., Corell, R.W., Fabry, V.J., Hansen, J., Walker, B., Liverman, D., Richardson, K., Crutzen, P. & Foley, J.A. (2009). A safe operating space for humanity. *Nature*, **461**, 472–475.

Tamura, K. & Aotsuka, T. (1988). Rapid isolation method of animal mitochondrial DNA by the alkaline lysis procedure. *Biochemical genetics*, **26**, 815–819.

Tang, M., Tan, M., Meng, G., Yang, S., Su, X., Liu, S., Song, W., Li, Y., Wu, Q., Zhang, A. & Zhou, X. (2014). Multiplex sequencing of pooled mitochondrial genomes-a crucial step toward biodiversity analysis using mito-metagenomics. *Nucleic acids research*, **42**, e166.

Weiss, M. & Leese, F. (2016). Widely distributed and regionally isolated! Drivers of genetic structure in Gammarus fossarum in a human-impacted landscape. *BMC evolutionary biology*, **16**, 153.

Zhou, X., Li, Y., Liu, S., Yang, Q., Su, X., Zhou, L., Tang, M., Fu, R., Li, J. & Huang, Q. (2013). Ultra-deep sequencing enables high-fidelity recovery of biodiversity for bulk arthropod samples without PCR amplification. *GigaScience*, **2**, 4.

Tables

Tab. 1: Mean coverage per genome without repetitive regions. 6,516,172 raw reads were mapped with Bowtie2.

Species	**Mean coverage**
Gammarus roeselii	2417.3 ± 295.8

Corbicula fluminea	3164.8 ± 452.6
Hydropsyche exocellata	3307.5 ± 410.0
Pomphorhynchus laevis	443.8 ± 85.0

Tab. 2: Mitochondrial reads (% of total reads) in samples not enriched (Complete 1-3) and enriched for mitochondria (Mito 1-3). Averages of five subsets of 100,000 hq reads per sample.

	Gammarus roeselii	***Corbicula fluminea***	***Hydropsyche exocellata***	***Pomphorhynchus laevis***	**Combined**
Complete 1	0.024	0.017	0.064	0	0.105
Complete 2	0.162	0.158	0.009	0	0.329
Complete 3	0.030	0.017	0.021	0.008	0.076
					0.17 %
Mito 1	2.845	4.209	5.039	0.322	12.415
Mito 2	3.910	3.400	3.904	0.570	11.784
Mito 3	1.888	2.947	1.645	0.260	6.740
					10.31 %
				Enrichment factor	**60.65 x**

Tab. 3: Percentage of reads of putative mitochondrial origin when performing BLAST searches with $10e^{-5}$ against reference genomes of the four analysed species including and reference genomes excluding repetitive regions.

	With repetitive regions	Without repetitive regions	Difference
Complete	0.251	0.173	**45.09 %**
Mito	12.456	8.996	**38.46 %**
All raw reads	10.226	7.385	**38.47 %**

Tab. 4 Percentage of reads of putative mitochondrial origin when performing BLAST search with $10e^{-5}$ against reference genomes of 177 related taxa (11 Trichoptera, 67 Amphipoda, 99 Heterodonta) compared to reads mapping against reference genomes of analysed species without repetitive regions.			
	177 reference genomes of taxa related to target species	Reference genomes without repetitive regions	Difference
Complete	0.350	0.173	**102.31 %**
Mito	20.050	8.996	**122.88 %**

Figures

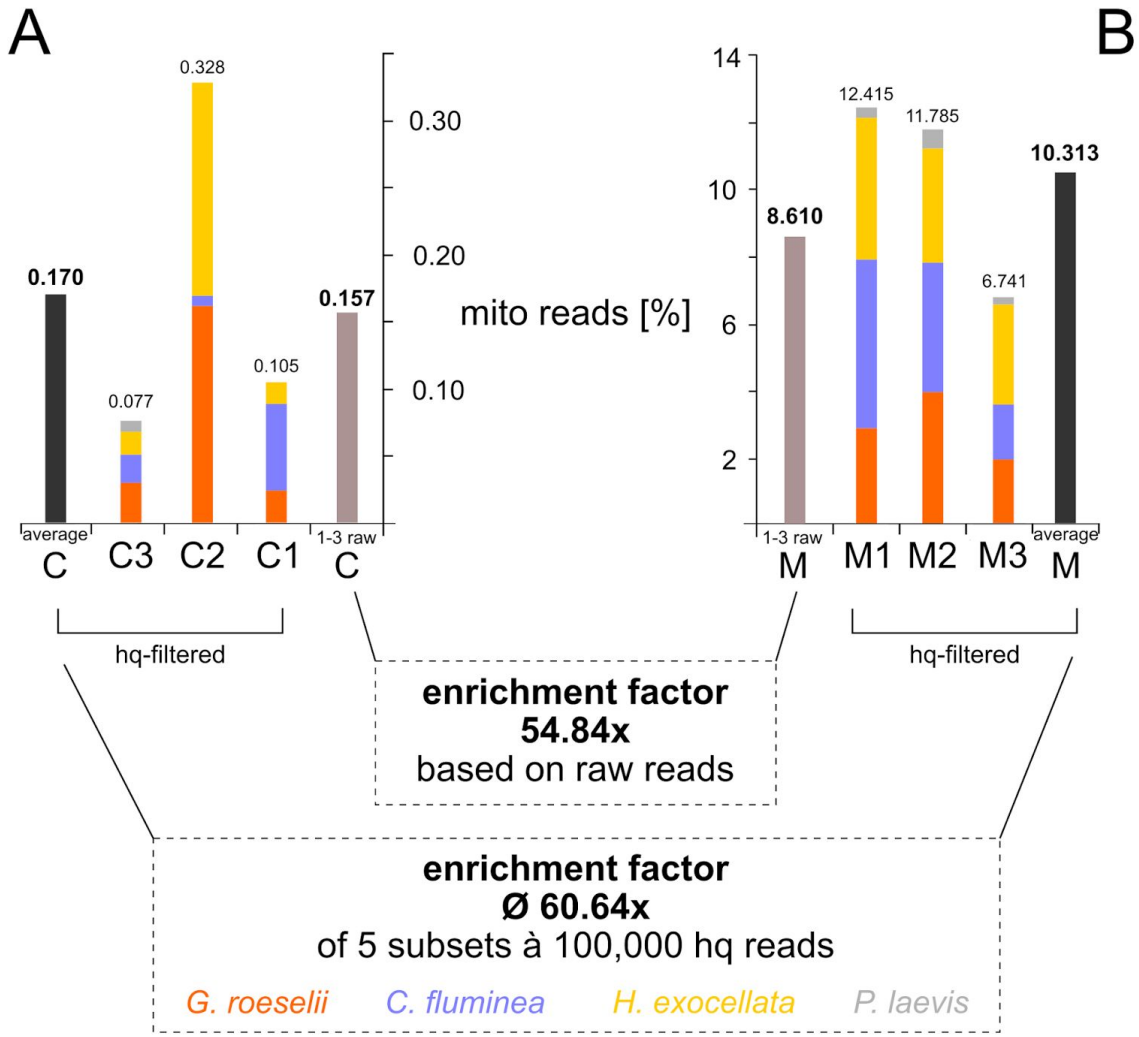

Figure 1: A) Percentage of mitochondrial reads in samples not enriched for mitochondria (C1 - C3). Average C = average of hq mitochondrial reads. 1-3 raw C = percentage of mitochondrial reads in raw reads from samples not enriched for mitochondria. B) Percentage of mitochondrial reads in samples enriched for mitochondria (M1 - M3, averages = M). Colors correspond to the four sequenced species. Average M = average of hq mitochondrial reads. 1-3 raw M = percentage of mitochondrial reads in raw reads from samples enriched for mitochondria. Small numbers above bars correspond to percentage of mitochondrial reads. Colors correspond to analysed species.

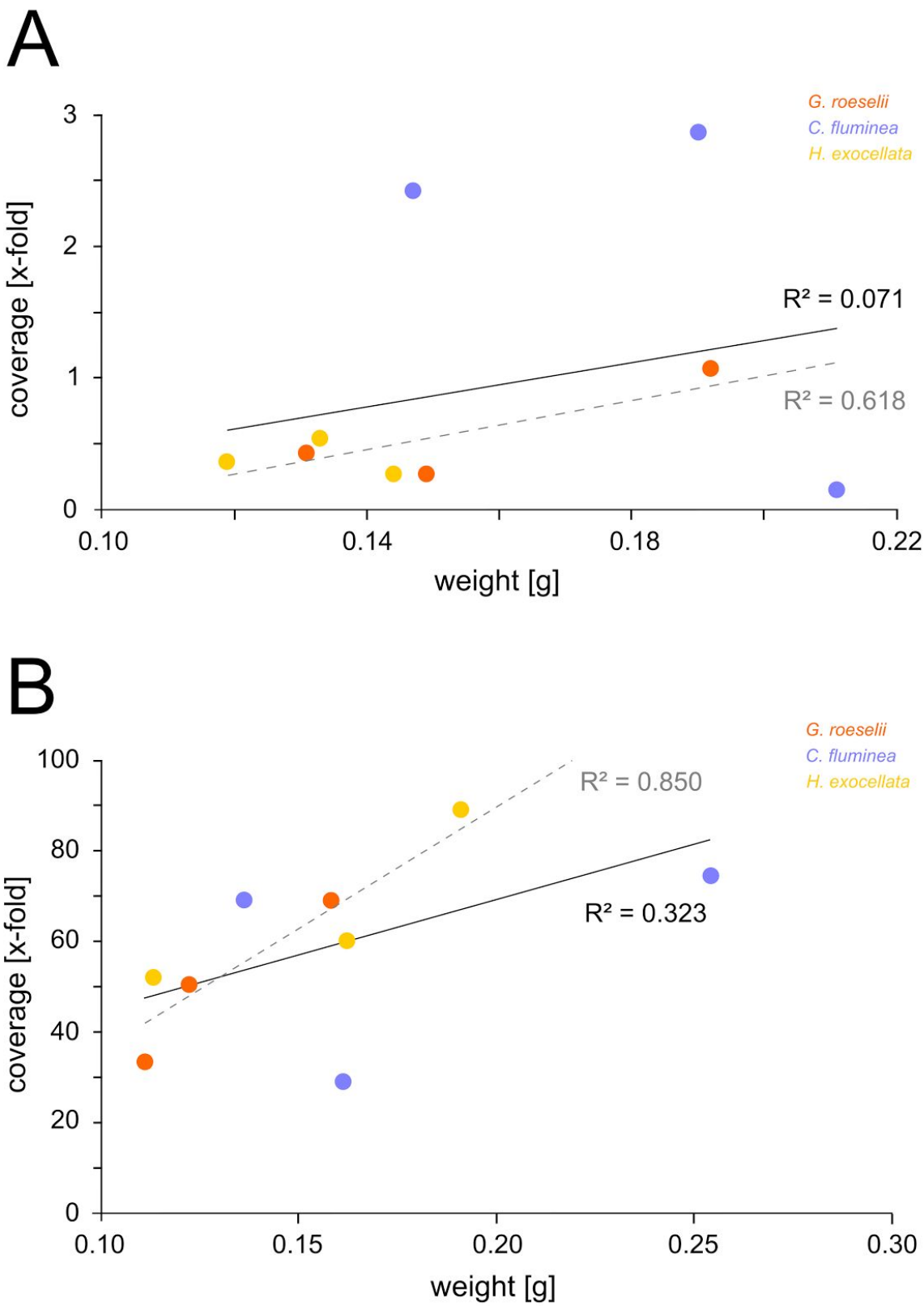

Figure 2: A) Correlation of mitochondrial genome coverage (without repetitive regions) and specimen weight for the three species in samples not enriched for mitochondria. Solid line

and black R^2 value: Correlation with *Corbicula fluminea* included in analysis. Dashed grey line and grey R^2 value: Correlation with *Corbicula fluminea* excluded from analysis. B) Correlation of mitochondrial genome coverage (without repetitive regions) and species weight in samples enriched for mitochondria