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-based methods are increasingly used to study and monitor biodiversity and DNA (meta)barcoding is readily available for standardised assessments. However, (meta)barcoding suffers from bias introduced by PCR and the use of primers, which can lead to the loss of taxonomic groups. PCR-free techniques such as metagenomics are therefore thought to be more suited for unbiased biodiversity assessments but currently suffer from the problem that reference libraries are poorly equipped with genomic information.
2. The technique of 'mitogenome-skimming' or 'mitogenomics', for which mitochondrial genomes are sequenced, is ideal for bridging the techniques of (meta)barcoding and metagenomics, as reference libraries for mitochondrial genes are relatively well established. 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 an easy to use centrifugation protocol that leads to an 55-fold enrichment of mitochondrial DNA. We sequenced six mock communities, each containing the freshwater invertebrate species *Corbicula fluminea*, *Gammarus roeselii* and *Hydropsyche exocellata*, and could recover the whole mitochondrial genomes of these species with a minimum coverage of 2417-fold. In addition, we were able to sequence the whole mitochondrial genome of the acanthocephalan endoparasite *Pomphorhynchus laevis*, which further strengthens the mitogenomics approach as unexpected taxa, for which primers have not been designed, are often not found by (meta)barcoding approaches.
5. Correlation between body mass and sequence number was moderate. Thereby analyses highlight potential problems with species that have a high weight due to their shells, but produce less than expected mitochondrial reads. The here presented protocol will greatly speed up the building of

reference libraries for whole mitochondrial genomes, as dozens of species could 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 mitochondrial enrichment.

Introduction

Biodiversity is of paramount importance for intact ecosystems and inevitable for human well being. However, it is facing ongoing pressures (Rockström *et al.* 2009) and it is therefore pertinent to assess, protect and regularly monitor global biodiversity. DNA-based approaches are increasingly applied for those reasons, the most prominent ones being DNA barcoding (Hebert *et al.* 2003) and more recently DNA metabarcoding (Hajibabaei *et al.* 2011). The latter technique has been repeatedly refined in recent years (e.g. Leray & Knowlton 2015, Elbrecht & Leese 2015) and is now applicable for the standardised assessment of biodiversity (e.g. Gibson *et al.* 2015, Pansu *et al.* 2015, Hänfling *et al.* 2016), i.e. the parallel identification of hundreds of species by the generation and analysis of millions of sequences. However, metabarcoding suffers from the central drawback of PCR primer bias (Piñol *et al.* 2015, Elbrecht & Leese 2015) thereby leading to unequal sequence amplification, which eventually can cause the complete loss of taxa and does not allow to correlate biomass and read abundance due to PCR stochasticity. In consequence, PCR-free techniques harbor the potential for future comprehensive biodiversity assessments (Zhou *et al.* 2013; Elbrecht & Leese 2015, Tang *et al.* 2014, Crampton-Platt *et al.* 2016, Coissac *et al.* 2016) by circumventing the aforementioned problems (i.e. no taxon-dependent PCR amplification bias and the possibility to correlate read numbers with biomass). In this context, two PCR-free methods have been invented. On the one hand, “bait-fishing” techniques (Dowle *et al.* 2016, Liu *et al.* 2015, Mayer *et al.* 2016) which effectively capture and hence enrich informative sequences for target species, but still suffer from bias introduced by the probe design and potentially lead to the loss of taxa. On the other hand, metagenomic approaches (Linard *et al.* 2015, Coissac *et al.* 2016) relying on the direct analysis of (mito)genomes. Since reference libraries are still largely incomplete for nuclear genomic information, but relatively comprehensive for mitochondrial genes (e.g. cytochrome c oxidase subunit I (COI) for animals), the consequent step towards a PCR-free analyses of biodiversity samples could be seen in mitochondrial metagenomics or mitogenomics (e.g. (Tang *et al.* 2014), (Crampton-Platt *et al.* 2015). This technique enables the comparison of newly generated mitogenomes or genome fragments with reference databases and thereby link genomic information to taxonomic data. However, the currently applied approaches are relatively ineffective in terms of sequencing capacity (and hence costs), with most of the ‘PCR-free’ mitogenomic libraries comprising less than 1 % of sequence reads of mitochondrial origin (Crampton-Platt *et al.* 2016). Yet, the major methodological disadvantage is its great sequencing depth needed (Zhou *et al.*, Crampton-Platt *et al.* 2016), 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 separate DNA according to its AT-content and therefore used for the enrichment of typically AT-rich mitochondrial genomes in biochemical studies (e.g. Garber & Yoder 1983). However, this approach requires a cost-intensive ultracentrifugation step. Further, mitochondria can have highly variable AT-contents, making an extraction from bulk

biodiversity samples in CsCl-gradients less straight-forward. The enrichment of mitochondria is most successful when organelles are intact, e.g. from live tissue by centrifugation (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. Here, we use a mock community of four 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 mitogenomic approaches when a) constructing mitochondrial reference libraries and b) assessing biodiversity by an approach which omits biases introduced by primers, probes and PCR.

Material and Methods

Sampling and laboratory protocol

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 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 S3) and assembled to six “mock“- communities, each containing three individuals of *G. roeselii* and *H. exocellata* and one individual of *C. fluminea*. 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 and Aotsuka 1988) in a mortar and crushed with a pestle until tissue was homogenized (70 strokes each). A total of 2 ml of homogenate was pipetted into a 2 ml eppendorf tube samples were treated after the following centrifugation protocols (4° C, centrifuge eppendorf 5427 R) (see supplement S6 for short protocol).

(1) Samples 1-3 (“Complete”- no enrichment of mitochondria): samples were centrifuged for 1 minute at 1,000g. This step was repeated four times. Final centrifugation was conducted for 10 minutes at 14,000g. 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.

(2) Samples 4-6 (“Mito”- enrichment of mitochondria): samples were centrifuged at 1,000g for 1 minute. Pelletized material was discarded, the supernatant transferred to a new tube and again centrifuged at 1,000g for 1 minute. This step was repeated three times. Samples were then centrifuged for 10 minutes at 14,000g. Supernatant was discarded and pelleted material was diluted in 600 µl TNES buffer.

For each sample 40 µl Proteinase K (300 U/ml) were added, vortexed and the solution 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 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”). 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, 250bp paired end) at GATC-Biotech (Konstanz, Germany).

Sequence analysis

Raw sequences were checked for remaining adapters and trimmed with BBDuk as implemented in Geneious. The complete mitochondrial genomes of *Corbicula fluminea*, *Gammarus roeselii*, *Hydropsyche exocellata* and *Pomphorhynchus laevis* (i.e. the acanthocephalan endoparasite, identified by the produced mito reads) were assembled from raw reads using MIRA 4.0.2 (Chevreux, 2014) as implemented in Geneious 10.0.9 (Settings: Genome, accurate) (Kearse *et al.* 2012). Annotations were performed with the MITOS (Bernt *et al.* 2013) web server and corrected 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) Raw 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; Langmead & Salzberg 2012), 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), 100,000 sequences were randomly selected with usearch and reads then 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 as in (Gómez-Rodríguez *et al.* 2015) with a threshold of <10e-5. This step was conducted six times: (1) With reads of samples enriched for mitochondria against reference genomes including repetitive regions, (2) with reads of samples not enriched for mitochondria against reference genomes including repetitive regions, (3) with reads of samples enriched for mitochondria against reference genomes excluding repetitive regions, (4) with reads of samples not enriched for mitochondria against reference genomes excluding repetitive regions, (5) with all raw reads against reference genomes including repetitive regions and (6) with all raw reads against reference genomes excluding repetitive regions.
4) BLASTn searches were conducted with <10e^-5 against a reference library of all 11 available trichoptera mitochondrial genomes, all 67 Amphipoda mitochondrial genomes and 99 mitochondrial genomes of Heterodonta (Supplements X), i.e. taxa related to the target taxa.

A correlation analysis between body weight of specimens 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 belonging to the samples Complete 1-3 and 6,516,172 to the samples Mito 1-3. 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) (Supplement 1).

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 raw reads from the mitochondrial enrichment samples were mapped against the reference genomes (Tab. 1).

We found a 54.84-fold enrichment for mitochondrial reads when complete raw reads were mapped against reference genomes without and a 54.82-fold enrichment when mapped against reference genomes with repetitive regions (table S1). When only high quality reads per sample were mapped against reference genomes, the enrichment was found to be 61.97-fold.

Mapping of subsets (100,000 reads) of high quality reads showed that samples not enriched for mitochondria (Comp 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 60.64-fold enrichment was achieved by the applied centrifugation protocol (Figure 1, table S6). All mentioned data were calculated under exclusion of repetitive elements (table 2). Analyses including repetitive elements showed a 57.9-fold enrichment of mitochondrial reads (Table S2).

Analyses with blastn against a reference library (Tables 3, S5) of the four assembled genomes showed that 12.456 % of all raw reads in samples enriched for mitochondria were putatively of mitochondrial origin when blasted with <10e^-5. When blasted against the reference genomes without repetitive regions, 8.996 % of all reads were found as potentially mitochondrial. The potential overestimate due to the inclusion of the repetitive regions was 38.46 % (Tables 3, S5).

Mapping of raw reads with blastn against a reference library of 11 trichopteran mitochondrial genomes, 67 Amphipoda mitochondrial genomes and 99 Heterodonta mitochondrial genomes (i.e. relatives of the target taxa) showed that putatively 20.05 % (Mito samples) and 0.35 % (Complete samples) of reads were of mitochondrial origin, respectively. The potential overestimate when compared to reads that mapped against the reference genomes of the four analysed species without repetitive regions was 55.132 % (Mito samples) and 102.312 % (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 only weakly correlated for samples enriched for mitochondria (R^2 = 0.323; p = 0.1106; rho = 0.633; p = 0.0760, Figure 2 B) (Table S4). 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) (Tables S3, S4)

Discussion

We developed and tested an easy-to-use centrifugation protocol for the enrichment of mitochondrial DNA from mock communities of common freshwater species. By using this technique and sequencing 6,516,172 raw reads for samples enriched for mitochondria, the full mitochondrial genomes of the common 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 444x. This highlights the great potential of the applied technique for fast sequencing of whole mitochondrial genomes. Subsequently, those genomes can be used for reference library development (Coissac *et al.* 2016) and biodiversity assessment based on bulk samples. By using our protocol, analysed bulk samples can be expected to be sequenced with a deep enough coverage that allows detecting mitochondrial sequences for many hundreds to thousands of specimens in parallel. However, taxonomic assignment of these reads requires the completion of reliable, well-curated full mitochondrial reference libraries. Our technique for mitochondrial enrichment can greatly speed up this process, as it is expected that full mitochondrial genomes of at least several dozens of species can be obtained in a single MiSeq run 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 an average of 10.31 % high quality mitochondrial reads, while samples not enriched for mitochondria contained averagely 0.17% of high quality mitochondrial reads, corresponding to an 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 showed that the amount of mitochondrial reads in our study was overestimated by 38.46 % in samples enriched for mitochondria, respectively, by 45.08 % in samples not enriched for mitochondria when performing blast searches against reference genomes including repetitive regions such as the control regions of mitogenomes. This value 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 % of reads from samples not enriched for mitochondria were reported as being of putative mitochondrial origin, suggesting an overestimate of at least 55.132 % and 102.312 %, respectively, even though relatives of the non-target parasite species *Pomphorhynchus laevis* were not included in the reference library. The results suggest that estimates of mitochondrial read number or at least of informative reads might have been

overestimated in previous studies, in cases 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 regions 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 biases 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. However, this was not observed for samples not enriched for mitochondria, for which only a weak correlation was detected. This observation might be explained by unequal homogenisation of some specimens (in particular the mussel *Corbicula fluminea*) or biased weight measurements, as the specimens could not be completely dried and thus 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 approx. 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 *Pomphorhynchus* tissue and mitochondria was transferred to the mitochondrial pellet. Correlation of weight and percentage of mitochondrial reads was weaker when *Corbicula fluminea* was included in the analyses. This is possibly due to the shell distorting the correlation, as it is heavy, but does not contain any living tissue with mitochondria. It also highlights the limitations of the technique, as proportion of reads and body mass might not correlate in certain taxa.

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 an easy to use 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 shows that even with an enrichment of mitochondria to 10 % of all sequences, hundreds of specimens could be sequenced and their mitogenomes assembled on a single MiSeq run. The discovery of the acanthocephalan parasite *Pomphorhynchus laevis* further strengthens the approach, as PCR primers 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, thus allowing to process the fresh tissue material needed for high(er) rates of mitochondrial enrichments. The ease of

application in combination with a minimized laboratory workload, greatly reduced costs in comparison to mitogenomic approaches without mitochondrial enrichment and high sequencing coverage per recovered mitogenome renders our mitochondrial enrichment protocol ideal for the fast generation of reference libraries ('genome skimming') and subsequent also for biodiversity assessments.

Acknowledgments
We thank Cristina Hartmann-Fatu for help in the lab and the Herzog-Sellenberg Foundation for financial support.

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**.

Dowle, E.J., Pochon, X., C Banks, J., Shearer, K. & Wood, S.A. (2016). Targeted gene enrichment and high-throughput sequencing for environmental biomonitoring: a case study using freshwater macroinvertebrates. *Molecular ecology resources*, **16**, 1240–1254.

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*.

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.

Gibson, J.F., Shokralla, S., Curry, C., Baird, D.J., Monk, W.A., King, I. & Hajibabaei, M. (2015). Large-Scale Biomonitoring of Remote and Threatened Ecosystems via High-Throughput Sequencing. *PloS one*, **10**, e0138432.

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.

Hänfling, B., Lawson Handley, L., Read, D.S., Hahn, C., Li, J., Nichols, P., Blackman, R.C., Oliver, A. & Winfield, I.J. (2016). Environmental DNA metabarcoding of lake fish communities reflects long-term data from established survey methods. *Molecular ecology*, **25**, 3101–3119.

Hebert, P.D.N., Cywinska, A., Ball, S.L. & deWaard, J.R. (2003). Biological identifications through DNA barcodes. *Proceedings of the Royal Society B: Biological Sciences*, **270**, 313–321.

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.

Leray, M. & Knowlton, N. (2015). DNA barcoding and metabarcoding of standardized samples reveal patterns of marine benthic diversity. *Proceedings of the National Academy of Sciences of the United States of America*, **112**, 2076–2081.

Linard, B., Crampton-Platt, A., Gillett, C.P.D.T., Timmermans, M.J.T.N. & Vogler, A.P. (2015). Metagenome Skimming of Insect Specimen Pools: Potential for Comparative Genomics. *Genome biology and evolution*, **7**, 1474–1489.

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.

Pansu, J., De Danieli, S., Puissant, J., Gonzalez, J.-M., Gielly, L., Cordonnier, T., Zinger, L., Brun, J.-J., Choler, P., Taberlet, P. & Cécillon, L. (2015). Landscape-scale distribution patterns of earthworms inferred from soil DNA. *Soil biology & biochemistry*, **83**, 100–105.

Piñol, J., Mir, G., Gomez-Polo, P. & Agustí, N. (2015). Universal and blocking primer mismatches limit the use of high-throughput DNA sequencing for the quantitative metabarcoding of arthropods. *Molecular ecology resources*, **15**, 819–830.

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 for mitochondria (Complete 1 - 3) and enriched for mitochondria (Mito 1 - 3). Averages of five subsets of 100,000 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	**60.64 x**

Tab. 3 Percentage of reads of putative mitochondrial origin when performing blast search 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 perfoming 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	0.35	0.173	**102.312 %**
Mito	20.05	8.996	**55.132 %**

Figures

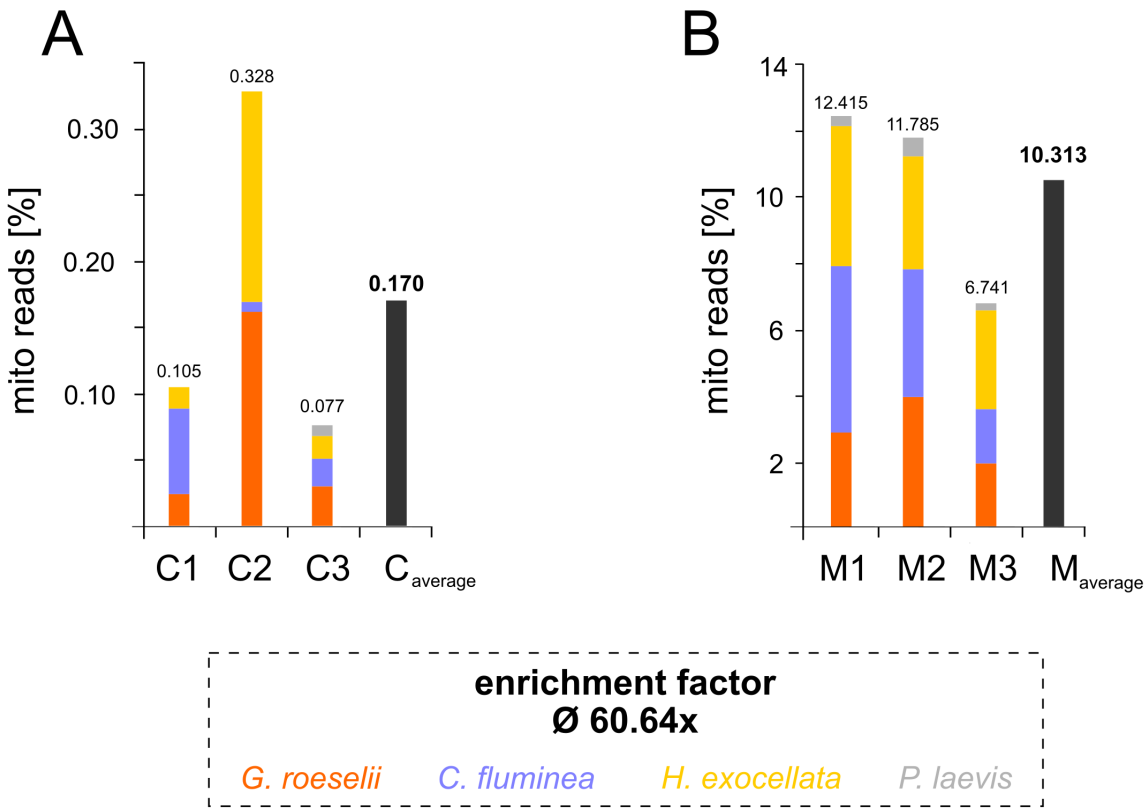

Figure 1: A) Percentage of mitochondrial reads in samples not enriched for mitochondria (C1 - C3, average = C). Colors correspond to the four sequenced species. B) Percentage of mitochondrial reads in samples enriched for mitochondria (M1 - M3, average = M) . Colors correspond to the four sequenced species.

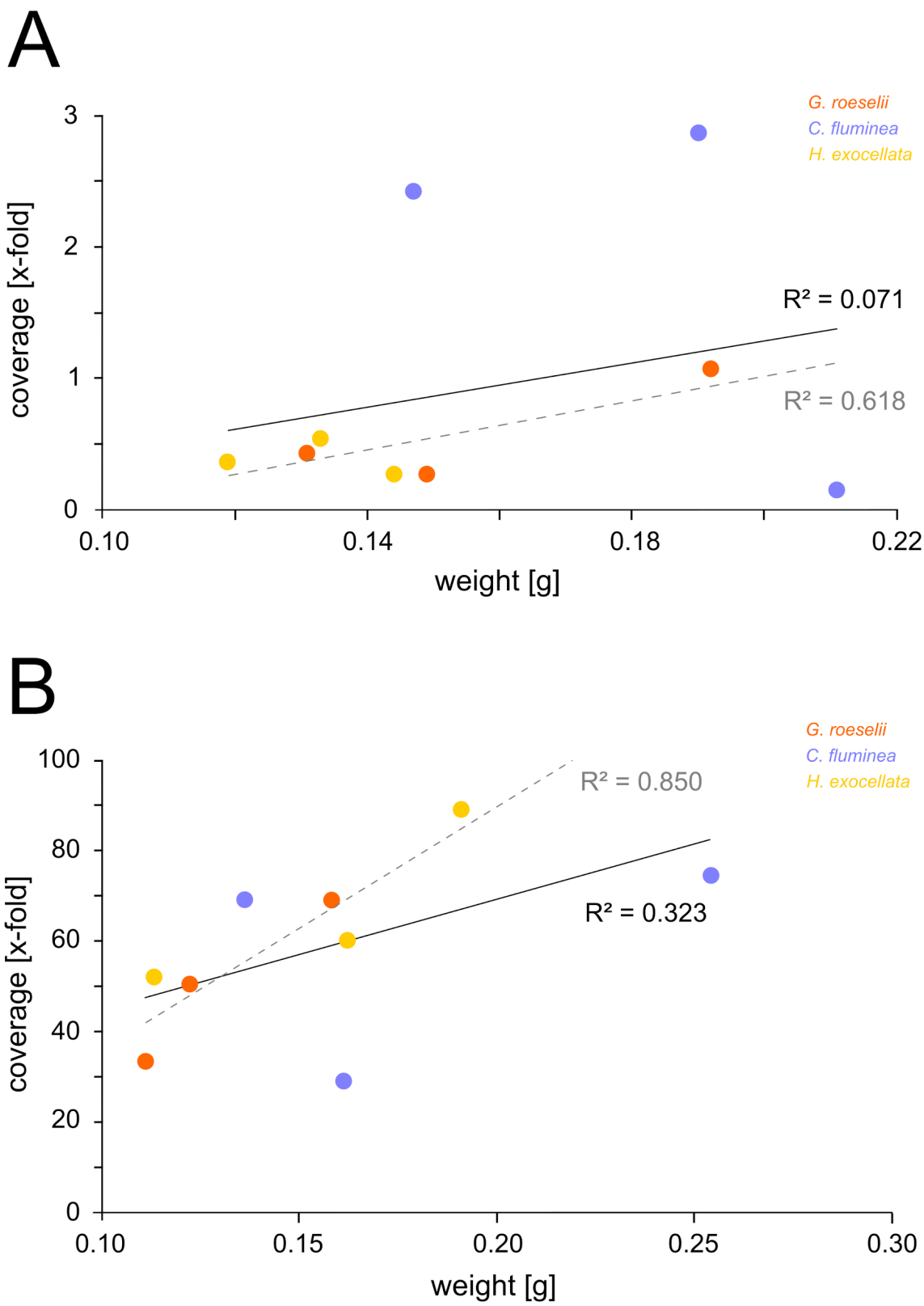

Figure 2: A) Correlation of mitochondrial genome coverage (without repetitive regions) and species weight in samples not enriched for mitochondria. Solid line and black R^2 value: Correlation with *Corbicula fluminea* included in analysis. Dashed grey line and grey R^2 value: Correlation with *Corbicula fluminea* excluded from analysis. B) Correlation of mitochondrial genome coverage (without repetitive regions) and species weight in samples

enriched for mitochondria