Received: 30 March 2020 | Revised: 12 July 2020 | Accepted: 13 July 2020

DOI: 10.1111/1755-0998.13228

RESOURCE ARTICLE

MOLECULAR ECOLOGY RESOURCES WILEY

Optimized and affordable high-throughput sequencing workflow for preserved and nonpreserved small zooplankton specimens

Jannik Beninde[1] | Markus Möst[2] | Axel Meyer[1]

[1]Department of Biology, University of Konstanz, Konstanz, Germany

[2]Department of Ecology, University of Innsbruck, Innsbruck, Austria

Correspondence
Axel Meyer, Department of Biology, Universitätsstraße 10, University of Konstanz, Konstanz, Germany.
Email: axel.meyer@uni-konstanz.de

Funding information
Deutsche Forschungsgemeinschaft, Grant/Award Number: 298726046/GRK2272; Austrian Science Fund, Grant/Award Number: P29667-B25; European Regional Development Fund; Swiss Confederation and Cantons

Abstract

Genomic analysis of hundreds of individuals is increasingly becoming standard in evolutionary and ecological research. Individual-based sequencing generates large amounts of valuable data from experimental and field studies, while using preserved samples is an invaluable resource for studying biodiversity in remote areas or across time. Yet, small-bodied individuals or specimens from collections are often of limited use for genomic analyses due to a lack of suitable extraction and library preparation protocols for preserved or small amounts of tissues. Currently, high-throughput sequencing in zooplankton is mostly restricted to clonal species, that can be maintained in live cultures to obtain sufficient amounts of tissue, or relies on a whole-genome amplification step that comes with several biases and high costs. Here, we present a workflow for high-throughput sequencing of single small individuals omitting the need for prior whole-genome amplification or live cultures. We establish and demonstrate this method using 27 species of the genus *Daphnia*, aquatic keystone organisms, and validate it with small-bodied ostracods. Our workflow is applicable to both live and preserved samples at low costs per sample. We first show that a silica-column based DNA extraction method resulted in the highest DNA yields for nonpreserved samples while a precipitation-based technique gave the highest yield for ethanol-preserved samples and provided the longest DNA fragments. We then successfully performed short-read whole genome sequencing from single *Daphnia* specimens and ostracods. Moreover, we assembled a draft reference genome from a single *Daphnia* individual (>50× coverage) highlighting the value of the workflow for non-model organisms.

KEYWORDS

DNA extraction, high-throughput sequencing, invertebrates, low-input, preserved samples, whole genome sequencing

1 | INTRODUCTION

Small-bodied species are particularly interesting for ecological and evolutionary research as they have the potential to rapidly adapt to environmental change due to higher levels of genetic variation (Ellegren & Galtier, 2016), shorter generation times (Blueweiss et al., 1978), and potentially faster molecular evolution (Martin & Palumbi, 1993; Thomas, Welch, Lanfear, & Bromham, 2010). Additionally, they often play an important role in food webs and ecosystems (Sommer et al., 2012; Sommer, Gliwicz, Lampert, & Duncan, 1986). Most phylogenetic groups have a right-skewed size distribution, meaning they typically have more small-bodied species (Kozłowski & Gawelczyk, 2002). There may also be practical reasons, for example fewer sampling restrictions with invertebrates or lower costs for storage, that can make it easier to work with small organisms.

In the last centuries, over 81 million animal specimens, more than 42 million of which are arthropods, have been collected and preserved by researchers and are stored by institutions all over the world (GBIF.org, 2020). These specimens represent an invaluable source of information on extant biodiversity as well as extinct populations and species. For example, the effect of natural selection on the demise of the Passenger Pigeon (*Ectopistes migratorius*) was reconstructed using mitochondrial and nuclear DNA that could be extracted from museum samples (Murray et al., 2017). Preserved samples can also be used to analyse time series to understand how populations react to environmental change, such as species invasions or increased human disturbance. For example, Hauser, Adcock, Smith, Bernal Ramirez, and Carvalho (2002) used decades-old archived fish scales to demonstrate the loss of genetic diversity as a consequence of human overexploitation of the New Zealand snapper (*Pagrus auratus*). Also today, samples are routinely preserved in ethanol during field trips to remote, inaccessible areas (Camacho-Sanchez, Burraco, Gomez-Mestre, & Leonard, 2013), when it is infeasible to bring individuals back to the laboratory alive or to cryopreserve them in the field.

Recently, high-throughput sequencing techniques have revolutionized the field of biology in general and evolutionary biology in particular (Schuster, 2008). These methods enable more accurate estimation of population structure, gene flow, and genetic variation compared to previous methods that relied on a limited number of markers (Gilbert et al., 2015). The mapping and characterization of genes involved in adaptation is now feasible even for non-model organisms (Ekblom & Galindo, 2011; Stapley et al., 2010).

However, to take full advantage of these exciting new possibilities, adequate quantities of DNA are required for the preparation of adapter-ligated libraries for high-throughput sequencing. The limited amount of available tissue from small-bodied species or valuable museum specimens, currently constrains their usability for cutting-edge genomic technologies. Small amounts of input DNA can be problematic for library preparation. They can lead to an incomplete representation of the genome in the sequencing data when haphazardly parts of the genome are not sufficiently amplified or sequenced due to low initial copy numbers in the input DNA or library, respectively. Moreover, more polymerase chain reaction (PCR) cycles are needed during the library preparation when initial copy numbers are low. This results in an increase of both PCR amplification errors and sequence duplication rates. The former issue affects error rates during genotyping whereas the latter problem increases sequencing efforts and costs as more sequencing is required to compensate for redundant, duplicate reads. Consequently, despite the wealth of samples and their ecological and evolutionary significance many preserved samples are not yet accessible to high-throughput sequencing methods and, hence, the full scientific potential of such collections cannot be utilized (Wandeler, Hoeck, & Keller, 2007).

While improvements of extraction methods specifically designed for museum samples exist, they are designed for dried samples of plants (Staats et al., 2013), avian tissues and egg shells (McCormack, Tsai, & Faircloth, 2016; Tsai, Schedl, Maley, & McCormack, 2019) or insects of comparably larger size (Sproul & Maddison, 2017). High-throughput sequencing of small organisms has so far been performed by including an WGA step (Cruaud et al., 2019; Grealy, Bunce, & Holleley, 2019; Lack, Weider, & Jeyasingh, 2017). This technique enables the use of small amounts of DNA, but introduces biases due to PCR selection, PCR artefacts, and PCR drift (Sabina & Leamon, 2015). Additionally, this extra step adds considerable costs and time. A commonly used alternative to WGA is collecting individuals in the field and establishing clonal (Innes & Ginn, 2014; Schaffner et al., 2019) or large inbred (Benesh, 2019) cultures in the laboratory. While this has enabled the first population genomic study in *Daphnia* (Lynch et al., 2017), culturing lineages in the laboratory has several major disadvantages. It only works for species that are clonal or easy to inbreed and can be kept in a laboratory setting. Furthermore, the survival of individuals in the laboratory and the establishment of clonal lineages is not random and introduces biases. Moreover, mutations can occur in cultures that introduce genetic variation among clonal individuals (Keith et al., 2016), even though this might be negligible over few generations (Dukić, Berner, Haag, & Ebert, 2019). A third alternative for sequencing of small organisms is pooling individuals for sequencing (Pool-seq), which captures allele frequencies but cannot be used for individual-based analyses, e.g., genome-wide association studies or pedigree analyses (Futschik & Schlötterer, 2010). Additionally, due to practical problems in the equimolar pooling of individuals, Pool-seq can suffer from inaccurate calling of rare variants, incorrect allele frequency estimation (Anand et al., 2016), and elevated estimates of population differentiation (Dorant et al., 2019). Hence, until now most population analyses in small-bodied zooplankton have been restricted to Sanger-sequencing-based methods (Koenders, Schön, Halse, & Martens, 2017; Ma, Hu, Smilauer, Yin, & Wolinska, 2019).

It is paramount that techniques of DNA extraction and library preparation are improved to leverage the power of high-throughput sequencing techniques for genomic studies with small-bodied and preserved individuals. Here, we used small bodied (typically <2 mm) aquatic crustaceans (Branchiopoda) of the *Daphnia longispina*-complex (Adamowicz, Petrusek, Colbourne, Hebert, & Witt, 2009;

Schwentner, Combosch, Pakes Nelson, & Giribet, 2017) to compare and modify several different DNA extraction methods in an effort to identify methods yielding high DNA quantities and qualities from small aquatic invertebrates. Further, we investigated the effect of ethanol preservation on extraction success and yield. We then produced a total of 24 high-quality whole genome sequencing (WGS) libraries from minimal DNA amounts without prior whole genome amplification for several *Daphnia* species and ostracods (another group of crustaceans typically between 0.5 and 2 mm in size). Next, we showed that a very basic de novo assembly can already be produced from a single individual sequenced to >50× coverage using our new workflow. We demonstrated the suitability and reliability of our protocol by mapping our libraries to different reference genomes and calculating the concordance of genotyped sites from sequenced libraries from the same individual and extraction. Additionally, we constructed a tree from whole-mitochondrial sequences and performed a principle component analysis on nuclear variants to demonstrate the validity of the generated data. Finally, we combined all these approaches into a workflow (Figure 1) detailing the steps from sample to high-quality sequencing result.

2 | MATERIALS AND METHODS

2.1 | Samples

A total of 27 *Daphnia* species and the ostracod species *Eucypris virens* were used for extractions (Table S1 and Methods S1). For *Daphnia* samples, we measured body length (BL) (henceforth referred to as "BL") of individuals as the distance from the top of the eye to the anterior base of the spine. Additionally, the number of eggs and embryos in the brood pouch ("eggs") was counted from pictures taken prior to extraction. For each extraction, a single individual, either living (henceforth referred to as "nonpreserved") or

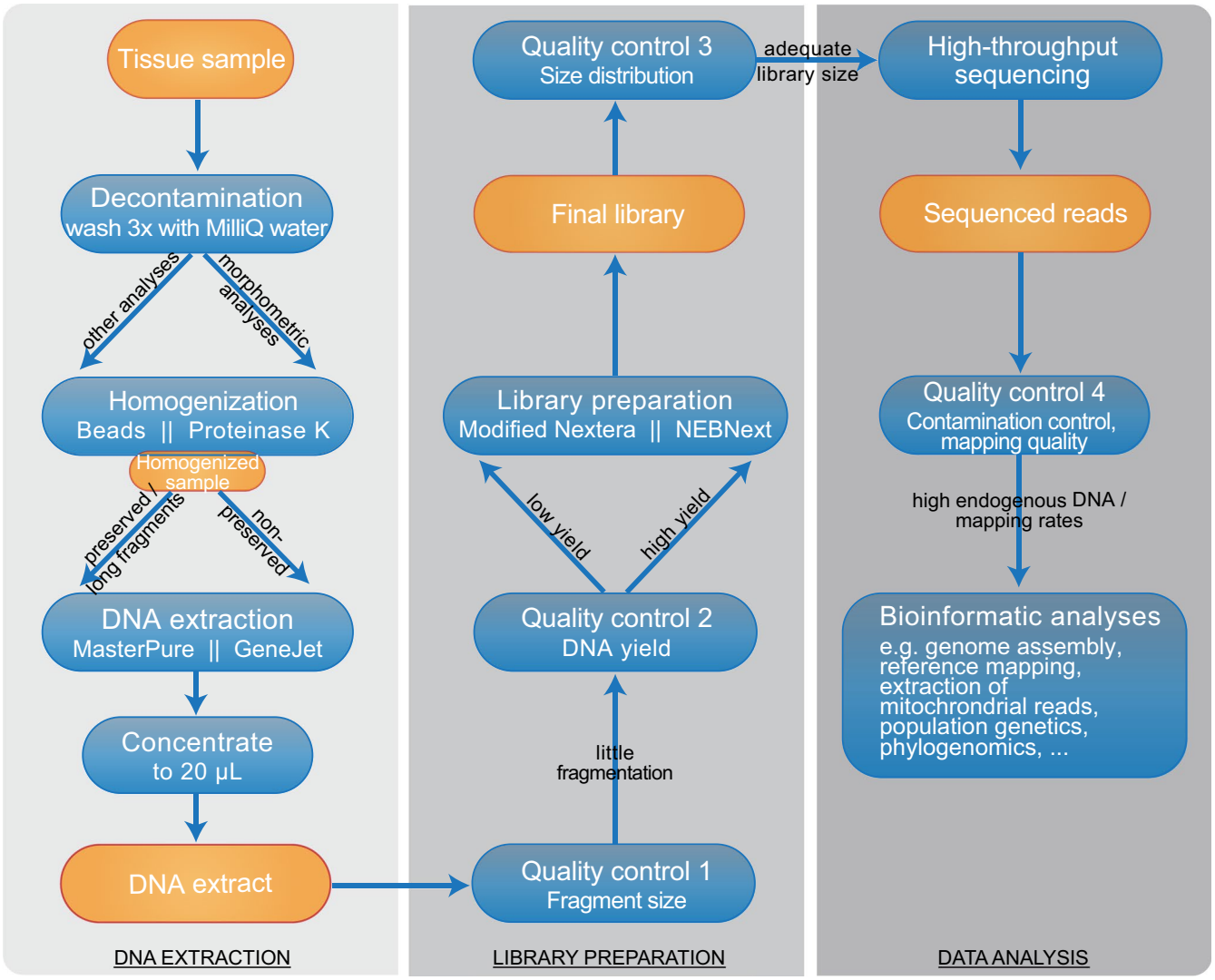


FIGURE 1 Flowchart depicting the workflow presented here. Blue boxes indicate steps of the workflow, and orange boxes show physical objects resulting from these. Several quality control steps are included in the workflow to ensure fragmented and contaminated samples can be removed or dealt with appropriately [Colour figure can be viewed at wileyonlinelibrary.com]

preserved in ethanol ("ethanol-preserved") was used. Nonpreserved samples were either obtained from laboratory cultures or collected from Lake Constance with 200 μm mesh size plankton net hauls. Ethanol-preserved samples comprised various species collected by collaborators and had been stored in ethanol between one and 29 years prior to extraction (Table S1). The exact storage conditions at collection for the samples are not known, but all were stored in >70% ethanol on receipt.

2.2 | Extraction kits and procedure

A few test samples were extracted with a Phenol-Chloroform Isopropanol (PCI) extraction protocol (Green & Sambrook, 2017); however, due to consistently poor results we did not continue with this protocol. Five further protocols (GeneJET Genomic DNA Purification Kit; Thermo Fisher Scientific), QIAamp DNA Micro Kit (Qiagen), Agencourt DNAdvance Kit (Beckman Coulter), MasterPure Complete DNA and RNA Purification Kit (Lucigen), and a modification of the HotSHOT protocol (Truett et al., 2000) by Montero-Pau, Gómez, and Muñoz (2008) were then tested on nonpreserved samples, covering commonly-used approaches for DNA extraction. The two most promising kits were then also tested with ethanol-preserved samples. An overview of all kits including the used protocols is given in Table S2; the modifications and additional information is given in Methods S1.

Before extraction all samples were washed three times in autoclaved Milli-Q water. Following the suggestion of Athanasio, Chipman, Viant, and Mirbahai (2016), individuals were then homogenized with Lysing Matrix D (MP Biomedicals) to break up the carapace. After the extraction, the elution volume was concentrated to 20 μl with a Concentrator Plus (Eppendorf) at 45°C and the V-AQ vacuum setting. DNA concentration was measured using fluorescence-based quantification with a Qubit 2 or 4 (Invitrogen) using the dsDNA HS Assay Kit. DNA fragment length of a subset of samples was measured with a Tapestation 4150 or 4200 (Agilent) using the genomic tape.

2.3 | Library preparation and sequencing

Three different methods were tested for whole-genome library preparation; the original NEBNext Ultra II FS DNA Library Prep Kit for Illumina (NEB) and two modifications of the Illumina Nextera DNA Library Prep Kit (Baym et al., 2015; Therkildsen & Palumbi, 2017). The protocol of Baym et al. (2015) was modified slightly to allow for lower DNA input and to optimize DNA fragment size distribution. Further information on the kits and modifications are available in Methods S1. Final libraries were either size-selected individually at 410–800 bp using a PippinPrep (Biozym Scientific), with a 1.5% cassette and internal markers, or after pooling at BGI Shenzhen. Libraries were then sequenced with unrelated libraries on three HiSeq X10 lanes at BGI Shenzhen in 150 bp paired-end mode. All sequenced libraries are listed with details on DNA input amount and library preparation protocol in Table S3.

2.4 | Analysis

2.4.1 | Extractions

To assess the effects of several different factors (extraction kit, sample preservation, duration of preservation, size of individuals, and number of eggs) on DNA extraction success (success: ≥0.4 ng total DNA; failure: <0.4 ng total DNA) and on DNA yield (ng total DNA), we compared several different combinations of generalized linear models. Models were compared with the Akaike (AIC) and Bayesian information criterion (BIC) and if they supported different models, we chose the model with the lowest BIC score as it should select the correct model when the sample size is much higher than the number of parameters used in the models (Aho, Derryberry, & Peterson, 2014). If the best model included the used extraction kits as a significant explanatory factor, post hoc tests were performed with the R package emmeans (Lenth, 2019) and corrected for multiple testing with Tukey's range test. All analyses were performed in R 3.6.1 (R Core Team, 2019) and the generalized linear models were fitted with lme() from the R package nlme (Pinheiro, Bates, DebRoy, Sarkar, & R Core Team, 2019). All compared models with their respective AIC and BIC values are given in Table S4.

Additionally, differences in the peak of the fragment size distribution between different kits were assessed with a Kruskal-Wallis rank sum test. Extractions of ethanol- and nonpreserved samples using the GeneJET and MasterPure kits were classified as separate groups. Pairwise tests were performed with Wilcoxon signed-rank tests, with correction for multiple testing according to the Benjamini-Hochberg procedure. The PCI extractions were not included in the statistics as sample size was too low.

2.4.2 | Whole genome sequencing

The major steps of all analyses are outlined below, for details see Methods S1. The expected genome sizes of the samples were estimated from the raw reads with Genomescope (Vurture et al., 2017) using the histogram calculated with the count and histo functions from Jellyfish 2.3.0 (Marçais & Kingsford, 2011) with canonical 21-mers. Contamination introduced during the library preparation from various pro- and eukaryotic sources (Methods S1) was estimated from raw reads using Kraken 2.0.9 (Wood, Lu, & Langmead, 2019) and FastQ Screen 0.14 (Wingett & Andrews, 2018). FastQ Screen was also used to test for the presence of endogenous DNA by estimating mapping rates of all libraries against our *Daphnia dubia* assembly (CWD21 v0.01) and against the published *Daphnia*

pulex reference genome (PA42 4.1; Genbank Accession Number GCA_900092285.2). Each sequenced library was individually processed using a GATK workflow. In short, raw reads were transformed to sam files and then Illumina adapters were marked with Picard 2.17.11 (Board Institute, 2019). Next reads were mapped to the *D. pulex* reference genome with BWA mem (Li & Durbin, 2009) and duplicates were marked with Picard. Variants of each library were then called with HaplotypeCaller from GATK 4.1.4 (Poplin et al., 2018). For the variant comparison, the libraries were then genotyped singly with GenotypeGVCF (GATK) and all genotyped positions were filtered and only high-quality SNPs were retained. The number of genotype mismatches between the two technical replicates was obtained with the concordance function of SnpSift 4.3t (Cingolani et al., 2012). For the PCA, prior to genotyping all libraries were combined using CombineGVCFs from GATK. The combined VCF was then sorted using SortVcf (GATK) and filtered with vcftools 0.1.15 to obtain one biallelic SNP every 1,000 basepairs. The PCA was performed with glPCA() from adegenet 2.1.2 (Jombart, 2008; Jombart & Ahmed, 2011) in R.

The mitochondrial sequences of the samples sequenced in this study were assembled using mitobim 1.9.1 (Hahn, Bachmann, & Chevreux, 2013), similar to the approach by Cornetti, Fields, Van Damme, and Ebert (2019). Sequences from *D. pulex* (NCBI Accession Number: AF117817) or *Daphnia galeata* (LC177072) were used as reference for *Daphnia* samples and from *Eucypris virens* (JN618101) for the ostracod samples (Table S3). Then, sequences of the 13 protein-coding genes as well as the 12S and 16S rRNA genes, as identified by MITOS (Bernt et al., 2013), were independently aligned with additional sequences from NCBI (Table S5) using the MAFFT online server with basic settings (Katoh, Rozewicki, & Yamada, 2019). A maximum likelihood tree was reconstructed with IQ-TREE 1.6.9 (Nguyen, Schmidt, von Haeseler, & Minh, 2015). The accuracy of the estimated tree was estimated with 10,000 rounds of both ultrafast bootstrapping (Hoang, Chernomor, von Haeseler, Minh, & Vinh, 2017) and Shimodaira-Hasegawa-like approximate likelihood ratio test (SH-aLRT), which compares the likelihood value of the current tree with that of the best alternative (Guindon et al., 2010). Percent sequence divergence between the whole mitochondrial sequences was estimated from MAFFT alignments in MEGA X (Kumar, Stecher, Li, Knyaz, & Tamura, 2018) using the Maximum Composite Likelihood substitution method.

The basic genome assembly of the *D. dubia* clone dubia_1 (CWD21 v0.01; NCBI Accession Number JAAVJA000000000) was produced with SOAPdenovo2 r241 (Luo et al., 2012) using default settings and reads from both libraries (dubia_1-1 & dubia_1-2). The completeness of BUSCO (Simão, Waterhouse, Ioannidis, Kriventseva, & Zdobnov, 2015; Waterhouse et al., 2018) genes was assessed with QUAST-LG 5.02 (Mikheenko, Prjibelski, Saveliev, Antipov, & Gurevich, 2018). The completeness of the assembly was calculated with KAT (Mapleson, Garcia Accinelli, Kettleborough, Wright, & Clavijo, 2016). Heterozygosity of the dubia_1 clone was estimated using ANGSD (Korneliussen, Albrechtsen, & Nielsen, 2014).

3 | RESULTS

3.1 | Extraction

In total 1,321 single individuals (1,044 nonpreserved and 277 ethanol-preserved samples) were extracted with five different kits (Table S1), of which 103 did not yield measurable DNA quantities (<0.4 ng total DNA). The likelihood of a sample to be successfully extracted was best explained by BL, with the second best explanatory model (ΔBIC > 2.6) supporting both *BL* and sample preservation (Preservation) as significant factors. The best model explaining DNA yield had BL and Preservation as fixed factors and all models that included *Preservation* as factor had lower BIC values than those without (ΔBIC > 15). When comparing nonpreserved and preserved samples separately, the DNA yield of successfully extracted nonpreserved samples (Figure 2a) was best explained by a model that included Kits and BL as fixed factors. Post hoc comparison of the kits showed that GeneJET yielded significantly more DNA than all other kits except for MasterPure. Further, MasterPure, DNAdvance, and Qiagen Micro did not differ from each other significantly. HotShot did not differ significantly from DNAdvance and Qiagen Micro. PCI extraction yielded significantly less DNA than all other kits (all corrected *p*-values < .01). When the number of developing eggs in the brood pouch (*Eggs*) was included as fixed factor in models, the top seven models included *Eggs* as significant factor. For ethanol-preserved samples with > 0.4 ng DNA the best model explaining DNA yield included both factors (*Kits* and *BL)* without interaction. The second best model had only *BL* as significant factor and was slightly less supported (ΔBIC ~ 0.8). MasterPure extractions gave significantly higher DNA yields than GeneJET ($p = .0037$; Figure 2b). All statistical models with their respective AIC and BIC are listed in Table S4.

The fragment sizes of extracted DNA varied significantly between kits ($p = 1.427^{*}10^{-7}$, $df = 5$, Kruskal-Wallis $\chi^2 = 40.098$), with column kit extractions (GeneJET and Qiagen Micro) resulting in the smallest fragments. The fragment length peak of samples extracted with noncolumn kits (DNAdvance and MasterPure) did not differ from each other, but were significantly longer than those extracted with column-based kits. Fragment lengths of ethanol-preserved samples were significantly smaller than those from nonpreserved samples for both GeneJET and MasterPure ($p < .05$ respectively), but MasterPure extractions of ethanol-preserved samples produced significantly larger fragment sizes than GeneJET extractions of nonpreserved samples ($p < .05$). The fragments typically showed a narrow distribution around the peak, even though the proportion of smaller fragments seemed to increase with preservation (Figure S1).

3.2 | Whole genome sequencing

All library preparation methods produced libraries with concentrations and size distributions suitable for sequencing (Figure S2). Libraries produced using the protocol by Therkildsen and

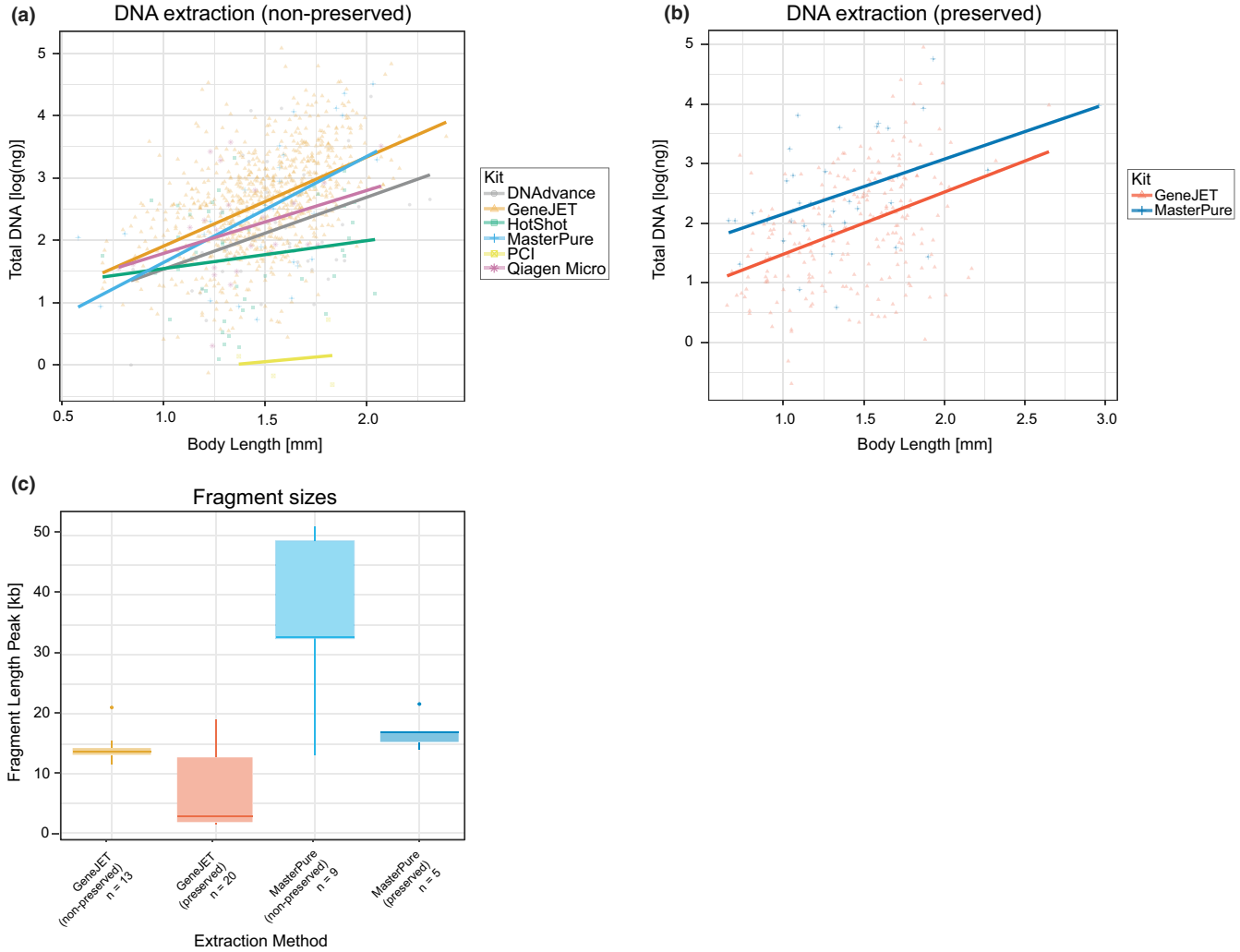


FIGURE 2 Comparison of different DNA extraction kits used for DNA extraction of both nonpreserved and ethanol-preserved individual *Daphnia* specimens. Log-transformed DNA yields of nonpreserved (a) and ethanol-preserved (b) samples in relation to body length, measured from eye to base of spine. Colours indicate different kits. Lines show linear regressions of log-transformed DNA yield onto body length for each kit. (c) Comparison of fragment length peak of ethanol- and nonpreserved samples from MasterPure and GeneJET extractions [Colour figure can be viewed at wileyonlinelibrary.com]

Palumbi (2017) showed an extra peak around 3,000 bp, for which we currently lack an explanation. This protocol also had longer hands-on time due to a second bead clean-up step compared to the protocol by Baym et al. (2015). The NEBNext libraries showed a very broad fragment size distribution, ranging from 150 to 8,000 bp. A total of 27 libraries were sequenced, one of which was produced with the NEBNext Ultra II FS kit and the others with the modified version of the protocol by Baym et al. (2015). We mainly focused on that library preparation method as it uses only 0.35 ng DNA for library preparation. Between 10,658,892 and 434,383,750 reads were obtained from the sequenced libraries, with a median of 31,165,512 reads. Summary statistics for each library are given in Table S3.

To account for technical variation introduced during the library preparation, for 10 *Daphnia* clones two separate libraries (technical replicates) were prepared from the same DNA extraction. The mismatch in genotyped sites between these replicates was calculated as the number of shared genotypes divided by the number genotyped sites. The library replicate pairs prepared from a single extraction had all very similar mapping proportions and very low proportions of mismatches between called genotypes (0.109 ± 0.099%; Table S3). There was a strong and significant negative correlation between genotype mismatch and average coverage ($r^2 = .64$, $p = .0055$).

To test for the presence of endogenous (sample-derived) and exogenous (not sample-derived) DNA in the samples, the mapping rates of sequencing reads to *Daphnia* references and possible contamination sources were estimated. All *Daphnia* libraries, except those of *Daphnia laevis*, had the highest mapping rate of reads to either the *Daphnia pulex* reference genome (PA42 4.1) or the *Daphnia dubia* assembly (CWD21 v0.01), with average mapping rates of 42.23 ± 18.04% (Figure 3). The mapping rate to the *Daphnia* genomes was strongly dependent on phylogenetic distance to the reference genome (Figure S3). Only libraries from samples with complete mitochondrial sequence divergence <30% from a reference had mapping rates above 40% to the respective nuclear

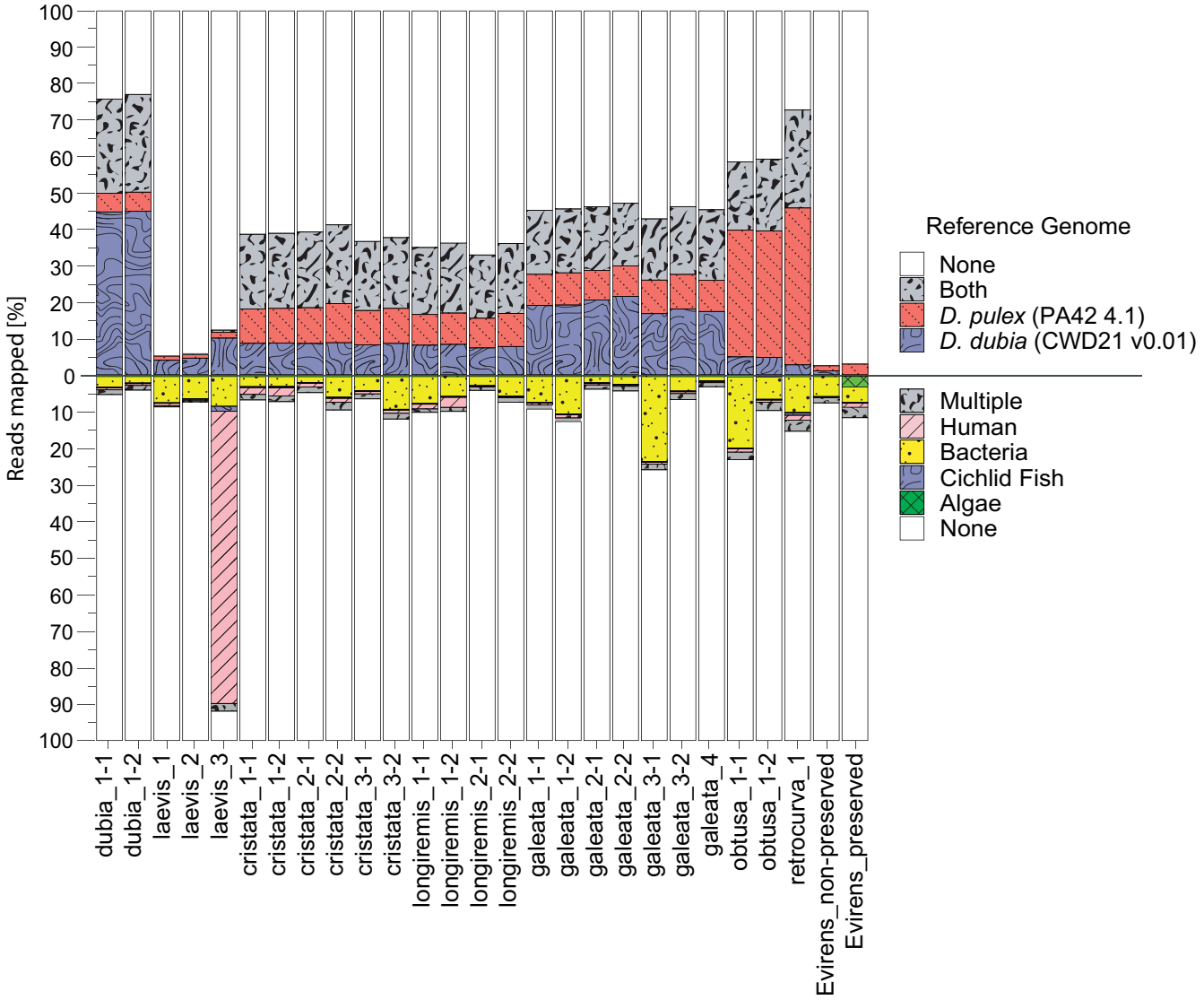


FIGURE 3 Results of contamination assessment of all whole-genome samples. All values were calculated with FastQ Screen, except for bacteria (Kraken2). Samples are given on the *x*-axis, with the *y*-axes indicating mapping rates to either *Daphnia* genomes (upper *y*-axis) or possible sources of contamination (lower *y*-axis) [Colour figure can be viewed at wileyonlinelibrary.com]

assembly. Samples laevis_1 and laevis_2 show little mapping to any of the used reference genomes (<9% mapping), while laevis_3 maps with >80% to the human reference genome. The contamination of all other samples was significantly lower, but still relatively high with 8.89 ± 5.58% on average. The highest contamination rate was 25.64% in sample galeata_3-1. Most contamination belonged to bacterial species (61.8 ± 16.0%).

To test for the completeness of sequencing the nuclear genome in the *Daphnia* samples, we estimated genome sizes of samples. The genome sizes were very similar within species and between 82 and 169 Mbp (Table S3), which is in the same size range as the *D. pulex* assemblies TCO (196 Mbp; Colbourne et al., 2011) and PA42 3.0 (156 Mbp; Ye et al., 2017). Additionally, using the short-read data from the two dubia-1 libraries we created a basic (fragmented and unannotated) assembly (CWD21 v0.01). The assembly had a total size of 98 Mb, falling between the calculated unique (91 Mb; total length of nonrepeated sequences) and haploid length of the sample (Jellyfish: 116 Mb; KAT: 118 Mb). The assembly consists of 11,749 contigs, the largest of which is 178,045 bp long with an N50 of 24,143 bp. The genome has an estimated completeness of 76%, which is slightly higher than the conservative estimate of BUSCO genes present in the assembly (49.83% complete, 19.8% partial). The k-mer spectra representing the assembly had a single peak, due to the low heterozygosity of the sample (0.08%).

To validate our results with external data, we constructed a phylogenetic tree from the individual protein-coding genes of the mitochondria combined with published sequences (Table S5). This analysis allowed us to evaluate whether the sequence data matched the morphological species definition, and how our data compared to the previously published phylogeny of the *Daphnia longispina*-complex (Adamowicz et al., 2009), which is limited to mitochondrial sequences. The mitochondrial tree of *Daphnia* was well resolved

(Figure 4a) and concordant with the tree presented by Adamowicz et al. (2009). All samples grouped according to the prior morphological species identification and all library pairs clustered together. In the mitochondrial tree of the ostracod family Cyprididae (Figure S4) both sequenced samples of *Eucypris virens* (Evirens_preserved & Evirens_nonpreserved) grouped with the reference sequence of *E. virens* with high support.

To assess the nuclear variants, which are more difficult to call due to their lower sequencing coverage compared to mitochondrial variants, a principal component analysis (PCA) was performed for the *Daphnia* samples. Similar to the phylogenetic tree this analysis should assess whether technical replicates are consistent and whether species and species-complexes can be resolved. In the PCA (Figure 4b), the first principal component explained 47.5% of the total variation and clearly differentiated species complexes within the *D. longispina* group sensu lato (*D. laevis*, *D. longispina*, and *Daphnia longiremis*-complexes). On the second principle component, explaining 21.9% of the variation, the *Daphnia pulex* group sensu lato and *D. longispina* group sensu lato were separated. The PCAs of *Daphnia galeata* and *Daphnia cristata* showed that the technical replicates are always closest to each other and that all samples could be clearly separated in PCA 1 and 2 (Figure S5).

4 | DISCUSSION

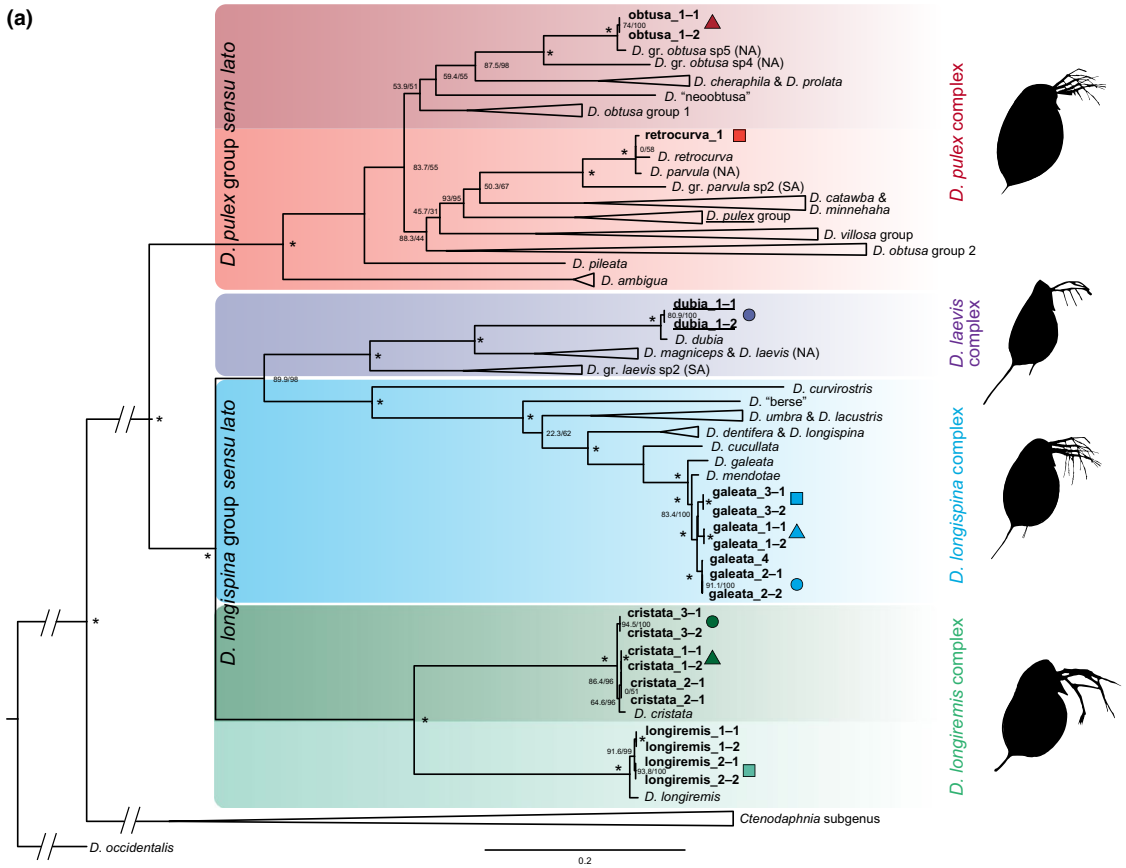
Genome-wide high-throughput sequencing of single individuals offers not only large improvements, such as better phylogenomic estimation, over previous techniques with fewer markers (Gilbert et al., 2015), it is also the basis for many analyses such as GWAS or QTL mapping (Korte & Farlow, 2013). Using single individuals instead of pooled samples improves estimates of allele frequencies (Dorant et al., 2019), aids the identification of genes associated with environmental variation (Rellstab, Gugerli, Eckert, Hancock, & Holderegger, 2015) or phenotypes (Kratochwil, Urban, & Meyer, 2019), and the identification of population structure (Ekblom & Wolf, 2014). Preserved samples from archives and collections stored in museums, institutes, or universities, offer vast opportunities for phylogenomic analyses (Evans et al., 2019) or to study temporal changes. Time series allow quantification of the effects of environmental variation or the strength of selection (Hauser et al., 2002; Schraiber, Evans, & Slatkin, 2016), the investigation of extinct taxa (Murray et al., 2017; Shapiro et al., 2002), and can lead to the identification of new species (Thandar, 2018), or clarification of species status which is relevant for conservation (Montano et al., 2018).

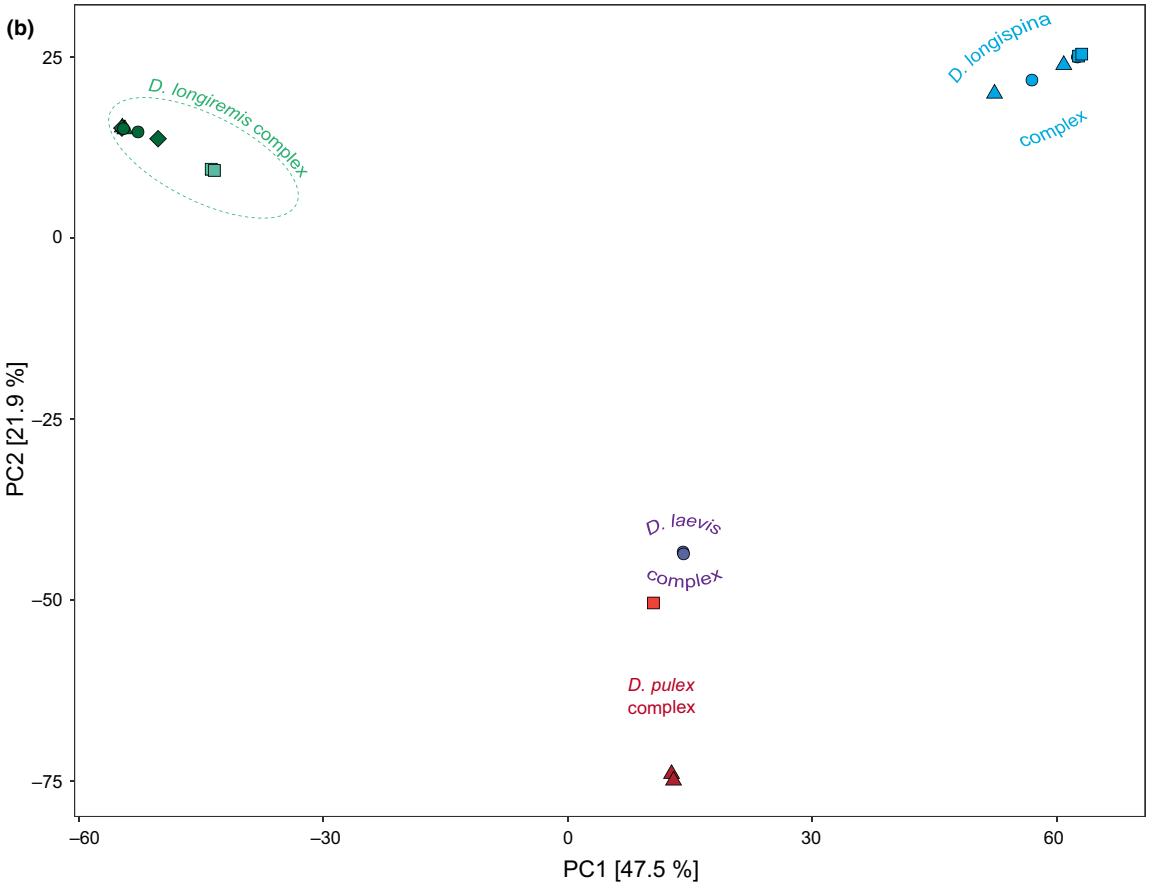
Despite these numerous research opportunities, samples of small-bodied individuals or museum samples are strongly underutilized (Cruaud et al., 2019; Derkarabetian, Benavides, & Giribet, 2019) for approaches using high-throughput sequencing techniques due to difficulties in extracting sufficient amounts of high-quality DNA (Grealy et al., 2019; Staats et al., 2013). In this study, we test which DNA extraction methods are best suited for different downstream applications and how sample preservation impacts the results. We successfully extract DNA from individual *Daphnia* and Ostracods from fresh material as well as specimens stored in ethanol for up to 29 years. Moreover, by reducing the required DNA input to 0.35 ng our workflow allows WGS without the need for whole genome amplification (Cruaud et al., 2019; Lack et al., 2017), cultures in the laboratory (Cornetti et al., 2019; Lynch et al., 2017), pooling of multiple individuals for extraction (Cornetti et al., 2019; Lynch et al., 2017), or using complete specimens for extractions (Scherz et al., 2019). We provide a workflow (Figure 1) that illustrates the process of getting high-quality sequencing results from single small-bodied and preserved samples.

4.1 | DNA extraction

Based on our results, we suggest different approaches depending on the downstream applications (Figure 1). In general, we recommend a homogenization step using a lysing matrix, as it improves DNA yield, for example by breaking the carapace of small crustaceans (Athanasio et al., 2016). However, if morphological features need to be preserved for later analyses, it is advisable to replace the homogenization step with a proteinase K digestion, which chitin exoskeletons can withstand unharmed (Cornils, 2015). For short-read sequencing (e.g., whole genome resequencing) or simple sequence repeat analyses of any type of sample we recommend the GeneJET DNA extraction kit, as it gave the highest DNA yields (Figure 2), had the shortest hands-on time, and lowest price per sample of all tested commercial kits. If the aim is to get the maximum yield from ethanol-preserved samples, MasterPure is better suited, as it produced higher yields from ethanol-preserved samples (Figure 2b). If long reads are needed due to their advantages in the characterization of structural variants and de novo assembly of genomes, we also recommend using MasterPure, which resulted in longer fragment sizes for both preserved and

FIGURE 4 (a) Phylogenetic tree of *Daphnia* based on mitochondrial sequences calculated with IQ-Tree using *Daphnia* samples sequenced in this study (in bold) and reference sequences from Adamowicz et al. (2009). Branches not including any samples sequenced in this study are collapsed. Colours indicate established species-complexes and shadings within a complex indicating different species. Tip labels are given in bold for samples sequenced in this study and species with reference genomes used in Figure 2 (*Daphnia dubia* and *Daphnia pulex*) are underlined. Ultra-fast bootstrap values and Shimodaira-Hasegawa-like approximate likelihood ratio test as calculated by IQ-TREE (uf-boot/SH-aLRT) are given, with support values higher than 95 for both shown as asterisks (*). For identical sequences a bootstrap value of 0 is given. (b) A PCA based on nuclear biallelic SNPs from all *Daphnia* samples. The first principal component explains 47.5% and the second 21.9% of the variation in the data. Colours correspond to the colour scheme used in (a). Shapes differentiate between samples within species and are identical between (a) and (b)


(a)
D. pulex group sensu lato
D. longispina group sensu lato
obtusa_1–1
obtusa_1–2
D. gr. obtusa sp5 (NA)
D. gr. obtusa sp4 (NA)
D. cheraphila & D. prolata
D. "neoobtusa"
D. obtusa group 1
retrocurva_1
D. retrocurva
D. parvula (NA)
D. gr. parvula sp2 (SA)
D. catawba & D. minnehaha
D. pulex group
D. villosa group
D. obtusa group 2
D. pileata
D. ambigua
dubia_1–1
dubia_1–2
D. dubia
D. magniceps & D. laevis (NA)
D. gr. laevis sp2 (SA)
D. curvirostris
D. "berse"
D. umbra & D. lacustris
D. dentifera & D. longispina
D. cucullata
D. galeata
D. mendotae
galeata_3–1
galeata_3–2
galeata_1–1
galeata_1–2
galeata_4
galeata_2–1
galeata_2–2
cristata_3–1
cristata_3–2
cristata_1–1
cristata_1–2
cristata_2–1
cristata_2–1
D. cristata
longiremis_1–1
longiremis_1–2
longiremis_2–1
longiremis_2–2
D. longiremis
Ctenodaphnia subgenus
D. occidentalis
0.2
D. pulex complex
D. laevis complex
D. longispina complex
D. longiremis complex


(b)
D. longiremis complex
D. longispina complex
D. laevis complex
D. pulex complex
PC2 [21.9 %]
PC1 [47.5 %]
25
0
−25
−50
−75
−60
−30
0
30
60

nonpreserved samples (Figure 2c). However, due to the high requirements of long read sequencing (minimum 10–100 ng DNA; >20 kb fragment size) only a single extraction out of all the extractions we performed would qualify for long-read sequencing, indicating that larger specimens or further modifications of the extraction methods are needed for this kind of sequencing.

We note that some ethanol-preserved samples yielded very small DNA fragments (Figure 2c) indicating strong degradation and we suggest checking the fragment size distribution prior to library preparation. While we did not test this, these samples could possibly be processed using our workflow, by adapting the fragmentation time of the library preparation. Alternatively, procedures specifically designed for degraded or ancient DNA, such as the use of base-repair enzymes (Carøe et al., 2018; Fulton & Shapiro, 2019; Gamba et al., 2016) could be used. Additionally, while we observed lower DNA quantities from ethanol-preserved samples (Figure 2b), duration of ethanol preservation had no effect, despite using samples that had been stored in ethanol for over 29 years. Storage in 95% ethanol is assumed to preserve samples well (Camacho-Sanchez et al., 2013; Vink, Thomas, Paquin, Hayashi, & Hedin, 2005), but several studies working with ethanol-preserved museum samples have shown that there is a decrease in the recovery of ultra-conserved elements with increasing preservation time (Blaimer, LaPolla, Branstetter, Lloyd, & Brady, 2016; Derkarabetian et al., 2019; McCormack et al., 2016). As the effects of preservation (lower DNA yield, smaller fragment sizes) are not correlated with time of preservation in our study, we speculate that handling before and during extraction are causing the increased fragmentation and that degradation over time is not detectable in this study due to relatively young age of samples. It is worth mentioning that there are commercially available DNA preserving solutions (e.g., Zymo DNA/RNA Shield, Monarch DNA/RNA Protection Reagent) and an increasing number of studies that propose to use RNAlater also for DNA preservation (Choo, Leong, & Rogers, 2015; Gray, Pratte, & Kellogg, 2013; Vink et al., 2005) due to its DNA-preserving properties.

4.2 | Whole genome sequencing

We present an improved protocol for the Nextera library preparation kit, that facilitates working with very small-bodied samples or small amounts of available starting material, for example tissue samples, or skin swabs from amphibians. To the best of our knowledge, we are the first research laboratory to routinely and successfully use only 0.35 ng of DNA for shotgun whole genome library preparation of small animals, considerably pushing the lower limit for input DNA from the 1 ng of DNA that Sproul and Maddison (2017) have used. We acknowledge that there are other sophisticated methods optimized that deal with low DNA inputs (Shapiro et al., 2019), in particular in the field of ancient DNA analysis working with samples thousands of years old (Willerslev & Cooper, 2005) or in "museomics" which typically deals with younger (up to 200 years) samples. However, these protocols are primarily optimized to deal with poor DNA quality due to contamination and fragmentation (Rohland, Harney, Mallick, Nordenfelt, & Reich, 2015), designed to enrich endogenous DNA over contaminations (Horn, 2012) and therefore target only specific parts of the genome (Knyshov, Gordon, & Weirauch, 2019; Suchan et al., 2016). Many of the protocols also still require higher amounts of input DNA or tissue than used in this study (Gamba et al., 2016; Shapiro et al., 2019; Tsai et al., 2019; Vershinina, Kapp, Baryshnikov, & Shapiro, 2020). Therefore, these methods are more targeted to fragmented DNA and less cost- and time-efficient compared to the presented workflow. The choice of method depends on the sample of interest and biological question and should always be re-evaluated at the different quality control steps (Figure 1).

We use technical replicates of libraries and a suite of analyses, to comprehensively assess the quality of our workflow. Technical variation introduced during the library preparation seems to have been very small, as the technical replicates from the extraction had very low mismatches in the genotyped sites ($0.11 \pm 0.10\%$) that significantly decreased with mean sequencing depth (Table S3). Our genotype mismatch fits the error rate that would be expected when filtering for a quality score of 30 (99.9% accuracy), as done in this study, and hence offers no indication of additional errors introduced during library preparation. Additionally, differences between the technical replicates could be expected due to heterozygous sites with skewed coverage, and to a lesser degree due to mutations between cells, that will stochastically be called differently between sequencing runs of the same individual. Moreover, errors are introduced during sequencing, especially for longer fragments in the reverse read (Tan, Opitz, Schlapbach, & Rehrauer, 2019), or during analysis, e.g., calling variants from single individuals with GATK rather than with population level sampling (Poplin et al., 2018) or due to hard-filtering of variants, which can lead to different sets of variants being filtered between the two samples.

Contamination levels were assessed, and they were highly variable (2.82%–25.64%; Figure 3), mostly due to bacterial contamination. Somewhat surprisingly, all samples had very low levels of algal contamination (<0.1%) although individuals were neither treated with antibiotics nor Sephadex beads. Such a laborious decontamination procedure is commonly implemented to remove food algae and the associated microbiome (Cooper & Cressler, 2020) in high-throughput sequencing studies on *Daphnia* water fleas (Cornetti et al., 2019; Fields et al., 2018). This additional step requires keeping and treating individuals in laboratory culture over several days prior to extraction. Our simple and time-saving approach of washing the samples in autoclaved water before extraction therefore seems to have reduced most contamination from algae. Mapping to eukaryotic contamination sources was low, with a single exception (laevis_3), which was contaminated with human DNA (>80%), most likely due to a handling error during extraction. Low levels of mapping to distant genomes is expected even in the absence of contamination due to highly and universally conserved regions between genomes, such as the ribosomal 16S and 23S sequences (Isenbarger et al., 2008) or UCEs (Meiklejohn, Faircloth, Glenn, Kimball, & Braun, 2016). In conclusion, we strongly

recommend to follow protocols to reduce contamination, such as using a sterile bench or specific rooms (Fulton & Shapiro, 2019), when working with low amounts of DNA.

We tested if the low input DNA amounts were sufficient to sequence the majority of the sample's genome. Estimated genome sizes for the samples (80–150 Mbp; Table S3) are mostly congruent with previous reports and are comparable to the 180 Mbp genome of *Daphnia pulex* (Ye et al., 2017). The de novo assembly of the sample dubia_1, while being arguably basic, has a size of 98 Mbp and high completeness (BUSCO: 70%; KAT: 76%) and can therefore be used to retrieve a large set of genes, enabling many different downstream applications, such as phylogenomic analyses (Cornetti et al., 2019) or obtaining estimates of population size and genetic variation for conservation biology (McMahon, Teeling, & Höglund, 2014). Even such a basic assembly helps to overcome a common limitation for non-model organisms for which no closely related reference is available, including most invertebrates. While all successful libraries showed the highest mapping rates either to the *Daphnia dubia* (CWD21 v0.01) de novo assembly or the well-resolved *D. pulex* reference (Ye et al., 2017), the mapping rates were rather low (<50%, Figure 3). We attribute this in large parts to the relatively high genetic divergence between the samples and the reference genomes (>30%) as mapping rates improve with reduced mitochondrial genetic divergence (Figure S3). If the low mapping rates were merely a result of low input DNA for the library preparation, we would expect a difference in mapping rate between the Nextera libraries created with 0.35 ng of DNA and the libraries produced with more DNA (galeata_4: 6.38 ng DNA; galeata_2-2:1 ng DNA) which is not the case. As expected, the ostracod samples had low mapping rates to any of the tested references.

To validate our data set against external data, we combined the most complete *Daphnia* phylogeny (Adamowicz et al., 2009), which is based on mitochondrial sequences, with mitochondrial sequences assembled from our samples. In the constructed mitochondrial tree (Figure 4a) the position of the samples matched the a priori classification based on morphological characters showing that it is possible to retrieve the mitochondrial sequence. The same result was achieved for the ostracod samples (Figure S4) for which no other comparable data is available. As whole genome data of only one of the *Daphnia* species (*Daphnia obtusa*) used in this study is available and mapping rates of more distantly-related species are very low the value of a phylogenetic approach for the nuclear variants is strongly reduced. Instead a PCA was used, which separated all species and the corresponding complexes in the first two axes of the PCA (Figure 4b). This approach also allows distinguishing closely related populations of the same species and shows the similarity of the technical replicates (Figure S5). This demonstrates the validity of our proposed method also for nuclear variants, which naturally have a much lower sequencing coverage than mitochondrial sequences. Sequencing to high coverage with more unique reads will facilitate further analysis of closely related populations, such as the *D. cristata* samples used in this study.

Despite the good results obtained by using our method, we advise using more than the minimum amount of 0.35 ng of DNA whenever possible, especially for species with larger genome sizes for which the risk of missing large parts of the genome is increased. Dedicated kits such as NEBNext Ultra II FS can be viable alternatives to the modified Nextera Kit, as its low-cost advantage (Baym et al., 2015; Therkildsen & Palumbi, 2017) diminishes when the amount of input DNA is increased due to more of the relatively expensive transposome being required. While, the Nextera kit is no longer manufactured, there are new protocols available for modifying the current Nextera DNA Flex Kit from Illumina (Gaio et al., 2019), which can be modified in a similar way to the changes proposed here to reduce DNA input below the suggested 10 ng.

In conclusion, in this study, we assessed, compared and optimized previously published methods for DNA extraction, various library preparation methods and their modifications. We suggest different kits depending on the type of starting material. The workflow presented here allows for the cost-efficient use of single individuals of small-bodied organisms collected during field trips or routine sampling, recent or historic, live or preserved samples. Instead of extracting a few loci from these samples using Sanger sequencing, the presented workflow allows extracting genome-wide information via reliable high-throughput sequencing. This is achieved without any laborious and costly intermediate steps, such as whole genome amplification or establishing laboratory cultures. While our protocols were tested and optimized using aquatic invertebrates, there is no reason to assume that similar approaches should not be applicable to other small-bodied taxa. It could be used for small insects, both aquatic and terrestrial, tissue samples of larger specimens when limited tissue is available, e.g., arthropod legs, or if multiple analysis are planned for each specimen. Therefore, we encourage other scientists to use and adapt the workflow we present in this study and to consider the application of high-throughput methods even for samples with limited material and projects with limited funds to take full advantage of the possibilities offered by genome-wide data.

ACKNOWLEDGEMENTS

Many thanks to all members of the Meyer-Laboratory for feedback and especially to Julián Torres-Dowdall and Sina Rometsch for statistical support as well as Paolo Franchini and Alexander Nater for help with the bioinformatics. Additionally, I want to thank very much Jeff Dudycha, Meghan Duffy, Anders Hobæk, Joachim Mergeay, Isabelle Schön, Piet Spaak, and Elena Zuykova for providing samples and Roland Fett for his help during fieldwork. Sabine Urban also helped greatly in the design of the figures. We additionally want to thank the three anonymous reviewers and Annabel Whibley for their helpful comments.

Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – 298726046/GRK2272 (to AM). MM was supported by the Austrian Science Fund (FWF): P29667-B25 and the grant "SeeWandel: Life in Lake Constance -the past, present and future" within the framework of the Interreg V programme

"Alpenrhein-Bodensee-Hochrhein (Germany/Austria/Switzerland/Liechtenstein)" whose funds are provided by the European Regional Development Fund as well as the Swiss Confederation and Cantons. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

AUTHOR CONTRIBUTIONS

J.B., and M.M. planned the study. J.B. collected and analysed the data with assistance from M.M. J.B. wrote the manuscript with improvements from all authors.

DATA AVAILABILITY STATEMENT

WGS sequences info and accession numbers are listed in Table S3. The accession numbers for mitochondrial sequences constructed in this study are given in Table S5. The assembly (CWD21 v0.01) is available on GenBank (JAAVJA000000000).

ORCID

Jannik Beninde https://orcid.org/0000-0001-7521-3105
Markus Möst https://orcid.org/0000-0003-2370-2788

REFERENCES

Adamowicz, S. J., Petrusek, A., Colbourne, J. K., Hebert, P. D. N., & Witt, J. D. S. (2009). The scale of divergence: A phylogenetic appraisal of intercontinental allopatric speciation in a passively dispersed freshwater zooplankton genus. *Molecular Phylogenetics and Evolution*, *50*(3), 423–436. https://doi.org/10.1016/j.ympev.2008.11.026

Aho, K., Derryberry, D., & Peterson, T. (2014). Model selection for ecologists: The worldviews of AIC and BIC. *Ecology*, *95*(3), 631–636. https://doi.org/10.1890/13-1452.1

Anand, S., Mangano, E., Barizzone, N., Bordoni, R., Sorosina, M., Clarelli, F., … De Bellis, G. (2016). Next generation sequencing of pooled samples: Guideline for variants' filtering. *Scientific Reports*, *6*(1), 1–9. https://doi.org/10.1038/srep33735

Athanasio, C. G., Chipman, J. K., Viant, M. R., & Mirbahai, L. (2016). Optimisation of DNA extraction from the crustacean *Daphnia*. *PeerJ*, *4*, e2004. https://doi.org/10.7717/peerj.2004

Baym, M., Kryazhimskiy, S., Lieberman, T. D., Chung, H., Desai, M. M., & Kishony, R. (2015). Inexpensive multiplexed library preparation for megabase-sized genomes. *PLoS One*, *10*(5), e0128036. https://doi.org/10.1371/journal.pone.0128036

Benesh, D. P. (2019). Tapeworm manipulation of copepod behaviour: Parasite genotype has a larger effect than host genotype. *Biology Letters*, *15*(9), 20190495. https://doi.org/10.1098/rsbl.2019.0495

Bernt, M., Donath, A., Jühling, F., Externbrink, F., Florentz, C., Fritzsch, G., … Stadler, P. F. (2013). MITOS: Improved de novo metazoan mitochondrial genome annotation. *Mitogenomics and Metazoan Evolution*, *69*(2), 313–319. https://doi.org/10.1016/j.ympev.2012.08.023

Blaimer, B. B., LaPolla, J. S., Branstetter, M. G., Lloyd, M. W., & Brady, S. G. (2016). Phylogenomics, biogeography and diversification of obligate mealybug-tending ants in the genus *Acropyga*. *Molecular Phylogenetics and Evolution*, *102*, 20–29. https://doi.org/10.1016/j.ympev.2016.05.030

Blueweiss, L., Fox, H., Kudzma, V., Nakashima, D., Peters, R., & Sams, S. (1978). Relationships between body size and some life history parameters. *Oecologia*, *37*(2), 257–272. https://doi.org/10.1007/BF00344996

Camacho-Sanchez, M., Burraco, P., Gomez-Mestre, I., & Leonard, J. A. (2013). Preservation of RNA and DNA from mammal samples under field conditions. *Molecular Ecology Resources*, *13*(4), 663–673. https://doi.org/10.1111/1755-0998.12108

Carøe, C., Gopalakrishnan, S., Vinner, L., Mak, S. S. T., Sinding, M. H. S., Samaniego, J. A., … Gilbert, M. T. P. (2018). Single-tube library preparation for degraded DNA. *Methods in Ecology and Evolution*, *9*(2), 410–419. https://doi.org/10.1111/2041-210X.12871

Choo, J. M., Leong, L. E., & Rogers, G. B. (2015). Sample storage conditions significantly influence faecal microbiome profiles. *Scientific Reports*, *5*(1), 16350. https://doi.org/10.1038/srep16350

Cingolani, P., Patel, V. M., Coon, M., Nguyen, T., Land, S. J., Ruden, D. M., & Lu, X. (2012). Using *Drosophila melanogaster* as a model for genotoxic chemical mutational studies with a new program, SnpSift. *Frontiers in Genetics*, *3*, 35. https://doi.org/10.3389/fgene.2012.00035

Colbourne, J. K., Pfrender, M. E., Gilbert, D., Thomas, W. K., Tucker, A., Oakley, T. H., … Boore, J. L. (2011). The ecoresponsive genome of *Daphnia pulex*. *Science*, *331*(6017), 555. https://doi.org/10.1126/science.1197761

Cooper, R. O., & Cressler, C. E. (2020). Characterization of key bacterial species in the *Daphnia magna* microbiota using shotgun metagenomics. *Scientific Reports*, *10*(1), 652. https://doi.org/10.1038/s41598-019-57367-x

Cornetti, L., Fields, P. D., Van Damme, K., & Ebert, D. (2019). A fossil-calibrated phylogenomic analysis of *Daphnia* and the Daphniidae. *Molecular Phylogenetics and Evolution*, *137*, 250–262. https://doi.org/10.1016/j.ympev.2019.05.018

Cornils, A. (2015). Non-destructive DNA extraction for small pelagic copepods to perform integrative taxonomy. *Journal of Plankton Research*, *37*(1), 6–10. https://doi.org/10.1093/plankt/fbu105

Cruaud, A., Nidelet, S., Arnal, P., Weber, A., Fusu, L., Gumovsky, A., … Rasplus, J.-Y. (2019). Optimized DNA extraction and library preparation for minute arthropods: Application to target enrichment in chalcid wasps used for biocontrol. *Molecular Ecology Resources*, *19*(3), 702–710. https://doi.org/10.1111/1755-0998.13006

Derkarabetian, S., Benavides, L. R., & Giribet, G. (2019). Sequence capture phylogenomics of historical ethanol-preserved museum specimens: Unlocking the rest of the vault. *Molecular Ecology Resources*, *19*(6), 1531–1544. https://doi.org/10.1111/1755-0998.13072

Dorant, Y., Benestan, L., Rougemont, Q., Normandeau, E., Boyle, B., Rochette, R., & Bernatchez, L. (2019). Comparing Pool-seq, Rapture, and GBS genotyping for inferring weak population structure: The American lobster (*Homarus americanus*) as a case study. *Ecology and Evolution*, *9*(11), 6606–6623. https://doi.org/10.1002/ece3.5240

Dukić, M., Berner, D., Haag, C. R., & Ebert, D. (2019). How clonal are clones? A quest for loss of heterozygosity during asexual reproduction in *Daphnia magna*. *Journal of Evolutionary Biology*, *32*, 619–628. https://doi.org/10.1111/jeb.13443

Ekblom, R., & Galindo, J. (2011). Applications of next generation sequencing in molecular ecology of non-model organisms. *Heredity*, *107*(1), 1–15. https://doi.org/10.1038/hdy.2010.152

Ekblom, R., & Wolf, J. B. W. (2014). A field guide to whole-genome sequencing, assembly and annotation. *Evolutionary Applications*, *7*(9), 1026–1042. https://doi.org/10.1111/eva.12178

Ellegren, H., & Galtier, N. (2016). Determinants of genetic diversity. *Nature Reviews Genetics*, *17*(7), 422–433. https://doi.org/10.1038/nrg.2016.58

Evans, B. J., Gansauge, M.-T., Stanley, E. L., Furman, B. L. S., Cauret, C. M. S., Ofori-Boateng, C., … Blackburn, D. C. (2019). *Xenopus fraseri*: Mr. Fraser, where did your frog come from? *PLoS One*, *14*(9), e0220892. https://doi.org/10.1371/journal.pone.0220892

Fields, P. D., Obbard, D. J., McTaggart, S. J., Galimov, Y., Little, T. J., & Ebert, D. (2018). Mitogenome phylogeographic analysis of a planktonic crustacean. *Molecular Phylogenetics and Evolution*, *129*, 138–148. https://doi.org/10.1016/j.ympev.2018.06.028

Fulton, T. L., & Shapiro, B. (2019). Setting up an ancient DNA laboratory. In B. Shapiro, A. Barlow, P. D. Heintzman, M. Hofreiter, J. L. A. Paijmans, & A. E. R. Soares (Eds.), *Ancient DNA* (pp. 1–13). New York, NY: Springer. https://doi.org/10.1007/978-1-4939-9176-1_1.

Futschik, A., & Schlötterer, C. (2010). The next generation of molecular markers from massively parallel sequencing of pooled DNA samples. *Genetics*, *186*(1), 207–218. https://doi.org/10.1534/genetics.110.114397

Gaio, D., To, J., Liu, M., Monahan, L., Anantanawat, K., & Darling, A. E. (2019). Hackflex: Low cost Illumina sequencing library construction for high sample counts. *BioRxiv*, https://doi.org/10.1101/779215

Gamba, C., Hanghøj, K., Gaunitz, C., Alfarhan, A. H., Alquraishi, S. A., Al-Rasheid, K. A. S., … Orlando, L. (2016). Comparing the performance of three ancient DNA extraction methods for high-throughput sequencing. *Molecular Ecology Resources*, *16*(2), 459–469. https://doi.org/10.1111/1755-0998.12470

GBIF.org (2020). *GBIF Home Page*. Retrieved from Global Biodiversity Information Facility website: https://www.GBIF.org

Gilbert, P. S., Chang, J., Pan, C., Sobel, E. M., Sinsheimer, J. S., Faircloth, B. C., & Alfaro, M. E. (2015). Genome-wide ultraconserved elements exhibit higher phylogenetic informativeness than traditional gene markers in percomorph fishes. *Molecular Phylogenetics and Evolution*, *92*, 140–146. https://doi.org/10.1016/j.ympev.2015.05.027

Gray, M. A., Pratte, Z. A., & Kellogg, C. A. (2013). Comparison of DNA preservation methods for environmental bacterial community samples. *FEMS Microbiology Ecology*, *83*(2), 468–477. https://doi.org/10.1111/1574-6941.12008

Grealy, A., Bunce, M., & Holleley, C. E. (2019). Avian mitochondrial genomes retrieved from museum eggshell. *Molecular Ecology Resources*, *19*(4), 1052–1062. https://doi.org/10.1111/1755-0998.13007

Green, M. R., & Sambrook, J. (2017). Isolation of high-molecular-weight DNA using organic solvents. *Cold Spring Harbor Protocols*, *2017*(4), 356–359. https://doi.org/10.1101/pdb.prot093450

Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., & Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Systematic Biology*, *59*(3), 307–321. https://doi.org/10.1093/sysbio/syq010

Hahn, C., Bachmann, L., & Chevreux, B. (2013). Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads—a baiting and iterative mapping approach. *Nucleic Acids Research*, *41*(13), e129. https://doi.org/10.1093/nar/gkt371

Hauser, L., Adcock, G. J., Smith, P. J., Bernal Ramirez, J. H., & Carvalho, G. R. (2002). Loss of microsatellite diversity and low effective population size in an overexploited population of New Zealand snapper (*Pagrus auratus*). *Proceedings of the National Academy of Sciences of the United States of America*, *99*(18), 11742–11747. https://doi.org/10.1073/pnas.172242899

Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q., & Vinh, L. S. (2017). UFBoot2: Improving the ultrafast bootstrap approximation. *Molecular Biology and Evolution*, *35*(2), 518–522. https://doi.org/10.1093/molbev/msx281

Horn, S. (2012). Target Enrichment via DNA hybridization capture. In B. Shapiro, & M. Hofreiter (Eds.), *Ancient DNA: Methods and protocols* (pp. 177–188). New York, NY: Humana Press. https://doi.org/10.1007/978-1-61779-516-9_21

Innes, D. J., & Ginn, M. (2014). A population of sexual *Daphnia pulex* resists invasion by asexual clones. *Proceedings of the Royal Society B-Biological Sciences*, *281*(1788), 20140564. https://doi.org/10.1098/rspb.2014.0564

Institute, B. (2019). *Picard toolkit*. Retrieved from http://broadinstitute.github.io/picard/

Isenbarger, T. A., Carr, C. E., Johnson, S. S., Finney, M., Church, G. M., Gilbert, W., … Ruvkun, G. (2008). The most conserved genome segments for life detection on earth and other planets. *Origins of Life and Evolution of Biospheres*, *38*(6), 517–533. https://doi.org/10.1007/s11084-008-9148-z

Jombart, T. (2008). adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics*, *24*(11), 1403–1405. https://doi.org/10.1093/bioinformatics/btn129

Jombart, T., & Ahmed, I. (2011). adegenet 1.3-1: New tools for the analysis of genome-wide SNP data. *Bioinformatics*, *27*(21), 3070–3071. https://doi.org/10.1093/bioinformatics/btr521

Katoh, K., Rozewicki, J., & Yamada, K. D. (2019). MAFFT online service: Multiple sequence alignment, interactive sequence choice and visualization. *Briefings in Bioinformatics*, *20*(4), 1160–1166. https://doi.org/10.1093/bib/bbx108

Keith, N., Tucker, A. E., Jackson, C. E., Sung, W., Lledo, J. I. L., Schrider, D. R., … Lynch, M. (2016). High mutational rates of large-scale duplication and deletion in *Daphnia pulex*. *Genome Research*, *26*(1), 60–69. https://doi.org/10.1101/gr.191338.115

Knyshov, A., Gordon, E. R. L., & Weirauch, C. (2019). Cost-efficient high throughput capture of museum arthropod specimen using PCR-generated baits. *Methods in Ecology and Evolution*, *10*(6), 841–852. https://doi.org/10.1111/2041-210X.13169

Koenders, A., Schön, I., Halse, S., & Martens, K. (2017). Valve shape is not linked to genetic species in the *Eucypris virens* (Ostracoda, Crustacea) species complex. *Zoological Journal of the Linnean Society*, *180*(1), 36–46. https://doi.org/10.1111/zoj.12488

Korneliussen, T. S., Albrechtsen, A., & Nielsen, R. (2014). ANGSD: Analysis of next generation sequencing data. *BMC Bioinformatics*, *15*(1), 356. https://doi.org/10.1186/s12859-014-0356-4

Korte, A., & Farlow, A. (2013). The advantages and limitations of trait analysis with GWAS: A review. *Plant Methods*, *9*(1), 29. https://doi.org/10.1186/1746-4811-9-29

Kozłowski, J., & Gawelczyk, A. T. (2002). Why are species' body size distributions usually skewed to the right? *Functional Ecology*, *16*(4), 419–432. https://doi.org/10.1046/j.1365-2435.2002.00646.x

Kratochwil, C. F., Urban, S., & Meyer, A. (2019). Genome of the Malawi golden cichlid fish (*Melanochromis auratus*) reveals exon loss of oca2 in an amelanistic morph. *Pigment Cell and Melanoma Research*, *32*(5), 719–723. https://doi.org/10.1111/pcmr.12799

Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution*, *35*(6), 1547–1549. https://doi.org/10.1093/molbev/msy096

Lack, J. B., Weider, L. J., & Jeyasingh, P. D. (2017). Whole genome amplification and sequencing of a *Daphnia* resting egg. *Molecular Ecology Resources*, https://doi.org/10.1111/1755-0998.12720

Lenth, R. (2019). *emmeans: Estimated marginal means, aka least-squares means*. Retrieved from https://CRAN.R-project.org/package=emmeans

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, *25*(14), 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., … Wang, J. (2012). SOAPdenovo2: An empirically improved memory-efficient short-read de novo assembler. *GigaScience*, *1*(1), 18. https://doi.org/10.1186/2047-217X-1-18

Lynch, M., Gutenkunst, R., Ackerman, M., Spitze, K., Ye, Z., Maruki, T., & Jia, Z. (2017). Population genomics of *Daphnia pulex*. *Genetics*, *206*(1), 315–332. https://doi.org/10.1534/genetics.116.190611

Ma, X., Hu, W., Smilauer, P., Yin, M., & Wolinska, J. (2019). *Daphnia galeata* and *D. dentifera* are geographically and ecologically separated whereas their hybrids occur in intermediate habitats: A survey of 44 Chinese lakes. *Molecular Ecology*, *28*(4), 785–802. https://doi.org/10.1111/mec.14991

Mapleson, D., Garcia Accinelli, G., Kettleborough, G., Wright, J., & Clavijo, B. J. (2016). KAT: A K-mer analysis toolkit to quality control NGS datasets and genome assemblies. *Bioinformatics*, *33*, 574–576. https://doi.org/10.1093/bioinformatics/btw663

Marçais, G., & Kingsford, C. (2011). A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics*, *27*(6), 764–770. https://doi.org/10.1093/bioinformatics/btr011

Martin, A. P., & Palumbi, S. R. (1993). Body size, metabolic rate, generation time, and the molecular clock. *Proceedings of the National Academy of*

Sciences of the United States of America, *90*(9), 4087–4091. https://doi.org/10.1073/pnas.90.9.4087

McCormack, J. E., Tsai, W. L. E., & Faircloth, B. C. (2016). Sequence capture of ultraconserved elements from bird museum specimens. *Molecular Ecology Resources*, *16*(5), 1189–1203. https://doi.org/10.1111/1755-0998.12466

McMahon, B. J., Teeling, E. C., & Höglund, J. (2014). How and why should we implement genomics into conservation? *Evolutionary Applications*, *7*(9), 999–1007. https://doi.org/10.1111/eva.12193

Meiklejohn, K. A., Faircloth, B. C., Glenn, T. C., Kimball, R. T., & Braun, E. L. (2016). Analysis of a rapid evolutionary radiation using ultraconserved elements: Evidence for a bias in some multispecies coalescent methods. *Systematic Biology*, *65*(4), 612–627. https://doi.org/10.1093/sysbio/syw014

Mikheenko, A., Prjibelski, A., Saveliev, V., Antipov, D., & Gurevich, A. (2018). Versatile genome assembly evaluation with QUAST-LG. *Bioinformatics*, *34*(13), i142–i150. https://doi.org/10.1093/bioinformatics/bty266

Montano, V., van Dongen, W. F. D., Weston, M. A., Mulder, R. A., Robinson, R. W., Cowling, M., & Guay, P.-J. (2018). A genetic assessment of the human-facilitated colonization history of black swans in Australia and New Zealand. *Evolutionary Applications*, *11*(3), 364–375. https://doi.org/10.1111/eva.12535

Montero-Pau, J., Gómez, A., & Muñoz, J. (2008). Application of an inexpensive and high-throughput genomic DNA extraction method for the molecular ecology of zooplanktonic diapausing eggs. *Limnology and Oceanography: Methods*, *6*(6), 218–222.

Murray, G. G. R., Soares, A. E. R., Novak, B. J., Schaefer, N. K., Cahill, J. A., Baker, A. J., … Shapiro, B. (2017). Natural selection shaped the rise and fall of passenger pigeon genomic diversity. *Science*, *358*(6365), 951–954.

Nguyen, L.-T., Schmidt, H. A., von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution*, *32*(1), 268–274. https://doi.org/10.1093/molbev/msu300

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., & R Core Team (2019). *nlme: Linear and nonlinear mixed effects models*. Retrieved from https://CRAN.R-project.org/package=nlme

Poplin, R., Ruano-Rubio, V., DePristo, M. A., Fennell, T. J., Carneiro, M. O., Van der Auwera, G. A., … Banks, E. (2018). Scaling accurate genetic variant discovery to tens of thousands of samples. *BioRxiv*, 201178, https://doi.org/10.1101/201178

R Core Team (2019). *R: A language and environment for statistical computing*. Retrieved from https://www.R-project.org/

Rellstab, C., Gugerli, F., Eckert, A. J., Hancock, A. M., & Holderegger, R. (2015). A practical guide to environmental association analysis in landscape genomics. *Molecular Ecology*, *24*(17), 4348–4370. https://doi.org/10.1111/mec.13322

Rohland, N., Harney, E., Mallick, S., Nordenfelt, S., & Reich, D. (2015). Partial uracil–DNA–glycosylase treatment for screening of ancient DNA. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *370*(1660), https://doi.org/10.1098/rstb.2013.0624

Sabina, J., & Leamon, J. H. (2015). Bias in whole genome amplification: causes and considerations. In T. Kroneis (Ed.), *Whole genome amplification: Methods and protocols* (pp. 15–41). New York, NY: Humana Press. https://doi.org/10.1007/978-1-4939-2990-0_2.

Schaffner, L. R., Govaert, L., De Meester, L., Ellner, S. P., Fairchild, E., Miner, B. E., … Hairston, N. G. (2019). Consumer-resource dynamics is an eco-evolutionary process in a natural plankton community. *Nature Ecology and Evolution*, *3*(9), 1351–1358. https://doi.org/10.1038/s41559-019-0960-9

Scherz, M. D., Hutter, C. R., Rakotoarison, A., Riemann, J. C., Rödel, M.-O., Ndriantsoa, S. H., … Glaw, F. (2019). Morphological and ecological convergence at the lower size limit for vertebrates highlighted by five new miniaturised microhylid frog species from three different Madagascan genera. *PLoS One*, *14*(3), e0213314. https://doi.org/10.1371/journal.pone.0213314

Schraiber, J. G., Evans, S. N., & Slatkin, M. (2016). Bayesian inference of natural selection from allele frequency time series. *Genetics*, *203*(1), 493–511. https://doi.org/10.1534/genetics.116.187278

Schuster, S. C. (2008). Next-generation sequencing transforms today's biology. *Nature Methods*, *5*(1), 16–18. https://doi.org/10.1038/nmeth1156

Schwentner, M., Combosch, D. J., Pakes Nelson, J., & Giribet, G. (2017). A phylogenomic solution to the origin of insects by resolving Crustacean-Hexapod relationships. *Current Biology*, *27*(12), 1818–1824.e5. https://doi.org/10.1016/j.cub.2017.05.040

Shapiro, B., Barlow, A., Heintzman, P. D., Hofreiter, M., Paijmans, J. L. A., & Soares, A. E. R. (Eds.). (2019). *Ancient DNA: Methods and protocols*. New York, NY: Humana Press https://doi.org/10.1007/978-1-4939-9176-1

Shapiro, B., Sibthorpe, D., Rambaut, A., Austin, J., Wragg, G. M., Bininda-Emonds, O. R. P., … Cooper, A. (2002). Flight of the Dodo. *Science*, *295*(5560), 1683. https://doi.org/10.1126/science.295.5560.1683

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, *31*(19), 3210–3212. https://doi.org/10.1093/bioinformatics/btv351

Sommer, U., Adrian, R., De Senerpont Domis, L., Elser, J. J., Gaedke, U., Ibelings, B., … Winder, M. (2012). Beyond the Plankton Ecology Group (PEG) model: Mechanisms driving Plankton succession. *Annual Review of Ecology, Evolution, and Systematics*, *43*(1), 429–448. https://doi.org/10.1146/annurev-ecolsys-110411-160251

Sommer, U., Gliwicz, Z. M., Lampert, W., & Duncan, A. (1986). The PEG-model of seasonal succession of planktonic events in fresh waters. *Archiv Für Hydrobiologie*, *106*(4), 433–471.

Sproul, J. S., & Maddison, D. R. (2017). Sequencing historical specimens: Successful preparation of small specimens with low amounts of degraded DNA. *Molecular Ecology Resources*, *17*(6), 1183–1201. https://doi.org/10.1111/1755-0998.12660

Staats, M., Erkens, R. H. J., van de Vossenberg, B., Wieringa, J. J., Kraaijeveld, K., Stielow, B., … Bakker, F. T. (2013). Genomic treasure troves: Complete genome sequencing of herbarium and insect museum specimens. *PLoS One*, *8*(7), e69189. https://doi.org/10.1371/journal.pone.0069189

Stapley, J., Reger, J., Feulner, P. G. D., Smadja, C., Galindo, J., Ekblom, R., … Slate, J. (2010). Adaptation genomics: The next generation. *Trends in Ecology and Evolution*, *25*(12), 705–712. https://doi.org/10.1016/j.tree.2010.09.002

Suchan, T., Pitteloud, C., Gerasimova, N. S., Kostikova, A., Schmid, S., Arrigo, N., … Alvarez, N. (2016). Hybridization capture using RAD probes (hyRAD), a new tool for performing genomic analyses on collection specimens. *PLoS One*, *11*(3), e0151651. https://doi.org/10.1371/journal.pone.0151651

Tan, G., Opitz, L., Schlapbach, R., & Rehrauer, H. (2019). Long fragments achieve lower base quality in Illumina paired-end sequencing. *Scientific Reports*, *9*(1), 2856. https://doi.org/10.1038/s41598-019-39076-7

Thandar, A. S. (2018). On some miscellaneous sea cucumbers (Echinodermata: Holothuroidea) in the collections of the South African Museum with three new species. *Zootaxa*, *4532*(1), 57. https://doi.org/10.11646/zootaxa.4532.1.3

Therkildsen, N. O., & Palumbi, S. R. (2017). Practical low-coverage genomewide sequencing of hundreds of individually barcoded samples for population and evolutionary genomics in nonmodel species. *Molecular Ecology Resources*, *17*(2), 194–208. https://doi.org/10.1111/1755-0998.12593

Thomas, J. A., Welch, J. J., Lanfear, R., & Bromham, L. (2010). A generation time effect on the rate of molecular evolution in invertebrates. *Molecular Biology and Evolution*, *27*(5), 1173–1180. https://doi.org/10.1093/molbev/msq009

Truett, G. E., Heeger, P., Mynatt, R. L., Truett, A. A., Walker, J. A., & Warman, M. L. (2000). Preparation of PCR-quality mouse genomic DNA with Hot Sodium Hydroxide and Tris (HotSHOT). *BioTechniques*, *29*(1), 52–54. https://doi.org/10.2144/00291bm09

Tsai, W. L. E., Schedl, M. E., Maley, J. M., & McCormack, J. E. (2019). More than skin and bones: Comparing extraction methods and alternative sources of DNA from avian museum specimens. *Molecular Ecology Resources*, 1–8. https://doi.org/10.1111/1755-0998.13077

Vershinina, A. O., Kapp, J. D., Baryshnikov, G. F., & Shapiro, B. (2020). The case of an arctic wild ass s the utility of ancient DNA for validating problematic identifications in museum collections. *Molecular Ecology Resources*, 131301–9. https://doi.org/10.1111/1755-0998.13130

Vink, C. J., Thomas, S. M., Paquin, P., Hayashi, C. Y., & Hedin, M. (2005). The effects of preservatives and temperatures on arachnid DNA. *Invertebrate Systematics*, *19*(2), 99–104. https://doi.org/10.1071/IS04039

Vurture, G. W., Sedlazeck, F. J., Nattestad, M., Underwood, C. J., Fang, H., Gurtowski, J., & Schatz, M. C. (2017). GenomeScope: Fast reference-free genome profiling from short reads. *Bioinformatics*, *33*(14), 2202–2204. https://doi.org/10.1093/bioinformatics/btx153

Wandeler, P., Hoeck, P. E. A., & Keller, L. F. (2007). Back to the future: Museum specimens in population genetics. *Trends in Ecology and Evolution*, *22*(12), 634–642. https://doi.org/10.1016/j.tree.2007.08.017

Waterhouse, R. M., Seppey, M., Simão, F. A., Manni, M., Ioannidis, P., Klioutchnikov, G., … Zdobnov, E. M. (2018). BUSCO applications from quality assessments to gene prediction and phylogenomics. *Molecular Biology and Evolution*, *35*(3), 543–548. https://doi.org/10.1093/molbev/msx319

Willerslev, E., & Cooper, A. (2005). Ancient DNA. *Proceedings of the Royal Society B: Biological Sciences*, *272*(1558), 3–16. https://doi.org/10.1098/rspb.2004.2813

Wingett, S., & Andrews, S. (2018). FastQ Screen: A tool for multi-genome mapping and quality control [version 2; peer review: 4 approved]. *F1000Research*, *7*(1338), 1–13. https://doi.org/10.12688/f1000research.15931.2

Wood, D. E., Lu, J., & Langmead, B. (2019). Improved metagenomic analysis with Kraken 2. *Genome Biology*, *20*(1), 257. https://doi.org/10.1186/s13059-019-1891-0

Ye, Z., Xu, S., Spitze, K., Asselman, J., Jiang, X., Ackerman, M. S., … Lynch, M. (2017). A new reference genome assembly for the Microcrustacean *Daphnia pulex*. *G3: Genes|genomes|genetics*. *7*, 1405–1416. https://doi.org/10.1534/g3.116.038638

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Beninde J, Möst M, Meyer A. Optimized and affordable high-throughput sequencing workflow for preserved and nonpreserved small zooplankton specimens. *Mol Ecol Resour.* 2020;20:1632–1646. https://doi.org/10.1111/1755-0998.13228