Edinburgh Research Explorer

Refining the evolutionary time machine: An assessment of whole genome amplification using single historical Daphnia eggs

RESOURCE ARTICLE

Refining the evolutionary time machine: An assessment of whole genome amplification using single historical *Daphnia* eggs

Christopher James O'Grady[1,2,3] | Vignesh Dhandapani[3] |
John K. Colbourne[3] | Dagmar Frisch[3,4]

[1]School of Life Sciences, University of Warwick, Coventry, UK

[2]Cell and Gene Therapy Catapult, London, UK

[3]School of Biosciences, University of Birmingham, Birmingham, UK

[4]Leibniz Institute of Freshwater Ecology and Inland Fisheries (IGB), Berlin, Germany

Correspondence
Dagmar Frisch, Leibniz Institute of Freshwater Ecology and Inland Fisheries (IGB), Berlin, Germany.
Email: dagmar.frisch@igb-berlin.de

Abstract

Whole genome sequencing is instrumental for the study of genome variation in natural populations, delivering important knowledge on genomic modifications and potential targets of natural selection at the population level. Large dormant eggbanks of aquatic invertebrates such as the keystone herbivore *Daphnia*, a microcrustacean widespread in freshwater ecosystems, provide detailed sedimentary archives to study genomic processes over centuries. To overcome the problem of limited DNA amounts in single *Daphnia* dormant eggs, we developed an optimized workflow for whole genome amplification (WGA), yielding sufficient amounts of DNA for downstream whole genome sequencing of individual historical eggs, including polyploid lineages. We compare two WGA kits, applied to recently produced *Daphnia magna* dormant eggs from laboratory cultures, and to historical dormant eggs of *Daphnia pulicaria* collected from Arctic lake sediment between 10 and 300 years old. Resulting genome coverage breadth in most samples was ~70%, including those from >100-year-old isolates. Sequence read distribution was highly correlated among samples amplified with the same kit, but less correlated between kits. Despite this, a high percentage of genomic positions with single nucleotide polymorphisms in one or more samples (maximum of 74% between kits, and 97% within kits) were recovered at a depth required for genotyping. As a by-product of sequencing we obtained 100% coverage of the mitochondrial genomes even from the oldest isolates (~300 years). The mitochondrial DNA provides an additional source for evolutionary studies of these populations. We provide an optimized workflow for WGA followed by whole genome sequencing including steps to minimize exogenous DNA.

KEYWORDS
ancient DNA, *Daphnia*, population genomics, SNP analysis, whole genome sequencing

1 | INTRODUCTION

Ancient or historical genomic data from natural populations is a key resource for an in-depth understanding of how organisms adapt to their environment. This is one of the most compelling and challenging tasks in evolutionary ecology, especially regarding the current unprecedented environmental change. A unique approach gaining momentum is the study of propagules of various plant or animal taxa preserved in layered aquatic sediments to reconstruct biological and environmental history (Ellegaard et al., 2020; Orsini et al., 2013). These propagules contain dormant embryos in early development (inside, e.g., eggs, seeds, cysts), providing DNA that is degraded to varying degrees, as well as intact DNA, and allowing the direct observation of evolutionary change across centuries or even millennia (Brede et al., 2009; Cordellier et al., 2021; Frisch et al., 2014; Härnström et al., 2011; Mergeay et al., 2006; Pollard et al., 2003; Weider et al., 1997). The exploitation of such resources together with modern molecular tools, targeting many key members of the aquatic food web, is instrumental in the study of evolutionary processes over thousands of generations in relation to environmental change, and can potentially be performed at genomic resolution of individual isolates.

With recent declines in sequencing costs due to development of high-throughput technologies, whole genome sequencing (WGS) has emerged as an important molecular tool in evolutionary biology and has been applied to a plethora of different biological systems (Dettman et al., 2012; Ellegren, 2014; Hohenlohe et al., 2018; Stiller & Zhang, 2019). WGS allows the analysis of genetic variation at thousands of genomic loci to test relationships between phenotypic and genotypic adaptations in genome-wide association studies (De La Torre et al., 2019; Rajpurohit et al., 2018; Sella & Barton, 2019). At the population level, and in particular if long-term time series data including from ancient DNA are available, WGS can provide invaluable genomic detail, shedding light on evolutionary patterns and processes (Leonardi et al., 2017; Parks et al., 2015).

As one of the notable examples, the population genetics of the ecological and genomic model *Daphnia* (Crustacea, Cladocera) has been studied over historical time frames and associated with changes in the lake environment, genotyping either individual eggs (Brede et al., 2009; Frisch et al., 2014, 2016; Limburg & Weider, 2003; Orsini et al., 2012) or by WGS of pooled egg DNA (Cordellier et al., 2021). WGS of individual eggs from sedimentary archives would allow long-term population genomic studies not only of *Daphnia*, but also of other key members of the aquatic foodweb with diapause stages that are preserved in aquatic sediments (e.g., various cladocerans, copepods, rotifers, large branchiopods, as well as algae or fungi) at high resolution.

A major obstacle for WGS of individual dormant eggs of key zooplankton taxa is their limited cell number, and thus minute amount of DNA. The dormant eggs of several planktonic crustaceans (copepods, cladocerans) contain an embryo in the late blastula or early gastrula stage with between 500 and 3000 cells depending on taxon (von Baldass, 1941; Chen et al., 2018; Reed et al., 2021), and even

fewer in rotifers (18–160 cells, Boschetti et al., 2011). For example, given a haploid genome size of ~200 Mb (Colbourne et al., 2011) and ~1000 cells in a dormant embryo of *Daphnia pulex* (von Baldass, 1941), the DNA content of a triploid embryo can be estimated at ~600 pg. Low-input library methods for WGS of individual zooplankton specimens are available but currently require an input of at least 0.35 ng DNA. While this is a small fraction of the DNA that can be extracted from adult zooplankton, averaging 16 ng per individual (Beninde et al., 2020), it exceeds the total amount of DNA available from dormant material. DNA loss during extraction and purification from historical material can be as high as 49%–90%, depending on DNA fragmentation (Barta et al., 2014). Extracting enough DNA from dormant *Daphnia* embryos thus poses technical limitations that are difficult to overcome, further reducing the likelihood that enough DNA can be extracted from dormant material. The situation is exacerbated for historical dormant *Daphnia* eggs or those of other taxa, due to DNA degradation, posing additional problems for DNA sequencing (Rizzi et al., 2012).

To overcome the problem of DNA limitation in individual eggs, it is possible to combine eggs from individual sediment strata for a pooled sequencing approach. Such a strategy has several main disadvantages (Schlötterer et al., 2014): it is not a suitable method to infer haplotypes and linkage disequilibrium; low-frequency variants can be difficult to distinguish from sequencing errors; and technical difficulties (pipetting, DNA quantification) can result in unequally represented individuals in the DNA pool, especially at small sample sizes. It can also lead to information loss on individual genotypes and accuracy of population genomic parameters such as F_{ST} estimates (Dorant et al., 2019).

An alternative approach to gain sufficient amounts of genetic starting material is by performing whole genome amplification (WGA). This method uses cell material without prior DNA extraction, thus minimizing potential loss of DNA during the extraction process, and amplifies genomic DNA from extremely low starting concentrations in the picogram range. However, a previous study that applied WGA to individual, dormant *Daphnia* eggs had limited success with only one of three eggs producing amplified *Daphnia* DNA (Lack et al., 2018). Multiple displacement amplification (MDA), a widely used PCR-free WGA method, utilizes a high-fidelity φ29 DNA polymerase which extends from hexamer primers that randomly bind to targets across the genomic template (Dean et al., 2001, 2002). This results in the generation of large DNA products, with an average length of ~10 kb (capable of reaching over 100 kb), that can have a strong coverage of the target genome (Blanco et al., 1989; Handyside et al., 2004; Lasken & Egholm, 2003; Paez, 2004). MDA is often favoured over PCR-WGA techniques, for example degenerate oligonucleotide-primed PCR, as PCR-based methods can result in the production of small DNA fragments (>1 kb; Telenius et al., 1992; Wells et al., 1999; Zhang et al., 1992) that contain several nonspecific amplification artefacts (Cheung & Nelson, 1996). Additionally, PCR-WGA methods can show a significant amplification bias towards specific loci, and consequently products may not give a complete coverage of loci (Dean et al., 2002). MDA is highly sensitive and

is particularly vulnerable to DNA contamination, which can compete or co-amplify with the desired DNA template during WGA and cause issues during downstream analyses (Blainey & Quake, 2011; Woyke et al., 2011). Great care must therefore be taken to eliminate sources of contamination during the amplification step.

In contrast to Beninde et al. (2020), who provided a detailed low-input library preparation protocol for adult zooplankton individuals from both fresh and maximally 29-year-old ethanol-preserved material, our goal was to develop an optimized WGA-WGS workflow for several-centuries-old, dormant egg isolates including improved decontamination steps. Extending a study of individual *Daphnia* where WGA of dormant eggs had limited success (Lack et al., 2018), we use recently produced *Daphnia magna* dormant eggs from laboratory cultures (days old) and *Daphnia pulicaria* dormant eggs isolated from lake sediment (between 10 and 300 years old), and compared two commercially available single-cell WGA kits based on MDA technology. We test the success of reducing exogenous DNA through the application of different concentrations and durations of bleach, or several washes with PBS (phosphate-buffered saline) to samples prior to WGA.

In a sequencing experiment, we analyse mapping efficiency and genome-wide read distribution and compare these between species and eggs of various age for a total number of 16 dormant eggs aged up to 300 years old. We compare read distribution patterns, coverage breadth and uniformity, and identify contaminants. Finally, we test the utility of these kits for detecting genomic variants in both nuclear and mitochondrial genomes.

2 | MATERIALS AND METHODS

2.1 | Egg collection

All eggs used for whole genome amplification were isolated from ephippia of two species: *Daphnia magna* Straus, 1820, and *Daphnia pulicaria* Forbes, 1893. For *D. magna*, we used eggs from recently produced ephippia (sexual eggs) that are routinely removed from laboratory cultures maintained in the *Daphnia* facility of the University of Birmingham, UK (DM1-DM7, unknown origin). Ephippia of Arctic, triploid populations of *Daphnia pulicaria* (asexually produced eggs) were collected in 2015 from sediment of two lakes in West Greenland (Kangerlussuaq area). Details on the lakes and sediment dating can be found in Dane et al. (2020). Briefly, we sampled ephippia from sediment corresponding to several historical time periods in two lakes: Lake SS4 (Braya Sø): *c*. 2010, *c*. 1880, *c*. 1720, and Lake SS381: *c*. 2010, *c*. 1840 (Table 1).

2.2 | Pre-WGA preparation and cleaning of *Daphnia* eggs

Eggs were removed from ephippia (decapsulated) and transferred to sterile 1× PBS shortly before use. Decapsulated eggs were inspected under a stereomicroscope to ensure that eggs were in good condition (judged by colour and appearance). Visually undamaged eggs were washed in a 5% or 10% wash solution made

TABLE 1 Species identity and specifics of dormant eggs used as template in whole genome amplification

Sample	Species	Ploidy	Age	Pretreatment	kit	DNA (μg)
DM1	*D. magna*	2*n*	Laboratory	10%, < 2 s	REPLI-g	22.27
DM2	*D. magna*	2*n*	Laboratory	10%, 20 s	REPLI-g	28.66
DM3	*D. magna*	2*n*	Laboratory	10%, < 2 s	REPLI-g	37.93
DM4	*D. magna*	2*n*	Laboratory	5%, < 2 s	REPLI-g	33.58
DM5	*D. magna*	2*n*	Laboratory	5%, 20 s	REPLI-g	22.80
DM6	*D. magna*	2*n*	Laboratory	10%, < 2 s	Trueprime	7.06
DM7	*D. magna*	2*n*	Laboratory	10%, < 2 s	Trueprime	7.48
DP1	*D. pulicaria*	3*n*	~10 years (SS4)	10%, < 2 s	REPLI-g	39.48
DP2	*D. pulicaria*	3*n*	~140 years (SS4)	5%, < 2 s	Trueprime	9.21
DP3	*D. pulicaria*	3*n*	~140 years (SS4)	10%, < 2 s	Trueprime	9.18
DP4	*D. pulicaria*	3*n*	~300 years (SS4)	10%, < 2 s	REPLI-g	42.85
DP5	*D. pulicaria*	3*n*	~300 years (SS4)	10%, < 2 s	REPLI-g	26.50
DP6	*D. pulicaria*	3*n*	~10 years (SS1381)	1× PBS	Trueprime	8.30
DP7	*D. pulicaria*	3*n*	~10 years (SS1381)	1× PBS	Trueprime	6.00
DP8	*D. pulicaria*	3*n*	~180 years (SS1381)	1× PBS	Trueprime	6.64
DP9	*D. pulicaria*	3*n*	~180 years (SS1381)	1× PBS	Trueprime	5.14

Note: Additional information is given on pretreatment (% of the bleach solution and exposure time, or 1× PBS buffer, e.g., "10% < 2 s" indicates exposure for less than 2 s to a 10% bleach solution made from 12% industrial bleach), the applied WGA kit and the total product of WGA-DNA obtained. SS4 and SS1831 are two lakes in West Greenland near Kangerlussuaq (for details see Section 2) from which sediment cores with ephippia were extracted.

from industrial strength (12%) bleach. Exposure to the wash solution was either instantly (<2 s) or for 20 s (Table 1), followed by five separate rinses in sterile 1× PBS to remove any remaining bleach. Alternatively, eggs were washed by five to eight rinses in 1× PBS, by placing a row of PBS droplets on a glass slide, and washing each egg individually by carefully and repeatedly drawing them up with a pipette (sterile tip) in each of the droplets. Rinse controls contained the PBS solution used for the last rinse, while negative controls contained sterile PBS. Bleaching (including the final PBS rinse) was performed in a SCANLAF Mars Safety Class 2 laminar flowhood in sterile conditions. The alternative procedure of PBS rinsing was performed in a clean, dedicated room with thorough bleaching of all surfaces prior to processing eggs. Bleached and rinsed eggs were kept on ice for brief periods in sterile 1× PBS until further processing.

2.3 | Whole genome amplification

WGA was performed using two PCR-free kits: Expedeon TruePrime Single Cell WGA kit (hereafter: TruePrime), and Qiagen REPLI-g Single Cell Kit (hereafter: REPLI-g). Positive controls (extracted *Daphnia* DNA) were included in WGA. Rinse controls and negative controls were included to monitor possible amplification of contaminating DNA. Prior to WGA, egg membranes were pierced with a sterile 10-μl pipette tip to allow exposure of embryonic cells, and kept in the respective amount of 1× PBS required for the first step of the reaction in each test kit. DNA concentration in WGA products was quantified with a microplate Reader (Tecan infinite F200 pro), or a Qubit 2.0 Fluorometer (Invitrogen) and the Qubit dsDNA HS Assay kit (Invitrogen). The size distribution of WGA products was determined by agarose gel electrophoresis to analyse the impact of pretreatment steps (bleaching or washing in PBS) on the DNA fragments produced by WGA.

2.4 | Whole genome library preparation and sequencing

WGA samples DM1–DM7 (*D. magna* eggs) and DP1–DP5 (*D. pulicaria* eggs) were used to prepare single-end (SE) libraries with an insert size of 300 bp using a PCR-free workflow with the KAPA HyperPrep Kit (Roche) following the manufacturer's instructions. Sequencing of 100-bp SE libraries was performed on the Illumina HiSeq2500 platform at the Environmental Omics sequencing facility, University of Birmingham, UK. WGA samples DP6–DP9 (*D. pulicaria* eggs) were used to prepare paired-end (PE) libraries with an insert size of 350 bp with the TruSeq DNA PCR-free gel-free library preparation kit (Illumina) according to the manufacturer's instructions. PE library preparation and sequencing (150 - bp PE libraries) was performed at Edinburgh Genomics, The University of Edinburgh, UK. The raw sequence files are available in the open-access repository Zenodo (O'Grady et al., 2021).

2.5 | Bioinformatic and statistical analysis

All analyses involving R packages were completed with R version 3.6.2 (R CoreTeam, 2019). The nature of the study system with limited access to historical eggs, led to some imbalance of the study design. We therefore refrained from a formal factorial statistical analysis for the comparison of different pretreatments and sequencing strategies.

2.5.1 | Quality control and mapping

We used FASTQC (Andrews, 2015) to check read quality, followed by adapter trimming and removal of leading and trailing low-quality bases with TRIMMOMATIC (Bolger et al., 2014). Following quality control, reads were mapped using BWA-MEM with default settings (Li & Durbin, 2009) to the respective reference genome assembly (*D. magna* genome assembly DAPHMAG2.4, GenBank accession GCA_001632505.1; *Daphnia pulex* genome assembly [http://genome.jgi.doe.gov/Dappu 1/Dappu1.download.html] [Colbourne et al., 2011]; *D. pulex* mitochondrial genome, GenBank Accession NC_000844 [Crease, 1999]). Mapping statistics were computed with QUALIMAP (García-Alcalde et al., 2012) prior to variant calling. Duplicate reads were removed from mapped reads using MARKDUPLICATES from the Picard Toolkit (Broad Institute, 2019).

2.5.2 | Nuclear DNA variant calling and analysis

Nuclear single nucleotide polymorphisms (SNPs) were called in *D. magna* using the available SE libraries. SNPs in the triploid Arctic *D. pulicaria* eggs were called only from PE samples because sequencing depth of SE samples was insufficient for calling variants in a triploid organism (Maruki & Lynch, 2017). Nuclear and mirochondrial variants were called with FREEBAYES version 1.3.2 (Garrison & Marth, 2012), excluding reads with a mapping quality <40, base quality <24 and a minimum alternate allele fraction of 0.01, ploidy = 2 (*D. magna* nuclear DNA [ncDNA], *D. pulicaria* mitochondrial DNA [mtDNA]) and ploidy = 3 in *D. pulicaria* ncDNA. After variant calling, nuclear SNPs were hard-filtered with VCFFILTER (Garrison, 2016) applying all of the following settings: "QUAL > 1" to ensure the exclusion of variants of very low quality, "QUAL/AO > 10" to include only variants where each observation contributes at least 10 log units (~Q10 per read), "SAF > 0 & SAR > 0" to avoid strand bias, and "RPR > 1 & RPL > 1" to require at least two reads on each side of the variant.

Genomic positions with high-confidence SNPs present in one or more samples were compared between selected samples to assess the percentage of loci that could be called in all selected samples (i.e., that were amplified and sequenced at the depth required for genotyping). This comparison was used primarily to estimate the repeatability of WGA and subsequent WGS, and thus for the resulting capacity to call variants at the multisample level. For *D. magna*, we compared the four samples with the highest number of SNPs

(two REPLI-g amplified samples: DM2, DM3, two TruePrime amplified samples: DM6, DM7). For *D. pulicaria*, we compared all four PE samples (only TruePrime amplified). Results were visualized with the R packages EULERR version 6.1.0 (Larsson, 2020) and GGVENNDIAGRAM version 0.3 (Gao & Yi, 2019).

Transition-to-transversion ratios for SNPs (Ti:Tv) were calculated after applying a minor allele frequency (MAF) threshold of 0.05, and not allowing missing data (R packages SEQARRAY 1.26.2, Zheng et al., 2017; and SEQVARTOOLS 1.24.1, Gogarten et al., 2021).

2.5.3 | mtDNA variant calling and analysis

To analyse mtDNA, we used all available *D. pulicaria* samples (DP1–DP9). For mitochondrial SNPs, the same filters as for nuclear SNPs were applied except "QUAL/AO > 10" to avoid filtering calls of the alternate allele from samples with PE sequencing due to their consistently higher depth compared to the SE samples. Identity-by-state (IBS) was calculated by applying an MAF threshold of 0.05, not allowing missing data (R package SEQARRAY 1.26.2, Zheng et al., 2017). SNPs in *D. pulicaria* mtDNA were visualized with the packages CIRCLIZE version 0.4.11 (Gu et al. 2014), and SNPRELATE (Zheng et al., 2021).

2.5.4 | Read distribution

This analysis focused on reads mapped to the N50 scaffolds of the *D. magna* and *D. pulex* genomes. Coverage depth was normalized between samples of binned reads (bin size 10 or 100 kb, normalized reads = number of reads per bin/average number of reads across bins). Normalized read coverage was visualized with the packages CIRCLIZE version 0.4.11 (Gu, 2014), and GGPLOT2 version 3.3.2 (Wickham, 2016). Read distribution was compared within species between samples by correlation analysis (Pearson's correlation coefficient) of normalized binned reads. For this purpose, we removed a single outlier present in all *D. pulicaria* samples (position 420,001–430,000, scaffold 38). We tested uniformity of read distribution according to the standard model for random sequencing by fitting the distribution of normalized read coverage to a Poisson distribution (Lander & Waterman, 1988). Uniformity of read distribution was quantified by the evenness score (Oexle 2016). In short, this metric is calculated as the coefficient of variation for non-normalized data.

2.5.5 | Outlier identification and exogenous DNA

The REPLI-g amplified *D. pulicaria* samples DP1, DP4 and DP5 contained obvious outliers with preferential amplification. They were defined as regions with a mapping rate 10 times higher than the mean normalized count (100-kb bins). Sequences belonging to these outlier regions were extracted and searched against the nucleotide database with BLASTN. Exogenous DNA was identified in two samples with low mapping efficiency (DP4, DP5). For this, unmapped

reads were called from bam files with SAMTOOLS version 1.4 (Li & Durbin, 2009) and converted to PE fastq files using BEDTOOLS version 2 (Quinlan & Hall, 2010). Following this step, SOAPDENOVO-127MER version 2 (Luo et al., 2012) was used to *de novo*-assemble the unmapped reads with kmer size of 23 bp and default parameters. The resulting contigs were searched against the NCBI nucleotide database with BLASTN (using the command line "*blastn -task megablast -db NCBI_nt_db -query infile -evalue 1e-100 -out outfile -max_target_seqs 1 -num_threads 10 -outfmt "6 qseqid sseqid sciname qlen slen qstart qend sstart send length evalue pident nident mismatch gaps*"). For graphical representation we used KRONA (Ondov et al., 2011).

3 | RESULTS

3.1 | Whole genome amplification

WGA products were separated by agarose gel electrophoresis to determine the impact of pretreatment steps (bleaching or washing in PBS) on WGA product fragment sizes. Regardless of the pretreatment process, strong bands for DNA fragments >10 kb were detected in all samples, suggesting a minimal impact of pretreatment on WGA product size (examples in Figure S1). WGA products obtained from one of the rinse controls (1× PBS, DNA amplified from the last PBS wash of an unbleached sample) also produced high-intensity bands (from 500 bp to >10 kb). This DNA probably resulted from amplified contaminant DNA carried over from one of the previous serial washes, suggesting that PBS washes do remove exogenous DNA from egg surfaces, but that a higher number of washes may be advisable for more complete removal. WGA of rinse controls obtained from the PBS after bleach-washing did not yield any product, suggesting that the application of bleach completely removes external, exogenous DNA. No amplified DNA was present in any of the negative controls (sterile PBS).

Mean WGA-DNA concentration was lower in the samples amplified by TruePrime (7.38 μg) in comparison with REPLI-g (31.76 μg; Table 1). These values were within the ranges suggested by the respective WGA kit manufacturers (~40 μg WGA-DNA for REPLI-g, and 3–4 μg when starting from a single cell for TruePrime).

3.2 | Read mapping to nuclear and mitochondrial reference genomes

Both SE (DP1–DP5) and PE (DP6–DP9) sequencing was performed for *D. pulicaria*, the latter with about 10-fold higher read numbers and related higher coverage depth (Table 2). For *Daphnia magna*, only SE libraries were sequenced (DM1–DM7). With one exception (DM4), all *D. magna* libraries mapped to the reference genome with high efficiency >98%, indicating no clear pattern of mapping efficiency to the nuclear genome related to the applied pretreatment in the several-days-old dormant eggs produced in cultures (DM1–DM7). Lower mapping efficiency and/or a smaller fraction of

TABLE 2 Mapping details and number of variants called in (a) nuclear DNA of *Daphnia magna* (DM), and *Daphnia pulicaria* (DP), and (b) mitochondrial DNA of *D. pulicaria*

Nuclear genome *D. magna* and *D. pulicaria*

Sample	Age	Pretreatment	Kit	Seq	Number of reads	Mapped reads (%)	Coverage breadth (%)	Coverage depth	Evenness score	SNP loci
DM1	Laboratory	10%, <2 s	R-g	SE	8,488,215	98.72	74.95	6	0.76	301,779
DM2	Laboratory	10%, 20 s	R-g	SE	17,742,882	98.76	77.43	13	0.76	414,782
DM3	Laboratory	10%, <2 s	R-g	SE	24,069,767	98.74	78.63	18	0.75	430,909
DM4	Laboratory	5%, <2 s	R-g	SE	38,765	55.73	1.47	0.02	0.55	NA
DM5	Laboratory	5%, 20 s	R-g	SE	10,167,137	98.64	76.36	7	0.77	339,023
DM6	Laboratory	10%, <2 s	TP	SE	6,392,740	99.18	66.75	5	0.68	99,927
DM7	Laboratory	10%, <2 s	TP	SE	14,800,247	99.04	78.05	11	0.71	357,847
DP1	~10 years	10%, <2 s	R-g	SE	12,990,625	48.36	41.45	3	0.38	NA[a]
DP2	~140 years	5%, <2 s	TP	SE	12,088,977	87.36	50.11	5	0.62	NA[a]
DP3	~140 years	10%, <2 s	TP	SE	16,643,733	84.62	51.82	7	0.66	NA[a]
DP4	~300 years	10%, <2 s	R-g	SE	8,579,357	0.36	0.33	0.01	0.23	NA[a]
DP5	~300 years	10%, <2 s	R-g	SE	7,950,283	2.55	0.93	0.07	0.16	NA[a]
DP6	~10 years	1× PBS	TP	PE	85,442,043	90.53	69.21	51	0.74	1,732,886
DP7	~10 years	1× PBS	TP	PE	134,031,117	90.95	69.58	78	0.74	1,734,987
DP8	~180 years	1× PBS	TP	PE	95,665,657	91.65	69.50	58	0.73	1,742,463
DP9	~180 years	1× PBS	TP	PE	93,262,683	91.85	69.55	57	0.73	1,741,613

(b) Mitochondrial genome *D. pulicaria*

Sample	Age (years)	Seq	Mapped reads (%)	Coverage breadth (%)	Coverage depth	SNP loci
DP1	~10	SE	73.87	99.97	61,542	200
DP2	~140	SE	0.68	99.94	498	200
DP3	~140	SE	0.25	99.93	252	200
DP4	~300	SE	0.13	99.93	71	200
DP5	~300	SE	4.55	99.95	2315	200
DP6	~10	PE	3.32	99.97	25,740	200
DP7	~10	PE	0.86	99.95	10,240	200
DP8	~180	PE	0.72	99.97	6155	200
DP9	~180	PE	0.62	99.95	5197	200

Note: Age = see Table 1; kit = WGA kit (R-g = REPLI-g, TP = TruePrime); Seq = Sequencing strategy (SE = single end, PE = paired end); Number of reads = total number of reads used for mapping; Mapped reads = fraction of reads mapped to the respective reference genome; Coverage breadth = fraction of reference genome with at least 1× coverage; Average coverage depth = mean number of reads per genomic position; Evenness score = evenness of coverage which quantifies uniformity of read distribution; SNP loci = number of single nucleotide variants compared to the respective reference genome.

[a]No variants called in these samples due to insufficient sequencing depth for triploid variants.

the genome covered was observed in the SE sequenced sedimentary eggs DP1–DP5 (Table 2). These eggs were pretreated with various bleach concentrations, suggesting a potential negative effect of bleaching for sedimentary eggs regardless of their age, compared to the several-days-old laboratory-produced eggs. Due to the experimental setup in which all bleached sedimentary samples were SE sequenced, and all nonbleached *Daphnia pulicara* eggs were PE sequenced, it is not possible to separate effects of pretreatment and sequencing effort on the mapping efficiency within this species. Nevertheless, contamination was low in the nonbleached sedimentary eggs DP6–DP9, with ~90% mapping rate, and ~70% coverage breadth (Table 2a), indicating efficient removal of exogenous DNA from egg surfaces with PBS and maintenance of DNA integrity before WGA.

High-throughput sequencing of libraries prepared from WGA-DNA from 13 of the 16 tested dormant *Daphnia* eggs successfully mapped with between 48% and 99% of reads (mean 88%, median 92%) in both SE and PE libraries to the respective *Daphnia* nuclear genomes, suggesting that WGA largely resulted in amplification of the target DNA (Table 2a). This was achieved for ephippia produced in laboratory cultures as well as those collected from lake sediment of different age with up to ~180-year-old eggs. Maximum coverage breadth (i.e., the fraction of the nuclear genome covered) was similar in both *Daphnia* species (between 70% and 80%, Table 2a). Moderate mapping efficiency was recorded for two relatively young eggs at 48% (DP1, from ~10-year-old sediment) and 56% (DM4, from a laboratory culture) and resulted in coverage of a lower fraction of the nuclear reference genomes (lower coverage breadth). Mapping to the nuclear genome failed almost entirely in the two oldest eggs where WGA was attempted (DP4 and DP5, ~300 years old, Table 2a), although the WGA-DNA yield was similar to that of other eggs (Table 1), suggesting amplification of contaminant DNA. However, egg age did not have a consistent effect on mapping efficiency; for example, SE libraries for the ~10-year-old *D. pulicaria* egg mapped with an efficiency of 48% (DP1), while the SE libraries from two ~140-year-old eggs mapped with an efficiency of 84%–87% (DP2, DP3). Likewise, PE libraries of *D. pulicaria* eggs were mapped with high efficiency (~90%) regardless of age (Table 2).

In contrast to the nuclear genome, we found that reads obtained from all historical eggs of *D. pulicaria* including the oldest samples (~300 years old) could be mapped to the mitochondrial genomes of the target species, resulting in high average coverage depth between 71× (DP4) and 60,000× (DP1) and a near to 100% coverage breadth of the mitochondrial genome (Table 2b). The extremely high coverage observed for DP1 indicates preferential amplification of mtDNA encoded in the mitochondrial and/or nuclear genome (see also next section).

3.3 | Read distribution

Patterns of genome-wide normalized read distribution differed between species, WGA kits and sequencing strategy (Figure 1). Results

for the evenness score that quantifies uniformity of read distribution did not suggest differences for libraries obtained with either WGA kit from the laboratory-produced eggs of *D. magna* (Figures 1a and S2). In contrast, the read distribution pattern differed markedly in *D. pulicaria* (Figure 1c–e). Samples from this species were represented by historical, sedimentary eggs and therefore the DNA may have been compromised to different degrees, providing inferior amplification substrate. The evenness scores obtained from sequencing sedimentary eggs (only *D. pulicaria*, Figure 1c,d) suggest lower uniformity of read distribution (Table 2a; Figure S2) in the REPLI-g amplified samples (DP1, DP4, DP5), while historical eggs amplified with TruePrime (DP2, DP3, DP6–DP9) provided a more uniform coverage with less pronounced outliers even in samples >100 years old (Figure 1c–e).

Read distribution patterns in *D. pulicaria* (Figure 1c–e) revealed several regions of the genome that were preferentially amplified. These outliers included genomic regions with sequences, for example, of the Pokey transposon or of several introns, but also for segments of mtDNA encoded in the nuclear genome (Table S1). The most extreme outlier identified in TruePrime-amplified DNA (a segment of nuclear mtDNA on scaffold 38, Table S1) was observed in both SE and PE libraries of the historical eggs.

To test the repeatability of read distribution patterns between samples within each species, we computed pairwise correlation coefficients (excluding the outlier on scaffold 38) for read counts within 10-kb bins (Figures 2, S3 and S4). For both species, samples amplified by the same WGA kit were strongly correlated (mean *r*: *D. magna* REPLI-g = .731; *D. magna* TruePrime = .767; *D. pulicaria* REPLI-g [SE] = .643; *D. pulicaria* TruePrime [SE] = .470, *D. pulicaria* TruePrime [PE] = .938). Weak correlations were observed between samples amplified using different kits (mean *r*: *D. magna* = .149; *D. pulicaria* = .085).

Owing to over-represented regions and preferentially amplified regions detailed above, the read distributions of none of the amplified samples had a significant fit to a Poisson distribution expected under the standard model for random sequencing (Table S2).

3.4 | Exogenous DNA

In samples where WGA failed almost entirely to produce target DNA (DP4 and DP5, Table 2a), but yielded a similar amount of DNA as for other eggs, we used a BLAST search to identify the taxon origin of the unmapped reads. This DNA was identified mostly as contaminant DNA from various bacterial and invertebrate taxa, as well as human DNA, but diversity and percentages of the contaminant taxa differed between the two eggs (Figure 3).

3.5 | Variant calling

For *D. magna* we called single nucleotide variants (SNPs) in the six samples with >6 million reads (DM1–DM3, DM5–DM7). For the
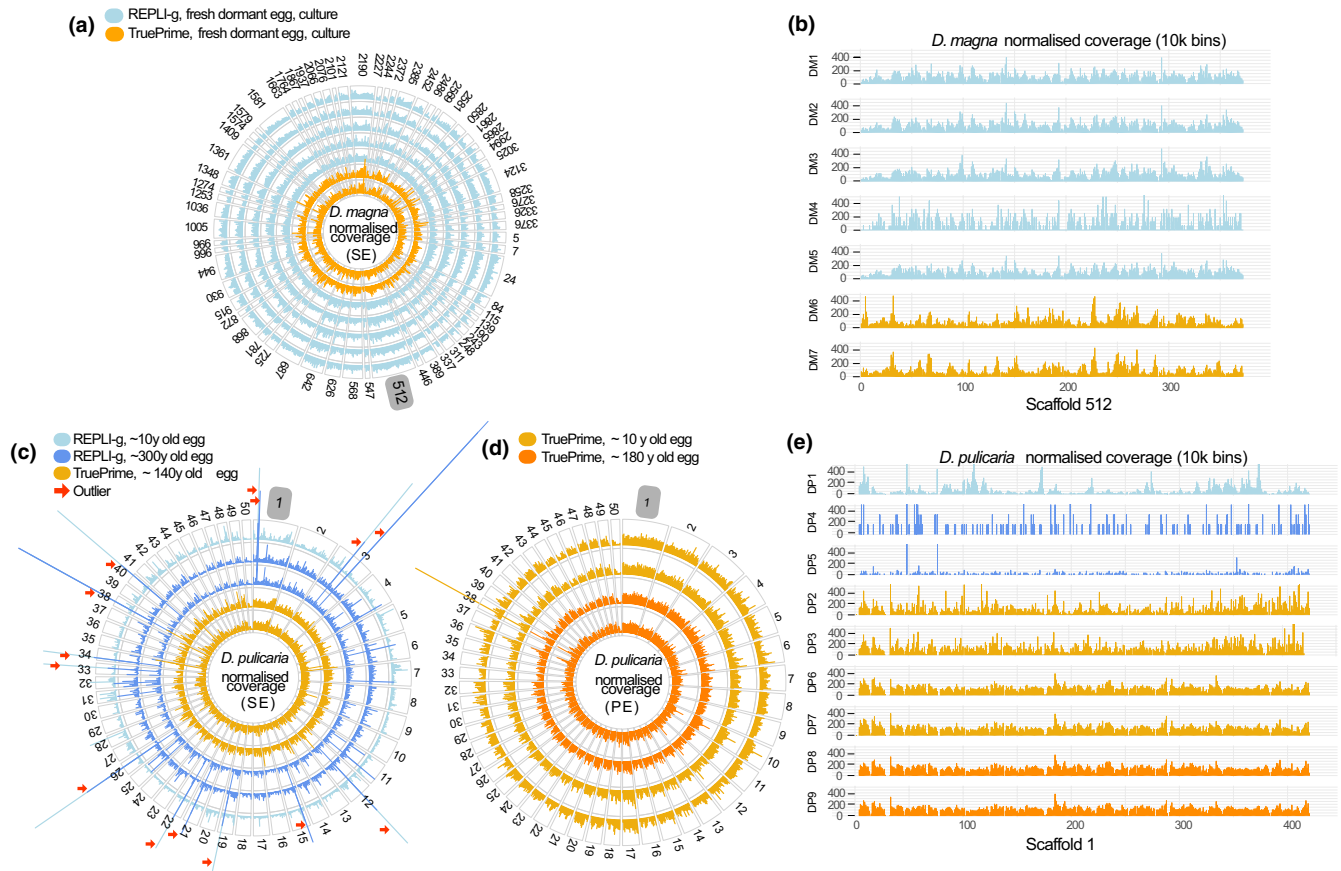
FIGURE 1 Visualization of the genome-wide normalized coverage pattern resulting from WGA-DNA of dormant eggs of *Daphnia magna* and *D. pulicaria*, with comparison of two commercial MDA kits. Bars in circular graphs (N50 scaffolds) represent normalized coverage in 100-kb bins, and 10-kb bins in linear graphs. (a) *D. magna* dormant eggs from cultures (SE libraries), amplified with REPLI-g (DM1–DM5, light blue) and TruePrime (DM6, DM7, orange). (b) Detail of (a) for the largest *D. magna* scaffold (scaffold 512). (c) *D. pulicaria* sedimentary eggs of different age (SE libraries). Amplified with REPLI-g (three outer rings: DP1, DP4, DP5) and with TruePrime (two inner rings: DP2, DP3). Red arrows point to outliers with preferred amplification in DP1, DP4 and DP5 (details in Table S1). (d) *D. pulicaria* sedimentary eggs of different age amplified with TruePrime (PE libraries). From outer to inner rings: DP6–DP9. (e) Detail of (c) for the largest *D. pulex* scaffold (scaffold 1). Sample order of circular graphs identical to linear graphs

triploid *D. pulicaria* eggs we could only use the four PE samples for confident SNP calling because deeper sequencing with higher coverage depth is needed for higher ploidy genomes.

The number of identified SNPs per sample in *D. magna* was between 99,927 and 430,909 in comparison with the *D. magna* reference genome (Table 2). In the *D. magna* samples with the highest number of called SNPs (REPLI-g: DM2, DM3, TruePrime: DM7, Table 2, Figure 4a), a total of 442,976 unique and shared SNP loci were recorded. All three samples could be genotyped at the required depth at 327,802 of these positions (74% of the potential genomic positions of SNPs in these three samples). SNP loci recorded in the lower-depth *D. magna* sample DM6 were nested almost entirely in those of the higher-depth sample DM7 (both amplified with TruePrime, Figure 4b). In *D. pulicaria*, the total number of SNP loci identified in the four PE samples in comparison to the *D. pulex* reference genome was 1,758,439. All four samples could be genotyped with the required depth at 1,712,889 (or 97.4%) of these genomic positions (Figure 4c).

We estimated the Ti:Tv ratio (Figure 4d) to gauge the variability of this metric between the samples studied and to assess whether

their values were within the ranges reported for *Daphnia* in general. We found that the ratios differed between species but that variation among samples of the same species was small (*D. pulicaria*: 1.30 in all four samples, *D. magna*: 1.43–1.47). Differences of ti:tv ratios between amplification kits (only *D. magna*) were small (REPLI-g: 1.43–1.45, TruePrime: 1.46–1.47).

SNP calling in the mitochondrial genomes (Figure 5) was performed with all available *D. pulicaria* samples. The analysis revealed the presence of 200 biallelic SNPs that differed between the two lake populations sampled. However, within-population samples were identical with the exception of DP1 (Lake SS4) and DP9 (Lake 1381) that differed from other samples of their respective population by a single SNP.

4 | DISCUSSION

WGA and subsequent WGS are staples of modern single-cell genome studies, and are widely applied to the study of human diseases
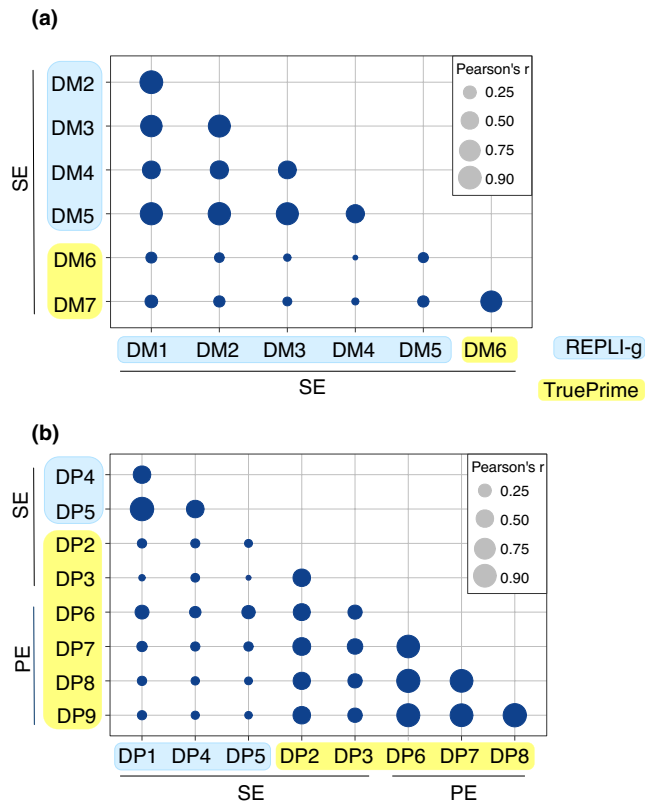
FIGURE 2 Correlation matrix (Pearson's *r*) of pairwise comparisons between normalized read counts of samples in 10-kb bins. (a) Pairwise comparisons of *Daphnia magna* samples. (b) Pairwise comparisons of *Daphnia pulicaria* samples. All pairwise comparisons had an associated *p*-value < .01

(Huang et al., 2015). Other promising but less common applications include phylogenomics (Ahrendt et al., 2018; Zhang et al., 2019) and metagenomics of microbial communities (Xu & Zhao, 2018). Suitable application for population genomics and evolutionary studies using *Daphnia* dormant eggs has been suggested (Lack et al., 2018), but to date a comprehensive study involving the comparison between multiple sedimentary eggs of different historical age and species, applying different pretreatments and amplification kits is not available. MDA has superior qualities when the goal is the discovery of single nucleotide variants, due to high fidelity of the φ29 polymerase and associated low error rates, while PCR-based WGA such as MALBAC may perform better for detecting copy number variation (de Bourcy et al., 2014; Chen et al., 2014). We therefore tested two MDA-WGA kits to provide a detailed workflow (Figure 6) for successful and repeatable amplification and sequencing of DNA for variant calls from dormant eggs of *Daphnia*.

4.1 | Pretreatment to minimize exogenous DNA

Sources of DNA contamination may originate from the WGA reagents, or are introduced when handling samples (Rinke et al., 2014). Contamination may also derive from nontarget exogenous DNA on the biological isolate (often of microbial origin), which is especially

common in historical samples containing highly degraded ancient DNA (Pilli et al., 2013). Decontamination procedures of equipment, which include the application of bleach and UV light, can be performed prior to WGA to prevent the amplification of exogenous DNA (Woyke et al., 2011). It is also recommended to use a thoroughly decontaminated laminar flowhood during all stages of WGA, preferably situated in a dedicated clean room.

Overall, our data suggest that amplification of exogenous DNA can be kept to a minimum when all careful steps are followed to avoid contamination. However, in historical samples where DNA is already damaged, exogenous DNA may be present in higher amounts than the target DNA, and thus be preferentially amplified, and even overwhelm the amplification process. This was particularly obvious in the oldest samples tested here (~300 years old). Apart from egg age, DNA integrity may also be highly dependent on the preservation conditions of the lake sediment: DNA preservation can vary strongly between different lakes and may be related to a number of variables, including temperature, salt concentration and pH (Ellegaard et al., 2020; Giguet-Covex et al., 2019).

Eggs produced in the laboratory that were only several days old did not show any signs of DNA degradation even after bleaching with a higher bleach concentration, and/or for a longer exposure time to bleach. In comparison to these samples, the DNA integrity and the resulting quality of WGA-DNA sedimentary eggs pretreated with diluted bleach appeared to be impaired. A possible explanation is the probable presence of microfissures in the egg membranes of historical, sedimentary eggs that are prone to increase with dormant egg age. To test this idea, microscopic studies comparing eggs of different age are needed.

Washing with PBS did not appear to affect the DNA in dormant eggs of *D. pulicaria* of different historical age (~10 and ~180 years old), which produced high-quality WGA-DNA. However, the observed higher quality of DNA libraries prepared from the PBS-treated eggs may also be largely due to a higher sequencing effort for these samples, as the PBS treatment was only performed prior to PE sequencing. It is also possible that these eggs had a generally higher quality than those used for bleaching, which were sampled from sediment of a neighbouring lake. Based on our findings, we cannot conclude with certainty that a pretreatment with PBS is superior to bleach, and further tests should be performed. However, serial washes with PBS appeared to effectively remove contamination in the tested samples, and thus may be a more cautious alternative to bleaching to remove possible contaminants on sedimentary egg surfaces that could overwhelm DNA amplification, particularly if the external structure of eggs and their DNA is more strongly impaired with increasing age.

4.2 | Mapping success and patterns of read distribution

Mapping of WGA-DNA produced with either of the two kits tested here was highly successful in most isolates, with a median of 92% of
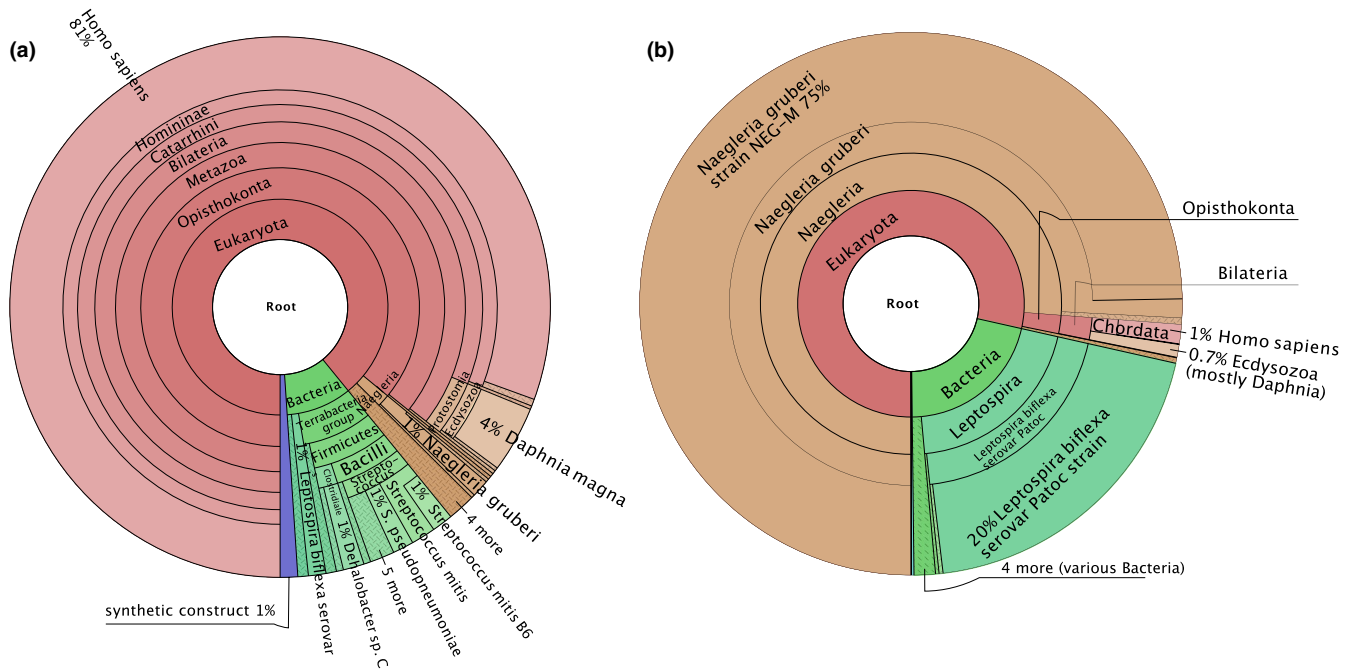
FIGURE 3 Taxon identity of nontarget DNA amplified from two *Daphnia pulicaria* dormant eggs. (a) DP5, (b) DP4. Labels only shown for those taxa that constitute >1% of the total exogenous sequences identified

reads mapped to their respective reference genomes, even of eggs as old as 180 years. Coverage breadth in most samples was between 70% and 80%. These values are similar to that previously reported for a dormant *Daphnia* egg, but below those obtained from *Daphnia* bulk sequencing (Lack et al., 2018), and higher than of MALBAC amplified individual sperm cells of *Daphnia* (53%, Xu et al., 2015). Indeed, incomplete genome coverage is commonly observed in WGA data in various taxa (e.g., de Bourcy et al., 2014; Huang et al., 2015; Picher et al., 2016).

Deviation of read distributions from a Poisson distribution expected under a random sequencing model (Lander & Waterman, 1988) was observed in all samples. However, this is not unexpected as it has been described previously that this model is inadequate for single-cell sequencing due to the possibility of locus dropout (Daley & Smith, 2014). Despite this, we found the regions of the genome that are amplified to be remarkably repeatable among samples of the same species, with highly significant correlation coefficients within amplification kits (average *r* of ~.7). The below-average *r* values were generally associated with failed target amplification or low read depth overall, specifically in the SE samples; but above-average values were found in the PE samples with high read depth. However, correlations between WGA kits were not strong, so for good comparability between samples, it is recommended to apply only one kit.

4.3 | Variant calling in nuclear and mitochondrial genomes

The highly reproducible patterns of read distribution within kits are largely responsible for the success of SNP calling across

samples, demonstrating the suitability of this method for population genomic applications, particularly for dormant eggs, and thus for its utility for studying genome evolution using sediment archives. In *D. magna*, >70% of the SNP positions could be called in all three samples with >14 million reads. Perhaps not surprisingly these values were even higher in the PE samples of *D. pulicaria* with >80 million reads, suggesting in general that sequencing strategy and depth strongly influence fidelity of the variant call also when applied to WGA-DNA.

Transition to transversion ratios have been suggested as a quality indicator for human SNP discovery (Wang et al., 2015). However, because these ratios vary between species, for example averaging 1.54 in several strains of *D. magna* (Ho et al., 2020), or 0.45 in *C. elegans* (Denver et al., 2009), comparisons should be made within species. Our range between 1.43 to 1.47 for nuclear DNA of *D. magna* was well within the ratio measured by Ho et al. (2020), and for *D. pulicaria* (Ti:Tv 1.30) was similar to that reported for the closely related *D. pulex* (1.58, Keith et al., 2016). In this study, we can also apply the Ti:Tv ratio as an indicator for the reliability of the WGA procedure within and across kits, with highly similar values for *D. magna* eggs, or almost identical values for *D. pulicaria*.

Other studies have identified possible limitations of the reliability of WGA, such as coverage uniformity, reproducibility and allelic dropout rate (de Bourcy et al., 2014; Huang et al., 2015). The substrates that these studies used were cell lineages from which individual cells were subjected to single cell WGA (scWGA) and compared to bulk sequencing of a multicellular sample from the same lineage. For historical isolates of dormant eggs from the sediment, such a strategy cannot be applied. However, an effective substitute to test the reliability of SNPs from these historical samples was the use of
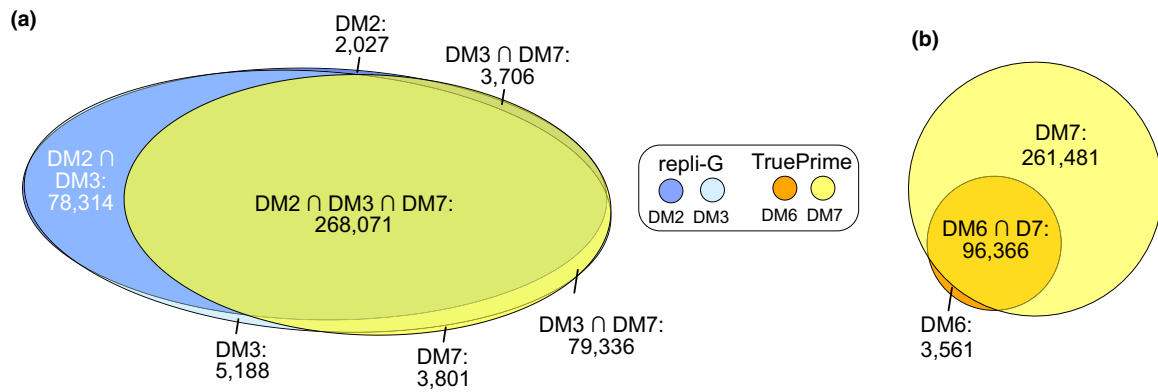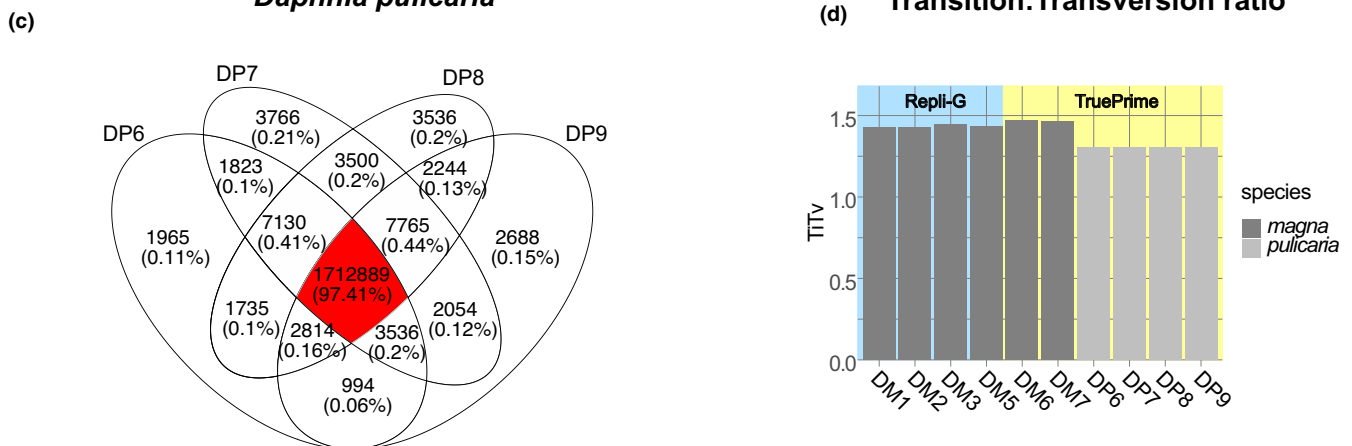
FIGURE 4 Results of SNP analysis in *Daphnia magna* and *D. pulicaria*. Venn diagrams (a–c) represent the number of genomic positions where SNPs were located in one or more samples and that were sequenced in all or a subset of samples (shared and unique genomic positions). This comparison was used to estimate the repeatability of whole genome amplification and thus for the resulting capacity to call variants at multisample level. (a) Comparison between three *D. magna* samples with the highest number of SNPs identified, amplified by REPLI-g (DM2, DM3) and TruePrime (DM7). (b) Comparison between the two *D. magna* samples amplified by TruePrime (DM6, DM7). (c) Comparison between four *D. pulicaria* samples, amplified with TruePrime. (d) Transition:transversion ratio analysed for SNPs of both *Daphnia* species

asexually produced dormant eggs such as those of the Arctic *D. pulicaria* population from which our samples originated. WGA of these samples allowed 97% of all SNP loci detected in the PE sequences to be genotyped in all triploid isolates. Preliminary analysis of variation between these genotyped eggs showed a maximum difference of 3% of the roughly 1.7 million SNPs between individuals (unpublished data). Encouraging results were also reported for MDA-amplified DNA from individual adults of a *D. pulicaria* clone compared to DNA from pooled individuals of the same clone (Lack et al., 2018). These authors found only a slight loss of heterozygosity in the amplified DNA, but concordance of structural variation (insertions, deletions, duplications and translocation, but not of inversions) between both sample types (Lack et al., 2018).

An added benefit of WGA with both MDA kits tested here was the possibility of obtaining high coverage of the full mitochondrial genome for both species. This is of particular interest for the historical samples from which full mitochondrial genomes and

high-quality SNP calls could be retrieved even of the oldest samples (~300-year-old eggs). This opens a promising avenue for gathering information of genome-wide mutation rates and spectra of the mitochondrial genome stored in sedimentary archives across extended time periods and thousands of generations, probably surpassing the time range tested here.

5 | CONCLUSION

Although the method described here was tested on dormant eggs of *Daphnia*, it could be widely applied to sedimentary dormant stages of a variety of taxa, potentially providing access to the evolutionary history not only of single taxa but also to that of entire aquatic communities. Such taxa could include key members of the plankton in the freshwater and marine to hypersaline ecosystems that produce dormant propagules at an early embryonic stage with a limited number
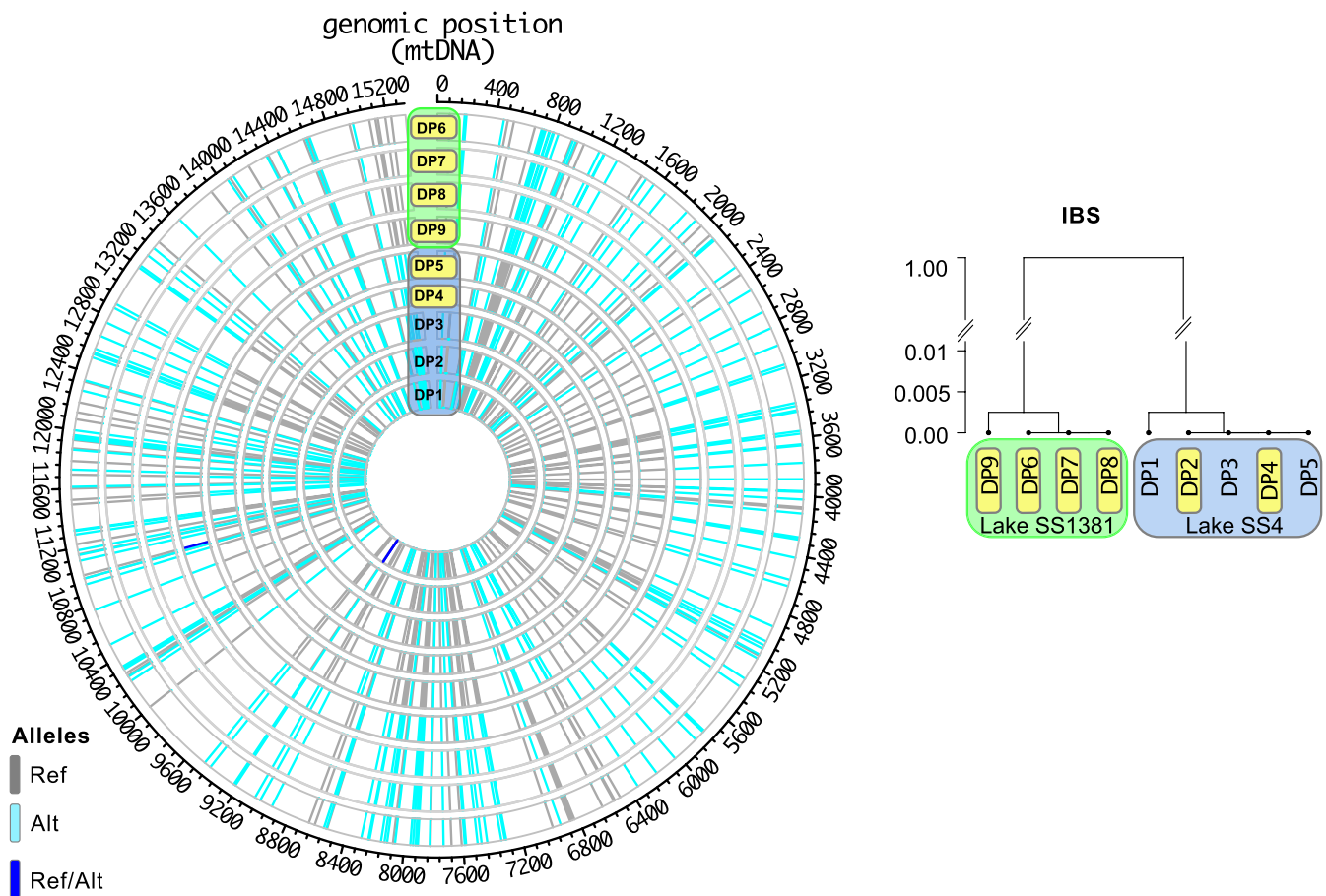
WILEY |

FIGURE 5 SNP positions and pairwise clustering of individuals by IBS of nine *Daphnia pulicaria* mitochondrial genomes recovered from historical sedimentary dormant eggs from two lakes in West Greenland (Lake SS1381, green label, and SS4, blue label). Samples with yellow labels were amplified with TruePrime, and the remaining three with REPLI-g. (a) Circular plot of detected SNP loci. The four outer rings represent samples from Lake SS1381, the five inner rings from SS4. SNP colours are labelled according to their state (turquoise = homozygous [variant allele], grey = homozygous [reference allele], blue = heteroplasmic [both reference and variant allele present]). (b) Hierarchical clustering tree resulting from the fraction of identical genomic positions (identity-by-state) in the nine mitochondrial genomes compared

of cells, such as the dormant eggs of calanoid copepods, cysts of the brine shrimp *Artemia*, algal dormant stages and plant seeds. Other methods that have successfully been applied to individual small planktonic crustaceans, such as low-DNA-input sequencing libraries (Beninde et al., 2020), could be tested on dormant stages isolated from historical sediment layers. However, their library preparation protocol, which was optimized for adult specimens, requires an amount of purified template DNA that may exceed the DNA content available from dormant embryos. In such cases, a technology which does not require a DNA extraction step such as the MDA methodology tested here could be a more practical approach. Since the protocol presented here does not require DNA extraction but uses the entire egg material as template, the risk of losing precious material can be minimized.

Our data reveal that both of the tested amplification kits provided high-quality DNA for most *Daphnia* egg isolates, and that the

amplified DNA could efficiently be applied to WGS and subsequent genome-wide studies at the population level. However, differences were observed with respect to the age of dormant eggs, where our results suggest a superior performance of TruePrime (compared with REPLI-g) for application to eggs of sedimentary origin and thus to prospectively degraded DNA. A possible explanation could be that the primase TthPrimPol used in the TruePrime kit shows translesion activity, allowing re-initiation of the replication fork when encountering damaged DNA, and thus continued amplification of damaged substrate (Picher & Blanco, 2014). Due to the similarity of TthPrimPol to human PrimPol, another mechanism could explain our results: human PrimPol can reprime DNA synthesis following a lesion, allowing the φ29 DNA polymerase to continue the amplification process close to the region in which the lesion was found (Mourón et al., 2013), possibly improving evenness during the process.
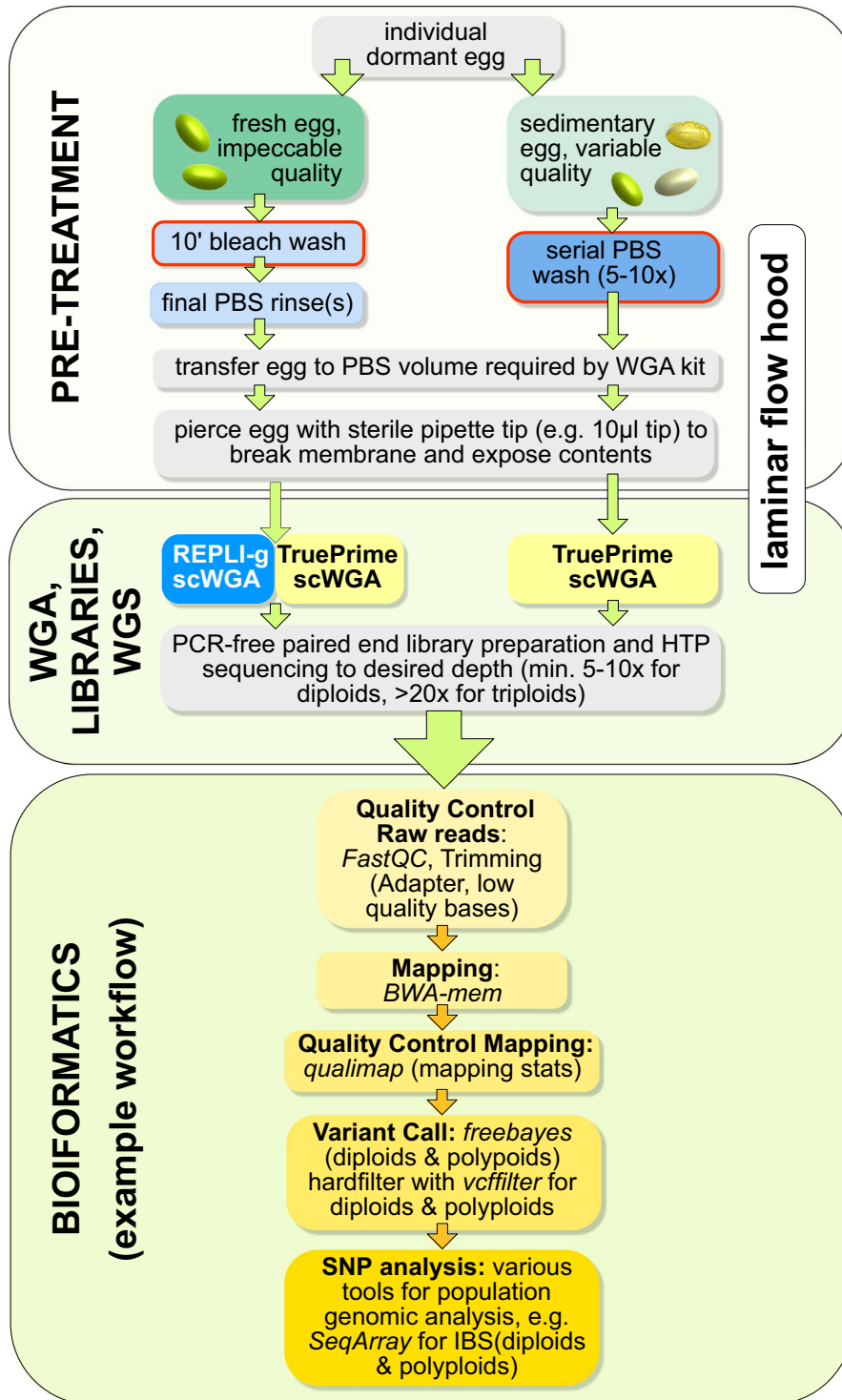
FIGURE 6 Recommended workflow for whole genome amplification and whole genome sequencing with downstream variant call and variant filtering. Recommendation for sequencing depth of higher ploidy genomes see Maruki and Lynch (2017). Steps that need further testing are marked by a red line. For further details see text

Ultimately, our data indicate that for optimal results, preliminary trials are recommended using both kits tested here (and if possible others) on the dormant egg population in question.

CONFLICT OF INTEREST

The authors declare no conflicts of interest.

AUTHOR CONTRIBUTIONS

D.F. conceived the idea and obtained funding for the study. D.F., C.O.G. and J.K.C. designed the experiments. C.O.G. performed the laboratory work. D.F. and C.O.G. analysed the data and wrote the paper with input from J.K.C. and V.D.

DATA AVAILABILITY STATEMENT

The sequence data supporting the findings reported here are available in the open access repository Zenodo at https://doi.org/10.5281/zenodo.5256276.

ORCID

Vignesh Dhandapani https://orcid.org/0000-0002-5745-2409
John K. Colbourne https://orcid.org/0000-0002-6966-2972
Dagmar Frisch https://orcid.org/0000-0001-9310-2230

REFERENCES

Ahrendt, S. R., Quandt, C. A., Ciobanu, D., Clum, A., Salamov, A., Andreopoulos, B., Cheng, J.-F., Woyke, T., Pelin, A., Henrissat, B., Reynolds, N. K., Benny, G. L., Smith, M. E., James, T. Y., & Grigoriev, I. V. (2018). Leveraging single-cell genomics to expand the fungal tree of life. *Nature Microbiology*, *3*(12), 1417–1428. https://doi.org/10.1038/s41564-018-0261-0

Andrews, S. (2015). *FASTQC a quality control tool for high throughput sequence data*. Babraham Institute. https://www.Bioinformatics.Babraham.Ac.Uk/Projects/Fastqc/

Barta, J. L., Monroe, C., Teisberg, J. E., Winters, M., Flanigan, K., & Kemp, B. M. (2014). One of the key characteristics of ancient DNA, low copy number, may be a product of its extraction. *Journal of Archaeological Science*, *46*, 281–289. https://doi.org/10.1016/j.jas.2014.03.030

Beninde, J., Möst, M., & Meyer, A. (2020). Optimized and affordable high-throughput sequencing workflow for preserved and nonpreserved small zooplankton specimens. *Molecular Ecology Resources*, *20*(6), 1632–1646. https://doi.org/10.1111/1755-0998.13228

Blainey, P. C., & Quake, S. R. (2011). Digital MDA for enumeration of total nucleic acid contamination. *Nucleic Acids Research*, *39*(4), e19. https://doi.org/10.1093/nar/gkq1074

Blanco, L., Bernad, A., Lázaro, J. M., Martín, G., Garmendia, C., & Salas, M. (1989). Highly efficient DNA synthesis by the phage ϕ 29 DNA polymerase. *Journal of Biological Chemistry*, *264*(15), 8935–8940. https://doi.org/10.1016/s0021-9258(18)81883-x

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Boschetti, C., Leasi, F., & Ricci, C. (2011). Developmental stages in diapausing eggs: An investigation across monogonont rotifer species. *Hydrobiologia*, *662*, 149–155. https://doi.org/10.1007/s10750-010-0490-6

Brede, N., Sandrock, C., Straile, D., Spaak, P., Jankowski, T., Streit, B., & Schwenk, K. (2009). The impact of human-made ecological changes on the genetic architecture of *Daphnia* species. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(12), 4758–4763. https://doi.org/10.1073/pnas.0807187106

Broad Institute. (2019). *Picard toolkit*. Broad Institute, GitHub Repository. https://Github.Com/Broadinstitute/Picard/Releases/Tag/2.25.0

Chen, L., Barnett, R. E., Horstmann, M., Bamberger, V., Heberle, L., Krebs, N., Colbourne, J. K., Gómez, R., & Weiss, L. C. (2018). Mitotic activity patterns and cytoskeletal changes throughout the progression of diapause developmental program in *Daphnia*. *BMC Cell Biology*, *19*(1), 30. https://doi.org/10.1186/s12860-018-0181-0

Chen, M., Song, P., Zou, D., Hu, X., Zhao, S., Gao, S., & Ling, F. (2014). Comparison of multiple displacement amplification (MDA) and multiple annealing and looping-based amplification cycles (MALBAC) in single-cell sequencing. *PLoS One*, *9*(12), e114520. https://doi.org/10.1371/journal.pone.0114520

Cheung, V. G., & Nelson, S. F. (1996). Whole genome amplification using a degenerate oligonucleotide primer allows hundreds of genotypes to be performed on less than one nanogram of genomic DNA. *Proceedings of the National Academy of Sciences*, *93*(25), 14676–14679. https://doi.org/10.1073/pnas.93.25.14676

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), 555–561. https://doi.org/10.1126/science.1197761

Cordellier, M., Wojewodzic, M. W., Wessels, M., Kuster, C., & von Elert, E. (2021). Next-generation sequencing of DNA from resting eggs: Signatures of eutrophication in a lake's sediment. *Zoology*, *145*, 125895. https://doi.org/10.1016/j.zool.2021.125895

Crease, T. J. (1999). The complete sequence of the mitochondrial genome of *Daphnia pulex* (Cladocera: Crustacea). *Gene*, *233*(1–2), 89–99. https://doi.org/10.1016/S0378-1119(99)00151-1

Daley, T., & Smith, A. D. (2014). Modeling genome coverage in single-cell sequencing. *Bioinformatics*, *30*(22), 3159–3165. https://doi.org/10.1093/bioinformatics/btu540

Dane, M., Anderson, N. J., Osburn, C. L., Colbourne, J. K., & Frisch, D. (2020). Centennial clonal stability of asexual *Daphnia* in Greenland lakes despite climate variability. *Ecology and Evolution*, *10*(24), 14178–14188. https://doi.org/10.1002/ece3.7012

de Bourcy, C. F. A., De Vlaminck, I., Kanbar, J. N., Wang, J., Gawad, C., & Quake, S. R. (2014). A quantitative comparison of single-cell whole genome amplification methods. *PLoS One*, *9*(8), e105585. https://doi.org/10.1371/journal.pone.0105585

De La Torre, A. R., Wilhite, B., & Neale, D. B. (2019). Environmental genome-wide association reveals climate adaptation is shaped by subtle to moderate allele frequency shifts in loblolly pine. *Genome Biology and Evolution*, *11*(10), 2976–2989. https://doi.org/10.1093/gbe/evz220

Dean, F. B., Hosono, S., Fang, L., Wu, X., Faruqi, A. F., Bray-Ward, P., Sun, Z., Zong, Q., Du, Y., Du, J., Driscoll, M., Song, W., Kingsmore, S. F., Egholm, M., & Lasken, R. S. (2002). Comprehensive human genome amplification using multiple displacement amplification. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(8), 5261–5266. https://doi.org/10.1073/pnas.082089499

Dean, F. B., Nelson, J. R., Giesler, T. L., & Lasken, R. S. (2001). Rapid amplification of plasmid and phage DNA using Phi29 DNA polymerase and multiply-primed rolling circle amplification. *Genome Research*, *11*(6), 1095–1099. https://doi.org/10.1101/gr.180501

Denver, D. R., Dolan, P. C., Wilhelm, L. J., Sung, W., Lucas-Lledo, J. I., Howe, D. K., Lewis, S. C., Okamoto, K., Thomas, W. K., Lynch, M., & Baer, C. F. (2009). A genome-wide view of *Caenorhabditis elegans* base-substitution mutation processes. *Proceedings of the National Academy of Sciences*, *106*(38), 16310–16314. https://doi.org/10.1073/pnas.0904895106

Dettman, J. R., Rodrigue, N., Melnyk, A. H., Wong, A., Bailey, S. F., & Kassen, R. (2012). Evolutionary insight from whole-genome sequencing of experimentally evolved microbes. *Molecular Ecology*, *21*(9), 2058–2077. https://doi.org/10.1111/j.1365-294X.2012.05484.x

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

Ellegaard, M., Clokie, M. R. J., Czypionka, T., Frisch, D., Godhe, A., Kremp, A., Letarov, A., McGenity, T. J., Ribeiro, S., & John Anderson, N. (2020). Dead or alive: Sediment DNA archives as tools for tracking aquatic evolution and adaptation. *Communications Biology*, 3(1), 169. https://doi.org/10.1038/s42003-020-0899-z

Ellegren, H. (2014). Genome sequencing and population genomics in nonmodel organisms. *Trends in Ecology and Evolution*, 29(1), 51–63. https://doi.org/10.1016/j.tree.2013.09.008

Frisch, D., Morton, P. K., Culver, B. W., Edlund, M. B., Jeyasingh, P. D., & Weider, L. J. (2016). Paleogenetic records of *Daphnia pulicaria* in North American lakes reveal the impact of cultural eutrophication. *Global Change Biology*, 23(2), 708–718. https://doi.org/10.1111/gcb.13445

Frisch, D., Morton, P. K., Roy Chowdhury, P., Culver, B. W., Colbourne, J. K., Weider, L. J., & Jeyasingh, P. D. (2014). A millennial-scale chronicle of evolutionary responses to cultural eutrophication in *Daphnia*. *Ecology Letters*, 17(3), 360–368. https://doi.org/10.1111/ele.12237

Gao, C.-H., & Yi, L. (2019). ggVennDiagram: A "ggplot2" implement of Venn Diagram. https://github.com/gaospecial/ggVennDiagram

García-Alcalde, F., Okonechnikov, K., Carbonell, J., Cruz, L. M., Götz, S., Tarazona, S., Dopazo, J., Meyer, T. F., & Conesa, A. (2012). Qualimap: Evaluating next-generation sequencing alignment data. *Bioinformatics*, 28(20), 2678–2679. https://doi.org/10.1093/bioinformatics/bts503

Garrison, E. (2016). *Vcflib, a simple C++ library for parsing and manipulating VCF files*. Retrieved from https://github.com/vcflib/vcflib

Garrison, E. & Marth, G. (2012). Haplotype-based variant detection from short-read sequencing. ArXiv Preprint ArXiv:1207.3907.

Giguet-Covex, C., Ficetola, G. F., Walsh, K., Poulenard, J., Bajard, M., Fouinat, L., Sabatier, P., Gielly, L., Messager, E., Develle, A. L., David, F., Taberlet, P., Brisset, E., Guiter, F., Sinet, R., & Arnaud, F. (2019). New insights on lake sediment DNA from the catchment: Importance of taphonomic and analytical issues on the record quality. *Scientific Reports*, 9(1), 14676. https://doi.org/10.1038/s41598-019-50339-1

Gogarten, S. M., Zheng, X., & Stilp, A. (2021). SeqVarTools: Tools for variant data. *R package version 1.30.0*. https://github.com/smgogarten/SeqVarTools

Gu, Z., Gu, L., Eils, R., Schlesner, M., & Brors, B. (2014). Circlize implements and enhances circular visualization in R. *Bioinformatics*, 30(19), 2811–2812. https://doi.org/10.1093/bioinformatics/btu393

Handyside, A. H., Robinson, M. D., Simpson, R. J., Omar, M. B., Shaw, M. A., Grudzinskas, J. G., & Rutherford, A. (2004). Isothermal whole genome amplification from single and small numbers of cells: A new era for preimplantation genetic diagnosis of inherited disease. *Molecular Human Reproduction*, 10(10), 767–772. https://doi.org/10.1093/molehr/gah101

Härnström, K., Ellegaard, M., Andersen, T. J., & Godhe, A. (2011). Hundred years of genetic structure in a sediment revived diatom population. *Proceedings of the National Academy of Sciences of the United States of America*, 108(10), 4252–4257. https://doi.org/10.1073/pnas.1013528108

Ho, E. K. H., Macrae, F., Latta, L. C., McIlroy, P., Ebert, D., Fields, P. D., & Schaack, S. (2020). High and highly variable spontaneous mutation rates in *Daphnia*. *Molecular Biology and Evolution*, 37(11), 3258–3266. https://doi.org/10.1093/molbev/msaa142

Hohenlohe, P. A., Hand, B. K., Andrews, K. R., & Luikart, G. (2018). Population genomics provides key insights in ecology and evolution. In O. Rajora (Ed.), *Population genomics*. Springer. https://doi.org/10.1007/13836_2018_20

Huang, L., Ma, F., Chapman, A., Lu, S., & Xie, X. S. (2015). Single-cell whole-genome amplification and sequencing: Methodology and applications. *Annual Review of Genomics and Human Genetics*, 16(1), 79–102. https://doi.org/10.1146/annurev-genom-090413-025352

Keith, N., Tucker, A. E., Jackson, C. E., Sung, W., Lucas Lledó, J. I., 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

Lack, J. B., Weider, L. J., & Jeyasingh, P. D. (2018). Whole genome amplification and sequencing of a *Daphnia* resting egg. *Molecular Ecology Resources*, 18(1), 118–127. https://doi.org/10.1111/1755-0998.12720

Lander, E. S., & Waterman, M. S. (1988). Genomic mapping by fingerprinting random clones: A mathematical analysis. *Genomics*, 2(3), 231–239. https://doi.org/10.1016/0888-7543(88)90007-9

Larsson, J. (2020). eulerr: Area-proportional euler and venn diagrams with ellipses. R package version 6.1.0. Retrieved from https://cran.r-project.org/package=eulerr.

Lasken, R. S., & Egholm, M. (2003). Whole genome amplification: Abundant supplies of DNA from precious samples or clinical specimens. *Trends in Biotechnology*, 21(12), 531–535. https://doi.org/10.1016/j.tibtech.2003.09.010

Leonardi, M., Librado, P., Der Sarkissian, C., Schubert, M., Alfarhan, A. H., Alquraishi, S. A., Al-Rasheid, K. A. S., Gamba, C., Willerslev, E., & Orlando, L. (2017). Evolutionary patterns and processes: Lessons from ancient DNA. *Systematic Biology*, 66(1), e1–e29. https://doi.org/10.1093/sysbio/syw059

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

Limburg, P. A., & Weider, L. J. (2002). "Ancient" DNA in the resting egg bank of a microcrustacean can serve as a palaeolimnological database. *Proceedings of the Royal Society of London Series B-Biological Sciences*, 269(1488), 281–287.

Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., He, G., Chen, Y., Pan, Q. I., Liu, Y., Tang, J., Wu, G., Zhang, H., Shi, Y., Liu, Y., Yu, C., Wang, B. O., Lu, Y., Han, C., … 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

Maruki, T., & Lynch, M. (2017). Genotype calling from population-genomic sequencing data. *G3: Genes|genomes|genetics*, 7(5), 1393–1404. https://doi.org/10.1534/g3.117.039008

Mergeay, J., Verschuren, D., & De Meester, L. (2006). Invasion of an asexual American water flea clone throughout Africa and rapid displacement of a native sibling species. *Proceedings of the Royal Society B-Biological Sciences*, 273(1603), 2839–2844.

Mourón, S., Rodriguez-Acebes, S., Martínez-Jiménez, M. I., García-Gómez, S., Chocrón, S., Blanco, L., & Méndez, J. (2013). Repriming of DNA synthesis at stalled replication forks by human PrimPol. *Nature Structural & Molecular Biology*, 20(12), 1383–1389. https://doi.org/10.1038/nsmb.2719

O'Grady, C., Dhandapani, V., Colbourne, J.K., & Frisch, D. (2021). Data from: Refining the evolutionary time machine: an assessment of whole genome amplification using single historical *Daphnia* eggs. *Molecular Ecology Resources*. https://doi.org/10.5281/zenodo.5256276

Ondov, B. D., Bergman, N. H., & Phillippy, A. M. (2011). Interactive metagenomic visualization in a Web browser. *BMC Bioinformatics*, 12(1), 385. https://doi.org/10.1186/1471-2105-12-385

Orsini, L., Schwenk, K., De Meester, L., Colbourne, J. K., Pfrender, M. E., & Weider, L. J. (2013). The evolutionary time machine: Forecasting how populations can adapt to changing environments using dormant propagules. *Trends in Ecology & Evolution*, 28, 274–282. https://doi.org/10.1016/j.tree.2013.01.009

Orsini, L., Spanier, K. I., & De Meester, L. (2012). Genomic signature of natural and anthropogenic stress in wild populations of the waterflea *Daphnia magna*: Validation in space, time and experimental evolution. *Molecular Ecology*, 21(9), 2160–2175. https://doi.org/10.1111/j.1365-294X.2011.05429.x

Paez, J. G. (2004). Genome coverage and sequence fidelity of 29 polymerase-based multiple strand displacement whole genome amplification. *Nucleic Acids Research*, 32(9), e71. https://doi.org/10.1093/nar/gnh069

Parks, M., Subramanian, S., Baroni, C., Salvatore, M. C., Zhang, G., Millar, C. D., & Lambert, D. M. (2015). Ancient population genomics and the study of evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 370, 20130381. https://doi.org/10.1098/rstb.2013.0381

Picher, Á. J., & Blanco, L. (2014). *Patent No. International Publication Number WO2014140309A1*. https://patents.google.com/patent/WO2014140309A1/en

Picher, Á. J., Budeus, B., Wafzig, O., Krüger, C., García-Gómez, S., Martínez-Jiménez, M. I., Díaz-Talavera, A., Weber, D., Blanco, L., & Schneider, A. (2016). TruePrime is a novel method for whole-genome amplification from single cells based on TthPrimPol. *Nature Communications*, 7(1), 13296. https://doi.org/10.1038/ncomms13296

Pilli, E., Modi, A., Serpico, C., Achilli, A., Lancioni, H., Lippi, B., Bertoldi, F., Gelichi, S., Lari, M., & Caramelli, D. (2013). Monitoring DNA contamination in handled vs. directly excavated ancient human skeletal remains. *PLoS One*, 8(1), e52524. https://doi.org/10.1371/journal.pone.0052524

Pollard, H. G., Colbourne, J. K., & Keller, W. (2003). Reconstruction of centuries-old *Daphnia* communities in a lake recovering from acidification and metal contamination. *Ambio*, 32(3), 214–218. https://doi.org/10.1579/0044-7447-32.3.214

Quinlan, A. R., & Hall, I. M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26(6), 841–842. https://doi.org/10.1093/bioinformatics/btq033

R CoreTeam. (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. http://www.R-project.org

Rajpurohit, S., Gefen, E., Bergland, A. O., Petrov, D. A., Gibbs, A. G., & Schmidt, P. S. (2018). Spatiotemporal dynamics and genome-wide association genome-wide association analysis of desiccation tolerance in *Drosophila melanogaster*. *Molecular Ecology*, 27(17), 3525–3540. https://doi.org/10.1111/mec.14814

Reed, K. A., Lee, S. G., Lee, J. H., Park, H., & Covi, J. A. (2021). The ultrastructure of resurrection: Post-diapause development in an Antarctic freshwater copepod. *Journal of Structural Biology*, 213(1), 107705. https://doi.org/10.1016/j.jsb.2021.107705

Rinke, C., Lee, J., Nath, N., Goudeau, D., Thompson, B., Poulton, N., Dmitrieff, E., Malmstrom, R., Stepanauskas, R., & Woyke, T. (2014). Obtaining genomes from uncultivated environmental microorganisms using FACS-based single-cell genomics. *Nature Protocols*, 9(5), 1038–1048. https://doi.org/10.1038/nprot.2014.067

Rizzi, E., Lari, M., Gigli, E., De Bellis, G., & Caramelli, D. (2012). Ancient DNA studies: New perspectives on old samples. *Genetics Selection Evolution*, 44(1), 21. https://doi.org/10.1186/1297-9686-44-21

Schlötterer, C., Tobler, R., Kofler, R., & Nolte, V. (2014). Sequencing pools of individuals—Mining genome-wide polymorphism data without big funding. *Nature Reviews Genetics*, 15(11), 749–763. https://doi.org/10.1038/nrg3803

Sella, G., & Barton, N. H. (2019). Thinking about the evolution of complex traits in the era of genome-wide association studies. *Annual Review of Genomics and Human Genetics*, 20, 461–493. https://doi.org/10.1146/annurev-genom-083115-022316

Stiller, J., & Zhang, G. (2019). Comparative phylogenomics, a stepping stone for bird biodiversity studies. *Diversity*, 11(7), 115. https://doi.org/10.3390/d11070115

Telenius, H., Carter, N. P., Bebb, C. E., Nordenskjöld, M., Ponder, B. A. J., & Tunnacliffe, A. (1992). Degenerate oligonucleotide-primed PCR: General amplification of target DNA by a single degenerate primer. *Genomics*, 13(3), 718–725. https://doi.org/10.1016/0888-7543(92)90147-K

von Baldass, F. (1941). Entwicklung von *Daphnia pulex*. *Zoologische Jahrbücher. Abteilung Für Anatomie Und Ontogenie Der Tiere*, 67, 1–60.

Wang, J., Raskin, L., Samuels, D. C., Shyr, Y., & Guo, Y. (2015). Genome measures used for quality control are dependent on gene function and ancestry. *Bioinformatics*, 31(3), 318–323. https://doi.org/10.1093/bioinformatics/btu668

Weider, L. J., Lampert, W., Wessels, M., Colbourne, J. K., & Limburg, P. (1997). Long-term genetic shifts in a microcrustacean egg bank associated with anthropogenic changes in the Lake Constance ecosystem. *Proceedings of the Royal Society of London Series B-Biological Sciences*, 264(1388), 1613–1618.

Wells, D., Sherlock, J. K., Handyside, A. H., & Delhanty, J. D. (1999). Detailed chromosomal and molecular genetic analysis of single cells by whole genome amplification and comparative genomic hybridisation. *Nucleic Acids Research*, 27(4), 1214–1218. https://doi.org/10.1093/nar/27.4.1214

Wickham, H. (2016). *ggplot2 Elegant Graphics for Data Analysis (Use R!)*. Springer. https://doi.org/10.1007/978-0-387-98141-3

Woyke, T., Sczyrba, A., Lee, J., Rinke, C., Tighe, D., Clingenpeel, S., Malmstrom, R., Stepanauskas, R., & Cheng, J.-F. (2011). Decontamination of MDA reagents for single cell whole genome amplification. *PLoS One*, 6(10), e26161. https://doi.org/10.1371/journal.pone.0026161

Xu, S., Ackerman, M. S., Long, H., Bright, L., Spitze, K., Ramsdell, J. S., & Lynch, M. (2015). A male-specific genetic map of the microcrustacean *Daphnia pulex* based on single-sperm whole-genome sequencing. *Genetics*, 201(1), 31–38. https://doi.org/10.1534/genetics.115.179028

Xu, Y., & Zhao, F. (2018). Single-cell metagenomics: Challenges and applications. *Protein & Cell*, 9(5), 501–510. https://doi.org/10.1007/s13238-018-0544-5

Zhang, F., Ding, Y., Zhu, C. D., Zhou, X., Orr, M. C., Scheu, S., & Luan, Y. X. (2019). Phylogenomics from low-coverage whole-genome sequencing. *Methods in Ecology and Evolution*, 10, 507–517. https://doi.org/10.1111/2041-210X.13145

Zhang, L., Cui, X., Schmitt, K., Hubert, R., Navidi, W., & Arnheim, N. (1992). Whole genome amplification from a single cell: Implications for genetic analysis. *Proceedings of the National Academy of Sciences of the United States of America*, 89(13), 5847–5851. https://doi.org/10.1073/pnas.89.13.5847

Zheng, X., Gogarten, S. M., Lawrence, M., Stilp, A., Conomos, M. P., Weir, B. S., Laurie, C., & Levine, D. (2017). SeqArray-a storage-efficient high-performance data format for WGS variant calls. *Bioinformatics*, 33(15), 2251–2257. https://doi.org/10.1093/bioinformatics/btx145

Zheng, X., Levine, D., Shen, J., Gogarten, S. M., Laurie, C., & Weir, B. S. (2012). A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics*, 28(24), 3326–3328. https://doi.org/10.1093/bioinformatics/bts606

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.