Sequencing complete mitochondrial and plastid genomes

Gertraud Burger[1], Dennis V Lavrov[1,2], Lise Forget[1] & B Franz Lang[1]

[1]Département de Biochimie, Robert Cedergren Centre, Program in Evolutionary Biology, Canadian Institute for Advanced Research, Université de Montréal, 2900 Boulevard Edouard-Montpetit, CP 6128, Montréal, Québec, H3T 1J4 Canada. [2]Present address: Department of Ecology, Evolution and Organismal Biology, Iowa State University, Ames, Iowa 50011, USA. Correspondence should be addressed to G.B. (Gertraud.Burger@UMontreal.CA) or B.F.L. (Franz.lang@Umontreal.ca).

Published online 22 March 2007; doi:10.1038/nprot.2007.59

Organelle genomics has become an increasingly important research field, with applications in molecular modeling, phylogeny, taxonomy, population genetics and biodiversity. Typically, research projects involve the determination and comparative analysis of complete mitochondrial and plastid genome sequences, either from closely related species or from a taxonomically broad range of organisms. Here, we describe two alternative organelle genome sequencing protocols. The "random genome sequencing" protocol is suited for the large majority of organelle genomes irrespective of their size. It involves DNA fragmentation by shearing (nebulization) and blunt-end cloning of the resulting fragments into pUC or BlueScript-type vectors. This protocol excels in randomness of clone libraries as well as in time and cost-effectiveness. The "long-PCR-based genome sequencing" protocol is specifically adapted for DNAs of low purity and quantity, and is particularly effective for small organelle genomes. Library construction by either protocol can be completed within 1 week.

INTRODUCTION

Organelle genomics has unearthed an unsuspected wealth of genetic diversity of mitochondrial DNAs (mtDNAs) and plastid (chloroplast) DNAs across eukaryotes. In addition, these genomes display intriguing molecular features such as genes and genomes in pieces[1–3], RNA editing[4,5], intron-mediated *trans*-splicing, catalytic introns and intron homing[6,7] (for reviews, see refs. 8–11).

The experimental challenges of organelle genomics are multiple. An important initial step, purification of organelle DNA, is addressed in another protocol of this issue. Moreover, the quantity of available organelle DNA (orgDNA) is often low and frequently cross-contaminated with other DNAs. Finally, the usually elevated A+T content of orgDNA significantly reduces cloning efficiency, and in extreme cases, sequence quality.

At present, two basic genomics strategies exist. The "whole-genome shotgun" strategy[12] has been used since the late 1980 in organelle genome sequencing. It implies random cloning and sequencing of genome fragments whose relative order is unknown. Reconstruction ("assembly") of the complete genome sequence relies on overlapping of the sequenced fragments. In contrast, a directed sequencing strategy, requiring *a priori* sequence information, is "primer walking"; a primer that binds to a known region generates new sequence, on the basis of which a second primer will be designed that proceeds further into the unknown sequence, and process continues. Further, PCR[13] is also a directed method and overcomes the necessity of cloning genome fragments. In practice, random and directed strategies are amalgamated in various combinations: for instance, cloning of random genome fragments and primer walking on these clones, or amplification of large genome fragments by long-PCR and cloning of random subfragments.

We have conducted whole-genome shotgun sequencing for nearly three decades and have established a robust and effective "random genome sequencing protocol." This protocol is best suited when microgram amounts of purified orgDNA are available and when the structure (number and topology of chromosomes) and size of the genome are unknown. More recently, we adopted and optimized the "long-PCR-based genome sequencing protocol," which is specifically adapted to small genomes of low DNA quantity and purity, but it presupposes either a single circular chromosome or detailed knowledge of the genome structure. There is no single fit-all approach in organelle genomics. The following type cases will help to select an optimal solution for a particular genome.

- When >1 μg orgDNA is available, contamination with DNAs from other sources is $<50\%$ and restriction analysis does not indicate a complex multichromosome genome organization, we recommend the random genome sequencing procedure. About 1 μg of orgDNA is usually ample for sequencing genomes up to 100 kbp with $>10\times$ coverage (contaminating sequences will be eliminated during the sequence assembly process).
- When the amount of orgDNA is between 0.2 and 1 μg, the random genome sequencing protocol is still suited, but cloning conditions must be meticulously optimized for yield. Extra-long readings from both ends of the insert and primer walking may compensate for a small number of clones.
- When the genome contains long (>1 kbp) repeat regions, the random genome sequencing protocol is recommended once more, but cloning conditions must be optimized for long inserts spanning the repeats. Readings must be long and from both ends of the insert; the assembly software should exploit positional information of forward and reverse readings. The final assembly should be confirmed by restriction analysis.
- When the organelle genome consists of a small, ideally circular-mapping chromosome and when sequences from closely related organisms are available for primer design (e.g., in animal mtDNAs), long-PCR is the strategy of choice. This protocol works well with total cellular DNA and minute quantities (~ 50 ng) of orgDNA.

Random genome sequencing protocol

Our protocol (**Fig. 1**) involves fragmentation of DNA by shearing with a nebulizer[14,15]. Fragment size is controlled by pressure, duration of treatment and viscosity of the solution. This technique is superior to sonication, which produces a high percentage of small fragments that are of little use for sequencing (<200 bp), and which preferentially break at A + T-rich regions. Fragmentation

with restriction enzymes is similarly unsuited, owing to the uneven size distribution of the resulting fragments.

DNA fragments are first sized by agarose gel electrophoresis to divide them into two size fractions, for example, small (0.5–1 kbp) and large (1–5 kbp) size fractions. The large-size fraction is preferred for library construction, whereas the small-size fraction is used only when the large one does not yield a sufficient number of clones or does not cover the entire genome due to cloning bias (regions including repeats are cloned more efficiently as smaller pieces).

DNA fragment termini are rendered blunt-ended by "end-polishing" using a combination of two types of DNA polymerases that fill/remove protruding ends. Then, fragments are phosphorylated, as phosphatases (potentially present in the agarose) may have removed 5′ phosphates that are essential for ligation with the vector. Note that the vector itself is dephosphorylated to avoid self-ligation. The ligation efficiency critically depends on the molar ratio of insert to vector, and the relative concentrations of DNA and ligase. Optimization of this step increases cloning efficiency most effectively.

Once a clone library is established, randomly chosen clones are sequenced, and dedicated software tools (such as Phred/Phrap[16]) assemble readings to reconstruct the whole-genome sequence. The presence of contaminant clones in a library (the extent of which depends on the purity of orgDNA) is usually compensated by sequencing of more clones to reach the desired level of sequence coverage (usually in the order of 10×). Yet, depending on the cost of sequencing reactions, it may be more economical to select "good" clones by colony hybridization.

The overall design of our random genome sequencing protocol is similar to that of the "random subclone generation" protocol developed by Roe *et al.*[17] (http://www.genome.ou.edu/protocol_book/protocol_partII.html). Yet, our protocol is optimized for small quantities and low purity of orgDNA. We recommend consulting Roe's protocols for additional information and additional references to the original literature.

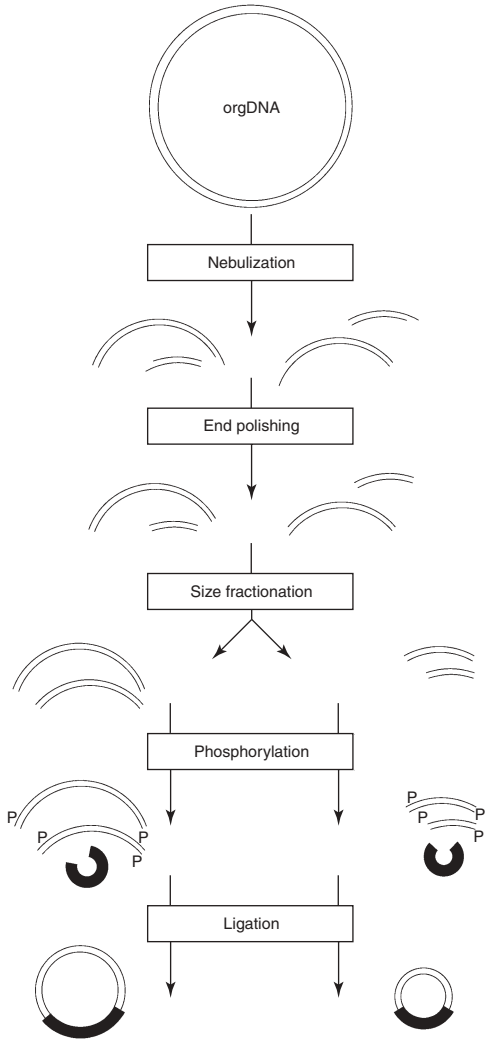

Figure 1 | Steps involved in the random genome sequencing protocol (protocol option A).

Long-PCR-based genome sequencing protocol

The long-PCR-based genome sequencing protocol (**Fig. 2**) has been developed primarily for animal mtDNAs[18–20]. Its success critically depends on the intactness of the DNA starting material. Therefore, mechanical stress should be avoided during cell disintegration and subsequent purification steps (i.e., CsCl gradient purification is not recommended). A simple procedure that works well for animal and protist mtDNA involves only two steps: (i) incubation of $\sim 2 \times 10^6$ cells for 1 h at 50 °C in PCR buffer containing 0.6 mg ml^{-1} proteinase K and 0.45% each of Nonident P40 and Tween 20 and (ii) inactivation of proteinase K at 95 °C for 10 min. Otherwise, the standard CTAB protocol[21] or more elaborate protocols[22] are sometimes required.

The long-PCR-based genome sequencing protocol starts with the amplification of "anchor" regions in the mtDNA (0.2–1.0 kbp). The heterologous "anchor primer" pairs target regions in most conserved mitochondrial genes such as *cox1* or *cob*. Primers that work well for metazoa are listed in **Table 1**. For other organismal groups, primer sequences can be either extracted from the literature or newly designed, based on multiple DNA sequence alignments of conserved genes from organisms closely related to the species of interest. The general rules for designing anchor primers are the same as for long-PCR primers described below, except that degenerate positions (e.g., N=A,T,G,C; R=A,G; Y=T,C) can be included.

The PCR product of the anchor region is sequenced and its mitochondrial origin is verified on the basis of expected sequence similarities, codon usage and A + T content. The anchor sequence serves for the design of long-PCR primers. These primer pairs are species specific, that is, they exactly match the anchor sequence and generally point outside of this region into the unknown portion of the genome (as shown in **Fig. 2**). Alternatively, they "face" each other, so that the genome is amplified in overlapping fragments and fully covering the anchor regions. Primers used for long-PCR are usually 23–25 nt long and contain 8–12 G or C residues. The nucleotide composition determines the temperature (T_m) at which 50% of a given primer is paired with the template. T_m (in °C) can be estimated by a simple formula: 2 × [(number of A's and T's) + 4 × (number of G's and C's)] (e.g., ref. 23). T_m-5 will be chosen as the annealing temperature in the PCR. The ideal primer has a balanced nucleotide composition, lacks internal repeats and tracts of homopolymers and is unable to form secondary structures and primer dimers. Primer pairs used in the PCRs should have similar T_m values. A number of computer programs are available for primer design (e.g., Primer3 http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi)[24]. Alternatively, primers can be easily designed manually,

using the principles described above and discussed in various publications[25]. For example, the manually designed "Tethya" primer pair (**Table 1**) allowed amplification of a complete demosponge mtDNA in two overlapping ~10 kb pieces[26].

Various combinations of long-PCR primers are usually tested, because it is typically uncertain which primers face each other on the genome and whether they are separated by an appropriate distance. In cases where amplification of a particular region proves difficult, a modified step-out PCR approach[27] can be used (see ref. 28): one primer is a species-specific long-PCR primer, whereas the other one is degenerate ("step-out primer"). The progressiveness of the DNA polymerase is critical. We preferentially use the Takara LA kit with an added hot start protocol similar to that devised for the Gene Amp XL PCR kit. If the Takara kit does not give satisfying results, we try other long-PCR kits.

After successful amplification, sequencing can proceed in two ways. Long-PCR products may be sequenced individually by primer walking. Alternatively, several PCR products may be combined in equimolar proportions for constructing a random library and sequencing of random clones.

Note that recent protocols (e.g., refs. 19,20) describe rolling circle amplification[29] for sequencing orgDNAs. Like long-PCR, this technique is best suited for intact and circular-mapping DNA, but it is more prone to generate multiple bands when orgDNA is contaminated with DNA from other sources.

Accessory protocols

Basic molecular cloning protocols are available in standard manuals (e.g., refs. 17,30).

Cloning vectors based on BlueScript or pUC have proven suitable for cloning inserts of up to 10 kbp. These vectors give high plasmid yields and are relatively tolerant toward repeats in the insert DNA. Nevertheless, any other vector systems may be used. The ability to distinguish colonies of insert-carrying clones from "empty" ones is essential because of the relatively low cloning efficiency of A + T-rich DNA. For BlueScript, distinction is possible by the white or blue color of colonies on YPTG/XGal-containing agar plates. We prepare vector free of *Escherichia coli* chromosomal DNA by CsCl/ethidium bromide gradient centrifugation. Vector DNA is linearized by restriction enzyme digestion (blunt-end cutter, e.g., *Eco*RV) and dephosphorylated to avoid religation. Finally, the vector is quality tested for the degree of dephosphorylation and cloning efficiency by using an orgDNA of known quality. Readily prepared vector DNA may be purchased.

DH5-α is an efficient, stable *E. coli* recipient strain for transformation. We prepare competent cells that, through addition of DMSO, can be stored frozen for several months[31]. The transformation efficiency is tested with a plasmid DNA standard. Commercially available competent cells may also be used, but their transformation efficiency is sometimes lower.

Cultures of *E. coli* recombinant clones require optimal aeration for high plasmid yield and purity; factors determining aeration are the surface to volume ratio of the culture medium and the shaking frequency. Plasmid preparation should avoid phenol, as traces of this reagent can lower the quality of the sequencing reaction. Most column-based plasmid isolation procedures work without phenol. Sequencing reactions using the dye-terminator technology yield high-quality readings. In our laboratory, the most frequent reasons for unsuccessful sequencing reactions have been traces of salts or ethanol in plasmid preparations. DNA samples containing an excess of remaining salt do not freeze at −20 °C; remaining ethanol can be detected by its scent. Finally, the cost of sequencing can be reduced at various steps. We regenerate columns used for plasmid purification up to 20 times without decrease in template quality. Regeneration (before use) consists of the following steps: wash columns with 1 ml each of H_2O (2×), renew buffer (2×), 1 N NaOH (1×) and ethanol (2×); for storage up to 1 week, wrap columns in a plastic foil and keep them at 4 °C. Finally, we downscale sequencing reactions with the dye-terminator kit to 1/16, without loss in read length.

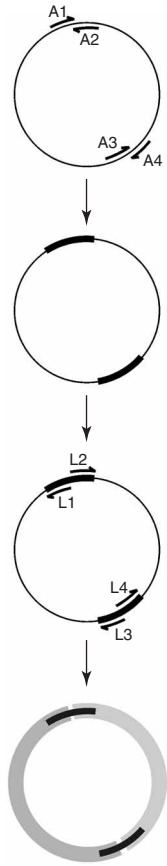

Figure 2 | PCR steps involved in the long-PCR-based genome sequencing protocol (protocol option B). The circle represents the organelle genome, A1–A4 are anchor primers and L1–L4 are long-PCR primers. Thick arcs indicate PCR-amplified regions of the genome: black arcs show anchor regions; gray arcs show long-PCR products. Anchor regions and long-PCR products together cover the entire organelle genome.

MATERIALS

REAGENTS

Option A

- Klenow polymerase
- T7 DNA polymerase, unmodified (New England Biolabs)
- T4 polynucleotide kinase
- T4 DNA ligase

TABLE 1 | Primers for animal mtDNAs (*Tethya* primers are specific for sponges).

Gene	Primer name	Sequence (5′→ 3′)	Reference
rnl	16Sar-5′	CGCCTGTTTATCAAAAACAT	23
	16Sbr-3′	CCGGTCTGAACTCAGATCACGT	
rns	12Sai-5′	AAACTAGGATTAGATACCCTATTAT	32
	12Sb-3′	GAGGGTGACGGGCGGTGTGT	
cob	cobF424	GGWTAYGTWYTWCCWTGRGGWCARAT	33
	cobR876	GCRTAWGCRAAWARRAARTAYCAYTCWGG	
cox1	LCO	GGTCAACAAATCATAAAGATATTGG	34
	HCO	TAAACTTCAGGGTGACCAAAAAATCA	
cox2	cox2F1	AAGCWAATWGGNCATCARTGRTATTG	This publication
	cox2R1	CTCCRCATATTTCNGARCATTGNCC	
cox3	cox3F	TGGTGGCGAGATGTKKTNCGNGA	This publication
	cox3R	ACWACGTCKACGAAGTGTCARTATCA	
nad4	nad4F	CCKAARGCYCAYGTKGARGCYCC	28
	nad4R	GARGAWCAKAWWCCRTGAGCAATYAT	
nad5	nad5F	TWYTATTAGGKTGAGATGGKYTNGG	35
	nad5R	TARAAKCCWGMTARAAAWGGKAWWCC	
rnl	Tethya-rnlF1	GAGAAGACCCCATTGAGCTTTAC	
	Tethya-rnlR1	TCGTTACTCTTTAGAAGGCAGTC	
cox1	Tethya-cox1F1	ATTGGGTTCTGCTTTTGTGGAAC	
	Tethya-cox1R1	TCTACTGATCCTCCTGAATGAGC	

- Dialysis tube 25 mm width, MWCO 12–14,000 (or smaller). The tubes are cut into desired sizes, boiled 2–3 times in TE to remove preservatives and stored refrigerated in TE
- TE: 10 mM Tris (pH 8) and 0.1 mM EDTA
- EtOH/AmAc ethanol (95% v/v), 0.5 M ammonium acetate (add anhydrous salt directly to ethanol)
- Klenow buffer 10×: 100 mM Tris (pH 8) and 50 mM $MgCl_2$ (store frozen)
- Kinase buffer 10×: 500 mM Tris (pH 7.6), 100 mM $MgCl_2$, 1 mM spermidine–HCl, 50 mM DTT and 1 mM EDTA (store frozen)
- Ligase buffer 10×: 200 mM Tris (pH 7.6), 50 mM $MgCl_2$ and 50 mM DTT (store frozen)
- Ligase dilution buffer: 1× ligase buffer plus 500 μg ml^{-1} BSA and 50% glycerol. Keep ≤3 months at −20 °C
- ATP (10 mM; store frozen)
- dNTP mix: 1 mM dATP, 1 mM dCTP, 1 mM dGTP and 1 mM dTTP (store frozen)
- NaCl (5 M)
- NaOH (50 mM)
- Glycogen (2 mg ml^{-1}; store frozen)
- Nebulized lambda DNA (5 ng μl^{-1}, in 1× loading buffer; store frozen)
- Loading buffer 5×: 10 mM EDTA, 50 mM Tris (pH 8), 0.04% bromophenol blue and 40% glycerol
- Agarose gel: 0.8–1% ordinary electrophoresis grade or low-melting agarose in 1× TAE; add 1 μl of ethidium bromide (stock: 10 mg ml^{-1}) per 10 ml agarose immediately before pouring the gel
- Electrophoresis buffer: 1× TAE (40 mM Tris acetate and 1 mM EDTA (pH 7.4–8.2)). Prepare 50× stock of 2 M Tris base, 0.95 M glacial acetic acid and 50 mM EDTA (pH 8)
- Electroelution buffer: 0.1× TBE (4.5 mM Tris, 0.05 mM EDTA and 3.5 mM borate (pH 8.3)). Prepare 10× stock, adjusting pH with boric acid.
- Agar plates: 1.5% agar, containing suitable antibiotics for selection of vector-containing colonies (e.g., 100 μg ml^{-1} ampicillin and 5 μg ml^{-1} tetracycline for BlueScript); 40 μl XGal (stock: 2% in dimethyl formamide) and 40 μl isopropyl β-D-1-thiogalactopyranoside (stock: 100 mM), if blue–white selection of clones is intended. The latter two reagents are spread onto the solidified agar
- Gel extraction column: QIAquick Gel Extraction kit (Qiagen, cat. no. 28104)
- PCR purification column: QIAquick PCR Purification kit (Qiagen, cat. no. 28704)
- Spin column: Montage PCR Centrifugal Filter Devices (Millipore, cat. no. UFC7PCR)

Option B

- Water: HPLC quality or filter (Millipore)-purified
- Long-PCR kit: TAKARA LA kit (Takara Bio Inc.) or Gene Amp XL PCR kit (Applied Biosystems)
- Paraffin cube: AmpliWax PCR Gem 100
- Oligonucleotide primers (standard quality)
- Gel extraction column: QIAquick Gel Extraction kit (Qiagen, cat. no. 28104)
- PCR purification column: QIAquick PCR Purification kit (Qiagen, cat. no. 28704)
- Spin column: Montage PCR Centrifugal Filter Devices (Millipore, cat. no. UFC7PCR)
- Loading buffer 5×: 10 mM EDTA, 50 mM Tris (pH 8), 0.04% bromophenol blue and 40% glycerol
- Agarose gel: 0.8–1% ordinary electrophoresis grade or low-melting-agarose in 1× TAE; add 1 μl of ethidium bromide (stock: 10 mg ml^{-1}) per 10 ml agarose immediately before pouring the gel
- Electrophoresis buffer: 1× TAE (40 mM Tris acetate and 1 mM EDTA (pH 7.4–8.2)). Prepare 50× stock of 2 M Tris base, ~57.1 ml glacial acetic acid and 50 mM EDTA (pH 8)

EQUIPMENT

Option A

- Nebulizer (Invitrogen, cat. no. K7025-05) or any (cheaper) medical nebulizer of similar design with a conical vessel. All these models require minor modification as described in the instructions of Invitrogen (http://www.invitrogen.com/content/sfs/manuals/nebulizer_man.pdf). Determine for each particular device the appropriate nebulization pressure (5–15 p.s.i.) and time (60–180 min) with a solution of 10 ng μl^{-1} lambda DNA and determine the degree of shearing by agarose gel electrophoresis
- Compressed N_2 container with pressure gauge, or pressure-adjustable compressed air
- Temperature-adjustable heating/cooling water bath
- Dry-heat block
- SpeedVac
- UV transilluminator
- Horizontal gel electrophoresis chamber
- Electroelution device or, alternatively dialysis tubes, closed on both ends, placed in electrophoresis chamber
- Refrigerated table-top centrifuge for microfuge tubes

Option B

- Thermocycler
- Horizontal gel electrophoresis chamber
- Refrigerated table-top centrifuge for microfuge tubes
- Thermocycler, for example, Applied Biosystems GeneAmp PCR System 9600 or 9700

PROCEDURE

1| As detailed in INTRODUCTION, option A is advised for orgDNA available in sufficient quantity (>1 μg) and if contamination with other DNA species is low (<50%). In cases where DNA quantity and purity are inferior, follow option B. All centrifuge steps are performed at maximum speed in a table top centrifuge (14,000*g* or more, depending on model).

(A) Random library construction

DNA fragmentation by nebulization ● TIMING 6–7 h

(i) Check quality, quantity and approximate size of organellar DNA by digesting 200–500 ng with a restriction enzyme (e.g., *Hin*dIII) and separate digested DNA on an agarose gel, together with a DNA marker of known concentration. Take a photo for documentation.

! CAUTION Ethidium bromide, which is added to agarose gel, is carcinogenic. It will diffuse from the gel into the migration buffer. Solutions should be handled with gloves.

(ii) If available, set aside 200 ng or more DNA for a later control (Step A(xxi) 21), or further restriction analysis, colony hybridization or PCR amplification for filling remaining sequencing gaps. Proceed with 0.2–5 μg DNA.

(iii) Adjust DNA solution (in TE) to 25% glycerol; final volume is 600 μl.

(iv) Fill DNA solution into the nebulizer chamber, connect device to compressed N_2 or air, and nebulize the DNA for 90–100 s at ~5 p.s.i. With our device, this condition produces ~70% fragments of 1–5 kbp in length.

(v) Verify the size distribution of nebulized DNA by electrophoresis on agarose gel together with a nebulized DNA standard of known concentration (e.g., phage lambda DNA). Load ~15–25 ng of the nebulized DNAs with a 1-kbp ladder DNA marker on each side. Take a photo for documentation.

▲ CRITICAL STEP In rare instances, the fragment sizes are too large and nebulization must be repeated. To increase the efficiency of fragmentation, pressure may be increased (or, less efficiently, the glycerol concentration decreased).

(vi) Recuperate DNA solution from chamber; recovery is ~500 μl or 85%.

(vii) Rinse chamber with ~100 μl H_2O, reduce rinsing liquid to 10–20 μl in SpeedVac and combine it with initially recuperated DNA solution in a microfuge tube.

▲ CRITICAL STEP Soak the nebulizer in water immediately after use. Otherwise, DNA sticking to the device will contaminate subsequent libraries. Wash the device by nebulizing with the following solutions (15 ml each; 20 p.s.i. for 2 min): 5 M NaCl; 50 mM NaOH; sterile H_2O, twice. Dry the device well before storage.

(viii) Add 1 μl of glycogen solution, precipitate DNA with EtOH/AmAc (at least for 1 h or overnight), centrifuge for 20 min at 4 °C and discard the supernatant. Wash the DNA pellet with 80% EtOH, centrifuge it for 10 min at 4 °C and discard the supernatant. Remove any remaining EtOH with a pipette. Dry pellet shortly and dissolve it in 10–20 μl TE.

▲ CRITICAL STEP Remaining traces of EtOH will inhibit DNA polymerases in subsequent steps.

■ PAUSE POINT DNA can be stored for weeks.

End repair ● TIMING 1.5–2 h

(ix) Use all the nebulized material. The following recipe is for 0.25–1.0 μg DNA. For >1 μg DNA, increase the volume and all ingredients accordingly.

End repair

Volume	Reagent	Final amount/concentration
… μl	Nebulized DNA in TE	~1.0 μg
1.5 μl	dNTP(10×)	1×
… μl	Klenow DNA polymerase	2.0 U
… μl	T7 DNA polymerase	2.0 U
1.5 μl	Klenow buffer (10×)	1×
… μl	H_2O	
15 μl	Total volume	

(x) Incubate for 30–60 min at 12 °C. Stop the reaction by adding EDTA to a final concentration of 10 mM.

(xi) Heat the DNA-containing tube for 10 min at 65 °C in a dry bath; remove metal block with the tube and let cool down to room temperature.

▲ CRITICAL STEP Cooling must take place slowly to allow proper reannealing of the DNA strands. This is especially important when DNA contains excessively long stretches of A + T or repeats.

Fragment sizing by agarose electrophoresis ● TIMING 3 h

(xii) Prepare a 0.8–1% low-melting agarose gel by using a comb that forms sufficiently large wells. For DNA quantities of up to 1.5 μg, wells holding ~20–25 μl are suitable (teeth width 0.5 cm). For larger DNA quantities, choose a comb with broader teeth.

▲ CRITICAL STEP To avoid contamination with DNA from previous experiments, clean comb and electrophoresis chamber by soaking them in 0.2% SDS, followed by several rinses with sterile H_2O. Use fresh electrophoresis buffer and change tips when loading DNA samples and marker.

(xiii) Adjust DNA quantities of ≤1.5 µg to a volume of 12 µl (with TE), and add 3 µl of loading buffer. For larger DNA quantities, scale up accordingly. If necessary, concentrate the DNA in a SpeedVac. Load samples on low-melting agarose gel, together with a 1-kbp DNA ladder as size marker on each side. Migrate at ~2 V cm^{-1} until the 0.5 and 3 kbp bands of the marker are separated by ~1 cm.
▲ CRITICAL STEP Size separation will be imperfect if DNA is too concentrated. The loading volume must easily fit into the well.

(xiv) Visualize ethidium bromide-stained DNA on a transilluminator in "preparation mode" (low-energy wavelength). Take a photo for documentation. Using a scalpel, cut three DNA size fractions from the gel: 1.0–5.0 kbp (large), 0.5–1.0 kbp (small) and 0.5–0.1 kbp (minuscule; this fraction will not be cloned but may be used as a hybridization probe; store the agar block in a freezer until use). Proceed with the large- and small-size fractions.
▲ CRITICAL STEP It is important to avoid UV-induced breakage of DNA by minimizing exposure time and intensity.

Electroelution ● TIMING 4 h

(xv) To electroelute DNA from agarose blocks, use a dedicated device or enclose each agarose block in a dialysis tube together with 200 µl TE. Use gloves when handling the dialysis tube. Here, the electrophoresis buffer is 0.1× TBE. Preferentially, place the agarose block opposite to the direction of the electrical field used for separation in Step 13.

(xvi) Electroelute the DNA at ~12 V cm^{-1} for 60–75 min depending on the fragment size. At the end of electroelution, invert electrode poles for 1 min.
▲ CRITICAL STEP Inverting poles detaches the DNA from the dialysis membrane, thus increasing DNA yield.

(xvii) Collect DNA solution and transfer it into a microfuge tube. Rinse the chamber with 100 µl TE. Also rinse the dialysis membrane with 100 µl TE by vortexing it in a microfuge tube; then fasten one corner of the membrane in the lid of the tube and centrifuge for 1 min to remove all liquid from the membrane. Combine all rinsing solutions with the initially recuperated DNA solution.

(xviii) Centrifuge the DNA solution at maximum speed for 15 min in a table-top centrifuge (room temperature) to remove agarose pieces. Transfer supernatant into a fresh tube.

(xix) Add 1 µl glycogen solution to the electroeluted DNA and precipitate by adding 2–2.5 volumes of EtOH/AmAc. Chill at −20 °C for at least 1 h.
■ PAUSE POINT DNA can be left in the freezer indefinitely.

(xx) Centrifuge precipitated DNA in a table-top centrifuge for 20–30 min at 4 °C and discard the supernatant. Wash pellet with 80% EtOH, centrifuge it for 10 min at 4 °C and discard the supernatant. Air-dry pellet for several minutes. Dissolve pellet in 10–30 µl TE depending on the DNA quantity.
▲ CRITICAL STEP Remove remaining EtOH thoroughly with a pipette, as residues of alcohol may inhibit subsequent enzymatic reactions.

Analytic agarose gel ● TIMING 2 h

(xxi) Assume 50–80% recovery of DNA after sizing and electroelution. Load an estimated 15–25 ng of the small- and large-size DNA fractions on agarose gel together with different concentrations (5, 10, 15, 20 and 25 ng) of a nebulized lambda DNA standard and 15–25 ng of unprocessed orgDNA. Estimate factual concentration of the small- and large-size DNA fractions. Take a photo for documentation.
■ PAUSE POINT DNA solution in TE can be left at 4 or −20 °C for several months.

Phosphorylation ● TIMING 1 h

(xxii) Conduct a test experiment (phosphorylation, ligation and transformation) with 50 ng large-size orgDNA fraction. Once all steps are optimized in subsequent experiments, the entire amount required for a desired number of clones can be processed, by scaling up components of the reactions proportionally. Occasionally the test experiment will already yield a sufficient number of clones for sequencing the complete genome.

Phosphorylation

Volume	Reagent	Final amount/concentration	Note
… µl	DNA	50 ng	DNA is dissolved in TE
… µl	Kinase buffer 10×	1×	Alternatively, ligase buffer
… µl	ATP (stock 10 mM)	1 mM	
… µl	T4 polynucleotide kinase	5 U kinase	
… µl	H_2O		
5–10 µl	Total volume		

Incubate phosphorylation reaction for 30 min at 37 °C. Inactivate kinase at 65 °C for 20 min, cool down slowly (see Step A(xi).

▲ CRITICAL STEP It is important to maintain a small reaction volume to allow the material to be used directly for ligation. Concentration of the sample in the SpeedVac is not advised, because ATP and spermidine concentrations would increase.

Ligation ● TIMING 2 h/overnight

(xxiii) Add the ingredients in the order shown in the table and mix well after each addition. The volumes given in the table are calculated for a 2.9-kbp BlueScript vector (dephosphorylated) and for genomic (insert) DNA (phosphorylated) of 1.5 kbp average size.

Ligation

Volume	Reagent	Final amount/concentration
1.2 μl	H_2O	
0.4 μl	Ligase buffer 10×	1×
0.3 μl	ATP (stock 10 mM)	Maximum 1 mM final
0.5 μl	Vector BlueScript (stock 20 ng μl^{-1})	10 ng
1.0 μl	Insert DNA (10 ng μl^{-1}; ~1.5 kbp)	Molar ratio: 2 insert + 1 vector
0.6 μl	T4 DNA ligase (1 U μl^{-1})	0.15 U μl^{-1}
4.0 μl	Total volume	Total DNA: 5 ng μl^{-1}

▲ CRITICAL STEP The ligation reaction should comply with the following requirements:

- the optimal final concentration of total DNA is 5 ng μl^{-1};
- the optimal molar ratio of phosphorylated insert DNA to vector is 2 insert to 1 vector;
- account for the carry-over of ATP and buffer from the kinase reaction.

Model calculation for the reaction setup:

1. 10 ng of vector is used for each ligation reaction.
2. The amount of insert needed for a molar ratio of 2 inserts to 1 vector depends on the size of the insert.
Example: for an average insert size of 1.5 kbp, the ratio is (2 × 1.5 kbp)/2.9 kbp = ~1 ng insert per ng vector.
3. The final reaction volume is defined by the envisaged final DNA concentration (5 ng μl^{-1}) and the amount of vector (10 ng) used per ligation. Example: (10 ng vector + 10 ng insert)/5 ng μl^{-1} = 4 μl.
4. To reach a final concentration of 1 mM ATP, given that the insert DNA already contains ATP, the volume of ATP stock to be added is ($V_{final} - V_{insert}$)/[ATP stock]. Example: (4 μl − 1 μl)/10 mM = 0.3 μl.
5. To reach a final concentration of 1× ligation buffer, the volume of 10× ligation buffer to be added is (V_{final} – V_{insert})/10×. Example: (4 μl − 1 μl)/10 = 0.3 μl.
6. The ligase concentration should be 0.15 U μl^{-1}. Prepare a dilution, if necessary.

(xxiv) Incubate the reaction for 14 h at 14 °C (or overnight).

(xxv) Stop the reaction containing 10 ng vector by the addition of 96 μl TE. The final vector concentration should be 0.1 ng μl^{-1}; the final EDTA concentration should be 1 mM.

■ PAUSE POINT Ligation mix can be stored at 4 °C for several months; alternatively, it can be precipitated and stored at −20 °C indefinitely.

Transformation ● TIMING 2 h/overnight

(xxvi) Take 10 μl ligation mix (containing 1 ng of vector) and transform 200 μl competent *E. coli* cells following standard protocols[30]. Distribute transformed cells on three agar plates containing the appropriate antibiotics, and XGAL and IPTG if indicated. Expect 50–200 blue and 20–150 white colonies in a ratio of 1:1 to 3:1. If the result is inferior, repeat ligation by adapting amounts and ratios of insert and vector DNAs.

? TROUBLESHOOTING

Assessment of cloning efficiency

(xxvii) Prepare plasmid DNA from 20 or more white colonies to determine insert sizes. Any rapid extraction protocol for screening[17,30] or a regular column-based protocol is suited for this step. When the vector used is BlueScript, also prepare plasmid from a few blue colonies as a size reference. Sequence about 50 insert-carrying clones and analyze the sequence *in silico* to determine the proportion of "good" versus unknown versus contaminant inserts. Calculate the number of ligation reactions required to obtain the desired number of clones. If the large-size DNA fraction is insufficient in quantity, use the small fraction.

In case cloning of precious DNA material results in a low relative yield of orgDNA-containing clones, colony hybridization will help to identify them. This procedure involves the binding of recombinant plasmid DNA to a membrane (e.g., in a 96-hole microtiter plate pattern), *in situ* denaturation of the plasmid DNA and hybridization of the membranes (up to 50 membranes at a time) with a labeled orgDNA probe. The probe may have low purity (down to 20%), as the complexity of the orgDNA is usually much lower than that of potential contaminants. Alternatively, insert-carrying clones can be

enriched by bulk plasmid preparation of all clones of a library, sizing of plasmid DNA and retransformation of the large plasmid fraction.

? TROUBLESHOOTING

DNA sequencing

(xxviii) Sequence randomly selected clones until genome assembles into ~2 or ~10 large contigs (for ~20 or ~200 kbp genomes, respectively). Fill remaining gaps by primer walking. If no clones that span the gaps exist, amplify the uncloned regions by PCR and proceed as in Step 12 of the long-PCR-based genome sequencing protocol.

(B) Long-PCR-based genome sequencing

Amplification of anchor regions ● TIMING **6 h**

(i) Amplify three to five anchor regions, that is, 200–500-bp-long genomic regions from parts of highly conserved genes by using 0.01–1 μg of total cellular DNA per reaction. Some successful primers for animal mtDNA are listed in **Table 1**. For primer design, follow rules described in INTRODUCTION.

(ii) Load 1/10 volume of the PCR product on a 1% agarose gel, together with a size marker, and estimate the size and quantity of the PCR product.

! CAUTION Ethidium bromide is carcinogenic. It will diffuse from the gel into the migration buffer. Wear gloves.

? TROUBLESHOOTING

(iii) Purify the PCR product of each anchor region with a spin column.

(iv) Determine the DNA sequence of the anchor regions by direct sequencing the PCR product with the PCR primers.

(v) For each amplified and sequenced anchor region, design long-PCR primers that are in opposite direction compared to the "anchor" primers (see **Fig. 2**).

Long-PCR setup ● TIMING **30 min to 1 h**

(vi) Choose a combination of primers and include positive and negative controls. If amplification is unsuccessful, try different primer combinations.

Set up the long-PCR reaction mix for hot start amplification. In the reaction tube, a wax plug will separate the lower reagent mix (primers) from the upper reagent mix (enzyme + DNA).

Final PCR volume is 100 μl. (For multiple simultaneous PCRs, prepare master mixes of the lower and upper reagent mixes without primers and DNA, aliquot master mixes and then add the particular primer pairs and DNA to the aliquots.)

Standard long-PCR (for an alternative long-PCR protocol, see **Box 1**)

(vii) Add the following components (lower PCR reagent mix) into a PCR tube.

Lower PCR reagent mix

Volume	Reagent	Final amount/concentration (Step 8)
20.0 μl	H_2O	
4.0 μl	LA buffer 10x	1×
8.0 μl	dNTPs (stock 10 mM each)	800 μM each
4.0 μl	Primer no. 1	40 pmol
4.0 μl	Primer no. 2	40 pmol
40.0 μl	Total volume	

▲ CRITICAL STEP To avoid potential contamination, use cotton-plugged tips, wear disposable gloves and aliquot all PCR reagents.

(viii) Add one wax cube to the PCR tube and melt it by incubating the tube at 80 °C for 5 min. Allow the wax to solidify at room temperature for 5 min.

(ix) Combine the following components (upper PCR reagent mix) in a separate microfuge tube and mix them.

Upper PCR reagent mix

Volume	Reagent	Final amount/concentration
53.0 μl	H_2O	
6.0 μl	LA buffer 10×	1×
0.5 μl	Takara LA *Taq* polymerase	2.5 U
0.5 μl	DNA	0.01–1 μg of total cellular DNA
60.0 μl	Total volume	

(x) Layer the upper reagent mix onto the wax-covered lower reagent mix, without piercing the wax.

Long-PCR amplification ● TIMING **8–10 h**

(xi) Incubate the PCR in a thermocycler. For the Applied Biosystems GeneAmp PCR System 9600 and 9700, the following times and temperatures work well. Note that the temperature decrement is optional but it helps to avoid mispriming[36].

BOX 1 | ALTERNATIVE STEP-OUT LONG PCR (ALTERNATIVE TO STEPS B(vii)–B(xii))

In cases where long-PCR amplification of a particular region does not give a satisfying product, use a modified step-out PCR procedure. Successful step-out primers are listed in table shown below.

Step-out primers

Primer	Sequence (5′→3′)
Step-out 1	GTCAGTCAGTCAGANNNNAGA
Step-out 2	TCAGGAACGATCGTNNNNTCA
Step-out 3	AACAAGCCCACCAAAATTTNNNATA
Step-out 4	TTGTTCGGGTGGTTTTAAANNNTAT

1. Add the following components (lower PCR reagent mix) into a PCR tube.

Lower PCR reagent mix

Volume	Reagent	Final amount/concentration (Step 10)
20.0 μl	H_2O	
4.0 μl	LA buffer 10×	1×
8.0 μl	dNTPs (stock 10 mM each)	800 μM each
4.0 μl	Step-out primer	25 pmol
36.0 μl	Total volume	

2. Add one wax cube and melt it by incubating the reaction tube for 5 min at 80 °C. Allow the wax to solidify at room temperature for 5 min.

3. Combine in a separate microfuge tube the following components (upper PCR reagent mix) and mix them.

Upper PCR reagent mix

Volume	Reagent	Final amount/concentration
53.0 μl	H_2O	
6.0 μl	LA buffer 10×	1×
0.5 μl	Takara LA *Taq* polymerase	2.5 U
0.5 μl	DNA	0.01–1 μg total cellular DNA
60.0 μl	Total volume	

4. Layer the upper reagent mix onto the wax-covered lower reagent mix without piercing the wax.

Step-out long-PCR amplification ● TIMING 8–10 h

5. Incubate the PCR in a thermocycler. For the Applied Biosystems GeneAmp PCR System 9600 and 9700, use the following settings:
   Hold: 94 °C for 1 min
   Cycle: (94 °C for 20 s; 30 °C for 2 min; (ramp time = 2 min); 68 °C for 8 min) × 1 cycle
   Pause to add 4 μl species-specific primer
   Auto: (94 °C for 20 s; 65 °C (decrement = 0.3 °C per cycle) for 20 s; 68 °C for 8 min) × 16 cycles
   Auto: [94 °C for 20 s; 60 °C for 20 s; 68 °C for 8 min (increment = 15 s per cycle)] × 19 cycles
   Hold: 68 °C for 12 min
   Hold: 4 °C non-stop
   ▲ CRITICAL STEP It is important to pause rather than stop the thermocycler to maintain the high temperature that prevents mispriming. For the same reason, do not remove PCR tubes from the heating block when adding the primer.

6. Load 5 μl of the PCR on a 1% agarose gel together with a size marker and verify the size and quantity of the PCR product.

Hold: 94 °C for 1 min
Auto: (94 °C for 20 s, 65 °C (decrement = 0.3 °C per cycle) for 20 s, 68 °C for 8 min) × 16 cycles
Auto: [94 °C for 20 s, 60 °C for 20 s, 68 °C for 8 min (increment = 15 s per cycle)] × 19 cycles
Hold: 68 °C for 12 min
Hold: 4 °C non-stop

(xii) Load 5 μl of the PCR on a 1% agarose gel together with size marker, and verify the size and quantity of the PCR product. If the expected product is obtained, move to Step B(xiii).
? TROUBLESHOOTING

DNA sequencing

(xiii) If the long-PCR generates a single product, purify the product with a spin column according to the supplier's instructions. Directly sequence the product with the PCR primers and subsequently by primer walking[37]. If multiple products occur in a single reaction, separate them on a low-melting agarose gel, cut out bands and purify them with a gel extraction column according to the supplier's instructions. Cloning of this material is advised, because direct sequencing of gel-extracted material often gives low-quality readings.
The long-PCR products from multiple reactions can be combined in equimolar proportions for construction of a random library (see Random genome sequencing protocol).

● **TIMING**

Step A(i–xxvii): a library can be constructed in 1 working week
Monday: determine DNA concentration on agarose gel (3 h), nebulize and precipitate DNA (2 h); pour low-melting agarose gel (1 h)
Tuesday: end repair (2 h), separate DNA by agarose gel electrophoresis and excise size fractions from gel (3 h)
Wednesday: electroelute size fractions (4 h), precipitate and determine size and quantity of DNA on analytical agarose gel (2 h)
Thursday: calculate recipe for phosphorylation and ligation (1 h), phosphorylate (2 h), set up ligation (30 min); incubate reaction overnight
Friday: transformation (2 h); incubate plates for 16–18 h overnight
Saturday: store agar plates at 4 °C
Monday: count colonies (1 h); inoculate cultures of white (and a few blue) clones (2 h)
Tuesday: prepare plasmid DNA (2 h), determine yield and insert sizes on agarose gel (2 h)
Wednesday: sequencing
Thursday: If the outcome is satisfying but the number of clones obtained in the transformation is insufficient, transform the residual ligation mix. See ANTICIPATED RESULTS for guidelines.
Three additional days are needed to evaluate the library.

Step B(i), standard PCR setup for amplification of anchor regions: 1–2 h
Step B(i), amplification of anchor regions: 3–5 h
Step B(ii), agarose gel: 2–3 h
Step B(vi), long-PCR setup: 1–2 h
Step B(vii–xi), long-PCR: 8–10 h, usually overnight

? TROUBLESHOOTING

Troubleshooting advice can be found in **Table 2**.

TABLE 2 | Troubleshooting table.

Step	Problem	Possible reason	Solution
A(xxvi)	The number of blue colonies per ng vector transformed is elevated (>250)	Vector is not sufficiently dephosphorylated	Compare test ligation of phosphorylated and dephosphorylated vectors. The efficiency should differ by a factor of >50 and the number of white colonies should be <2%
A(xxvi)	The number of blue colonies per ng vector transformed is acceptable (100–200), but the percentage of white colonies is low <20%	(i) Inefficient phosphorylation of insert DNA; (ii) insufficient amount of insert DNA in ligation experiment; (iii) impurities in orgDNA	(i) Redo phosphorylation with new enzyme and new buffers; (ii) check DNA concentrations and repeat experiment; (iii) repurify orgDNA. A more general option is the use of the small-size fraction
A(xxvi)	The number of both blue and white colonies per ng insert is too small	Inefficient ligation. Transformation is inefficient due to insufficiently competent cells	Redo ligation with new enzyme and solutions. Test transformation with a plasmid of known concentration
A(xxvii)	Elevated proportion (>80%) of white colonies without insert	Insufficient antibiotics (degraded with time) in agar plates	Test plates with plasmid-less *E. coli*. Prepare fresh plates
A(xxvii)	Elevated proportion (>80%) of white colonies without insert and many plasmids are smaller than vector	Abortive inserts due to repeats, or sequences not tolerated by *E. coli*	Clone small insert size fraction or use long-PCR-based genome sequencing protocol

TABLE 2 | Troubleshooting table (continued).

Steps	Problem	Possible reason	Solution
B(ii, xii)	Low yield of PCR product	Inhibitory contaminants in DNA (test by adding DNA to an otherwise functioning PCR)	Reduce the amount of DNA or reamplify the amplified DNA in a second PCR
B(xii)	Low yield of PCR product	Large non-coding ("control") region	Try different polymerases and/or different cycling conditions; use step-out protocol
B(xii)	No products in long-PCR	Too large distance between primers	Change polymerase or generate more anchor regions
		Both primers bind to same DNA strand	Try different primer combinations or generate more anchor regions
		One of the two primers was designed on the basis of contaminant DNA sequence	Try different primer combinations or generate more anchor regions
		Linear chromosome	Use step-out PCR to sequence ends of linear chromosome
		Multichromosome genome	Use random genome sequencing protocol
B(xii)	Multiple bands	One of the primers has several binding sites	Increase annealing temperature. Try different primer combinations or generate more anchor regions
		Strong secondary structures interfere with amplification	Increase denaturation temperature and time (96 °C for 30 s)

ANTICIPATED RESULTS

Option A

Sizing and electroelution (Step 21). About 50–80% DNA is recovered after sizing and electroelution of the size fractions. **Figure 3** shows a photo of an analytical agarose gel.

Cloning (Steps 22–24). The number of white colonies per ng of insert DNA is $\geq$20, and the relative number of blue colonies is $\leq$3:1.

Insert size determination (Step 25). Number of insert-containing versus insert-less clones $\geq$5:1. Number of clones with insert sizes within the expected size range versus smaller sizes $\geq$80%.

DNA sequencing (Step 26). Number of recognized organelle sequences ("good" clones) versus unidentified sequences versus contaminant sequences $\geq$1:1:1. There is no strict rule as to an acceptable proportion of "good" versus contaminant clones. Evidently, the lower this proportion, the more costly DNA sequencing will be. As a rule of thumb, a genomic library should contain at least ten "good" clones per kbp of genome. This rule is based on a 10$\times$ sequence coverage, assuming on average 1,000 bp high-quality, new sequence per clone when sequencing from both ends. The exact number of clones to be sequenced is calculated on the basis of the following data (case of blue–white colony selection): the yield of white colonies in a test ligation/transformation, performed with 1 ng of vector DNA; the percentage among white clones that carry an insert (sample size $\geq$20

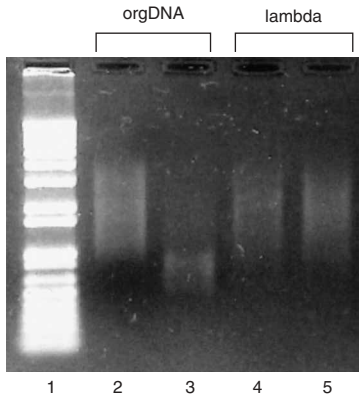

Figure 3 | Nebulized and size-separated orgDNA. Analytical agarose gel, Step A(xxi) (random genome sequencing protocol, protocol option A). Lane 1, 1 kbp + DNA ladder; lane 2, nebulized orgDNA, large-size fraction (1.0–5.0 kbp); lane 3, nebulized orgDNA, small-size fraction (0.5–1.0 kbp); lanes 4 and 5, 15 and 25 ng nebulized lambda DNA.

clones); and the ratio of clones whose inserts are of the desired organelle origin, to those of other origins, deduced from the DNA sequence (sample size ideally above 50 clones).

Option B

For animal mtDNAs, often three out of five anchor regions amplify and generally the complete genome can be amplified in two to three overlapping fragments of 6–12 kbp. The yield of standard and long-PCR is usually 3–10 µg product per reaction. In most cases, a single PCR product per reaction is obtained, which can be used directly for sequencing and/or random library construction.

ACKNOWLEDGMENTS We thank Shona Teijeiro for critical reading of the manuscript and for help in the preparation of the figures. This work was supported by the Canadian Institute of Health Research (B.F.L., grant MOP-42475; G.B., grant MOP-79309; postdoctoral fellowship to D.V.L.). Interaction support from the Canadian Institute of Advanced Research, Program in Evolutionary Biology (G.B. and B.F.L.) is gratefully acknowledged.

COMPETING INTEREST STATEMENT The authors declare that they have no competing financial interests.

Published online at http://www.natureprotocols.com
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions

1. Boer, P.H. & Gray, M.W. Scrambled ribosomal RNA gene pieces in *Chlamydomonas reinhardtii* mitochondrial DNA. *Cell* **55**, 399–411 (1988).
2. Burger, G., Forget, L., Zhu, Y., Gray, M.W. & Lang, B.F. Unique mitochondrial genome architecture in unicellular relatives of animals. *Proc. Natl. Acad. Sci. USA* **100**, 892–897 (2003).
3. Marande, W., Lukes, J. & Burger, G. Unique mitochondrial genome structure in diplonemids, the sister group of kinetoplastids. *Eukaryot. Cell* **4**, 1137–1146 (2005).
4. Gray, M.W. Diversity and evolution of mitochondrial RNA editing systems. *IUBMB Life* **55**, 227–233 (2003).
5. Gray, M.W. & Covello, P.S. RNA editing in plant mitochondria and chloroplasts. *FASEB J.* **7**, 64–71 (1993).
6. Michel, F., Jacquier, A. & Dujon, B. Comparison of fungal mitochondrial introns reveals extensive homologies in RNA secondary structure. *Biochimie* **64**, 867–881 (1982).
7. Dujon, B., Colleaux, L., Jacquier, A., Michel, F. & Monteilhet, C. Mitochondrial introns as mobile genetic elements: the role of intron-encoded proteins. *Basic Life Sci.* **40**, 5–27 (1986).
8. Lang, B.F., Gray, M.W. & Burger, G. Mitochondrial genome evolution and the origin of eukaryotes. *Annu. Rev. Genet.* **33**, 351–397 (1999).
9. Gray, M.W., Lang, B.F. & Burger, G. Mitochondria of protists. *Annu. Rev. Genet.* **38**, 477–524 (2004).
10. Bonen, L. & Vogel, J. The ins and outs of group II introns. *Trends Genet.* **17**, 322–331 (2001).
11. Burger, G., Gray, M.W. & Lang, B.F. Mitochondrial genomes—anything goes. *Trends Genet.* **19**, 709–716 (2003).
12. Bankier, A.T., Weston, K.M. & Barrell, B.G. Random cloning and sequencing by the M13/dideoxynucleotide chain termination method. *Methods Enzymol.* **155**, 51–93 (1987).
13. Mullis, K.B. & Faloona, F.A. Specific synthesis of DNA *in vitro* via a polymerase-catalyzed chain reaction. *Methods Enzymol.* **155**, 335–350 (1987).
14. Okpodu, C.M., Robertson, D., Boss, W.F., Togasaki, R.K. & Surzycki, S.J. Rapid isolation of nuclei from carrot suspension culture cells using a BioNebulizer. *Biotechniques* **16**, 154–159 (1994).
15. Surzycki, S. in *The International Conference on the Status and Future of Research on the Human Genome.* Human Genome II (San Diego, CA), pp.51 (1990).
16. Ewing, B. & Green, P. Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res.* **8**, 186–194 (1998).
17. Roe, B., Crabtree, J.S. & Khan, A.S. *DNA Isolation and Sequencing* (John Wiley and Sons, New York, 1996).
18. Cheng, S., Fockler, C., Barnes, W.M. & Higuchi, R. Effective amplification of long targets from cloned inserts and human genomic DNA. *Proc. Natl. Acad. Sci. USA* **91**, 5695–5699 (1994).
19. Boore, J.L., Macey, J.R. & Medina, M. Sequencing and comparing whole mitochondrial genomes of animals. *Methods Enzymol.* **395**, 311–348 (2005).
20. Jansen, R.K. *et al.* Methods for obtaining and analyzing whole chloroplast genome sequences. *Methods Enzymol.* **395**, 348–384 (2005).
21. Saghai-Maroof, M.A., Soliman, K.M., Jorgensen, R.A. & Allard, R.W. Ribosomal DNA spacer-length polymorphisms in barley: mendelian inheritance, chromosomal location, and population dynamics. *Proc. Natl. Acad. Sci. USA* **81**, 8014–8018 (1984).
22. Berntson, E.A., France, S.C. & Mullineaux, L.S. Phylogenetic relationships within the class Anthozoa (phylum Cnidaria) based on nuclear 18S rDNA sequences. *Mol. Phylogenet. Evol.* **13**, 417–433 (1999).
23. Palumbi,, S.R. in *Molecular Phylogenetics* (eds. Hillis, D.M., Moritz, C. & Mable, B.K.) 205–247 (Sinauer Associates Inc., Sunderland, MA, 1996).
24. Rozen, S. & Skaletsky, H. Primer3 on the WWW for general users and for biologist programmers. *Methods Mol. Biol.* **132**, 365–386 (2000).
25. Robertson, J.M. & Walsh-Weller, J. An introduction to PCR primer design and optimization of amplification reactions. *Methods Mol. Biol.* **98**, 121–154 (1998).
26. Lavrov, D.V., Forget, L., Kelly, M. & Lang, B.F. Mitochondrial genomes of two demosponges provide insights into an early stage of animal evolution. *Mol. Biol. Evol.* **22**, 1231–1239 (2005).
27. Wesley, U.V. & Wesley, C.S. Rapid directional walk within DNA clones by step-out PCR. *Methods Mol. Biol.* **67**, 279–285 (1997).
28. Shao, Z., Graf, S., Chaga, O.Y. & Lavrov, D.V. Mitochondrial genome of the moon jelly *Aurelia aurita* (Cnidaria, Scyphozoa): a linear DNA molecule encoding a putative DNA-dependent DNA polymerase. *Gene* **381**, 92–101 (2006).
29. Fire, A. & Xu, S.Q. Rolling replication of short DNA circles. *Proc. Natl. Acad. Sci. USA* **92**, 4641–4645 (1995).
30. Sambrook, J., Fritsch, E.F. & Maniatis, T. *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory Press, New York, NY, 1989).
31. Inoue, H., Nojima, H. & Okayama, H. High efficiency transformation of *Escherichia coli* with plasmids. *Gene* **96**, 23–28 (1990).
32. Kocher, T.D. *et al.* Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers. *Proc. Natl. Acad. Sci. USA* **86**, 6196–6200 (1989).
33. Boore, J.L. & Brown, W.M. Mitochondrial genomes of *Galathealinum*, *Helobdella*, and *Platynereis*: sequence and gene arrangement comparisons indicate that Pogonophora is not a phylum and Annelida and Arthropoda are not sister taxa. *Mol. Biol. Evol.* **17**, 87–106 (2000).
34. Folmer, O., Black, M., Hoeh, W., Lutz, R. & Vrijenhoek, R. DNA primers for amplification of mitochondrial cytochrome *c* oxidase subunit I from diverse metazoan invertebrates. *Mol. Mar. Biol. Biotechnol.* **3**, 294–299 (1994).
35. Lavrov, D.V., Brown, W.M. & Boore, J.L. Phylogenetic position of the Pentastomida and (pan)crustacean relationships. *Proc. Biol. Sci.* **271**, 537–544 (2004).
36. Don, R.H., Cox, P.T., Wainwright, B.J., Baker, K. & Mattick, J.S."Touchdown" PCR to circumvent spurious priming during gene amplification. *Nucleic Acids Res.* **19**, 4008 (1991).
37. Kusukawa, N., Uemori, T., Asada, K. & Kato, I. Rapid and reliable protocol for direct sequencing of material amplified by the polymerase chain reaction. *Biotechniques* **9**, 66–68 70, 72 (1990).