Construction of Small-Insert and Large-Insert Metagenomic Libraries

Carola Simon and Rolf Daniel

Abstract

The vast majority of the Earth’s biological diversity is hidden in uncultured and yet uncharacterized microbial genomes. The construction of metagenomic libraries is a cultivation-independent molecular approach to assess this unexplored genetic reservoir. In the last few years, a high number of novel biocatalysts have been identified by function-based or sequence-based screening of metagenomic libraries. Here, we describe detailed protocols for the construction of metagenomic small-insert and large-insert libraries in plasmids and fosmids, respectively, from environmental DNA.

Key words: Metagenomic DNA, Small-insert library, Large-insert library, Plasmid, Fosmid, Whole genome amplification

1. Introduction

The construction and screening of metagenomic libraries that have been generated from DNA directly isolated from environmental samples have been proven to be a powerful tool for the recovery of novel biomolecules of biotechnological importance (1, 2). In principle, metagenomic libraries provide access to the entire gene content of a habitat (2). The construction of metagenomic libraries involves the same steps as the cloning of genomic DNA derived from individual microorganisms. The required steps include fragmentation of environmental DNA by restriction digestion or shearing, insertion into an appropriate vector system, and transformation of the recombinant vectors into a suitable host, which is in almost all published studies on construction of metagenomic libraries *Escherichia coli* (3). Although the generation of metagenomic libraries is conceptually simple, the community
sizes of most metagenomes such as those derived from soil and sediment samples and, correspondingly, the large number of clones that is necessary for a significant coverage of the metagenome are great technological challenges (4, 5). Two types of libraries with respect to average insert size can be generated: small-insert libraries in plasmid vectors (less than 10 kb) and large-insert libraries in cosmid and fosmid vectors (up to 40 kb) or BAC vectors (more than 40 kb). The selection of a vector system for library construction depends on the quality of the isolated environmental DNA, the desired average insert size of the library, the copy number required, the host, and the screening strategy that will be used (3, 5). Environmental DNA that is contaminated with humic or matrix substances after purification or DNA sheared during purification is only suitable for the generation of small-insert libraries (3). Small-insert metagenomic libraries are useful for the isolation of single genes or small operons encoding novel biomolecules. To identify complex pathways encoded by large gene clusters or large DNA fragments for the partial genomic characterization of uncultured microorganisms, the generation of large-insert libraries is the appropriate method. Here, we describe one protocol for the construction of small-insert libraries and one for large-insert fosmid libraries. Both methods have been proven to be suitable for cloning of DNA purified from various environmental samples, including soil, ice, and compost (6–8).

2. Materials

2.1. Metagenomic DNA

The construction of metagenomic libraries derived from environmental samples and cloning of functional genes is dependent on the high quality of the extracted DNA, since the enzymatic modifications required during the construction of the libraries are sensitive to contamination by various biotic and abiotic components. High molecular environmental DNA is especially required for the construction of large-insert libraries. To start with library construction 5–10 µg of purified environmental DNA are required.

---

2.2. Generation of Small-Insert Metagenomic Libraries

1. Illustra GenomiPhi V2 DNA Amplification Kit (GE Healthcare, Munich, Germany).
2. Phi29 DNA polymerase (10 U/µL) and reaction buffer (10×) (Fermentas, St. Leon-Rot, Germany).
3. S1 nuclease (100 U/µL) and reaction buffer (5×) (Fermentas, St. Leon-Rot, Germany).
4. DNA polymerase I (10 U/µL) and reaction buffer (10×) (Fermentas, St. Leon-Rot, Germany).
5. Nebulizer (Invitrogen, Karlsruhe, Germany).
6. Shearing buffer: 10 mM Tris–HCl, pH 7.5, 1 mM EDTA, 10% (w/v) glycerol. Store at room temperature.
7. Low melting point (LMP) Biozym Plaque GeneticPure Agarose (Biozym Scientific GmbH, Hessisch Oldendorf, Germany).
8. Tris-acetate-ethylenediamine tetraacetic acid (TAE) buffer (50×): 242 g Tris-base, 57.1 mL acetic acid, 100 mL 0.5 M EDTA, pH 8. Add H₂O to 1 L. Store at room temperature.
9. GELase Agarose Gel-Digesting Preparation (EPICENTRE Biotechnologies, Madison, WI).
10. 3 M sodium acetate, pH 5.
11. 5 M NH₄OAc, pH 7.
12. T4 DNA polymerase (5 U/μL) (Fermentas, St. Leon-Rot, Germany).
13. 10 mM dNTP Mix (Fermentas, St. Leon-Rot, Germany).
14. Klenow Fragment (10 U/μL) (Fermentas, St. Leon-Rot, Germany).
15. Buffer O (10×) (Fermentas, St. Leon-Rot, Germany).
16. SureClean (Bioline, Luckenwalde, Germany).
17. Taq DNA polymerase and reaction buffer with (NH₄)₂SO₄ (10×) (Fermentas, St. Leon-Rot, Germany).
18. 25 mM MgCl₂.
19. 100 mM dATP.
21. Topo® XL PCR Cloning Kit (Invitrogen, Karlsruhe, Germany).
22. Bio-Rad Gene Pulser II (Bio-Rad, Munich, Germany).
23. Kanamycin stock solution: 25 mg/mL H₂O. Filter-sterilize and store at −20°C.
24. Isopropyl-β-d-thiogalactopyranoside (IPTG) stock solution: 24 mg/mL in H₂O. Filter-sterilize, divide into 2 mL aliquots and store at −20°C.
25. 5-Bromo-4-chloro-3-indolyl-β-d-galactoside (X-gal) stock solution: 20 mg/mL N,N'-dimethyl formamide. Filter-sterilize and store at −20°C.
26. Luria-Bertani (LB) agar: 10 g NaCl, 10 g tryptone, 5 g yeast extract per liter, pH 7.2. Add 1.5% agar. Sterilize by autoclaving.
27. LB agar supplemented with 50 μg/mL kanamycin, 48 μg/mL IPTG, and 40 μg/mL X-gal; add 1 mL of kanamycin, IPTG, and X-gal stock solution to 500 mL hot liquid LB agar after autoclaving.
2.3. Generation of Large-Insert Metagenomic Libraries

1. CopyControl™ Fosmid Library Production Kit (EPICENTRE Biotechnologies, Madison, WI). Store according to manufacturer’s instructions.
2. LMP Biozym Plaque GeneticPure Agarose (Biozym Scientific GmbH, Hessisch Oldendorf, Germany).
3. Biometra Rotaphor (Biometra, Goettingen, Germany).
4. Tris–borate–EDTA (TBE) buffer (5×): 54 g Tris-base, 27.5 g boric acid, 20 mL 0.5 M EDTA, pH 8. Add H₂O to 1 L. Store at room temperature.
5. SureClean (Bioline, Luckenwalde, Germany).
6. LB broth supplemented with 10 mM MgSO₄.
7. Chloramphenicol stock solution: 6.25 mg/mL ethanol. Store at −20°C.
8. LB agar supplemented with 12.5 μg/mL chloramphenicol; add 1 mL of chloramphenicol stock solution to 500 mL molten agar.
9. 3 M sodium acetate, pH 7. Store at room temperature.
10. Phage dilution buffer: 10 mM Tris–HCl, pH 8.3, 100 mM NaCl, 10 mM MgCl₂. Store at room temperature.

3. Methods

Library reconstruction comprises several separate steps. For successful cloning of environmental DNA, it is recommended to avoid storage of the isolated DNA for longer periods between the individual steps. If this is not applicable, the purified DNA can be stored 1 to several days at 4°C after each step. Before conducting the end-repair of insert DNA for construction of the plasmid library (see Subheading 3.1.5) or the size fractionation for fosmid library construction (see Subheading 3.2.2), the DNA can be stored at −20°C. However, after end-repair or size fractionation, the DNA should not be stored at −20°C, as freezing and thawing will break the DNA strands. Similarly, unnecessary pipetting of the prepared DNA should be avoided. Where possible, the reagents should be added to the DNA rather than transferring the DNA. When DNA has to be transferred to a fresh microcentrifuge tube, use only large bore or cut off pipette tips to avoid further shearing of the DNA.

After completion of each step, the DNA concentration should be measured to ensure that a sufficiently high DNA concentration is recovered to conduct the remaining steps. Preferably, a large amount of DNA should be used to start as performing the separate procedures will result in the loss of DNA. If less than 5 μg of
environmental DNA are available, for reconstruction of small-insert libraries the amount of DNA can be increased by employing whole genome amplification (WGA). To improve cloning efficiency and to avoid abnormal insert size distribution, hyperbranched structures generated during WGA are resolved as described recently (9) with modifications.

In Subheadings 3.1.1–3.1.3, a protocol for WGA of the environmental DNA and resolving hyperbranched structures is given. However, if a sufficient amount of environmental DNA is available, metagenomic library construction starts with Subheading 3.1.4.

1. Conduct WGA of environmental DNA by using, e.g., the Illustra GenomiPhi V2 DNA Amplification Kit according to manufacturer’s instructions (10).

2. Purify the DNA with SureClean according to manufacturer’s instructions (11). Do not air-dry the pellet for longer than 5–10 min.

3. Resuspend the DNA pellet in 30 μL H₂O (see Note 1).

1. Combine the following ingredients in a sterile microcentrifuge tube: the amplified and purified DNA from Subheading 3.1.1, step 3, 5 μL 10 mM dNTP Mix, 5 μL phi29 buffer (10×), and 1 μL phi29 DNA polymerase (10 U/μL). Add up to a final volume of 50 μL with H₂O. The reaction mix can be scaled up as needed.

2. Incubate at 30°C for 2 h.

3. Inactivate the enzyme at 65°C for 3 min.

4. Purify the DNA with SureClean (see Subheading 3.1.1, steps 2 and 3).

1. Set up the reaction mix as follows: the purified DNA from Subheading 3.1.2, step 4, 10 μL S1 nuclease buffer (5×), 0.5 μL S1 nuclease (100 U/μL). Add up to a final volume of 50 μL with H₂O.

2. Incubate at 37°C for 30 min.

3. Purify the DNA with SureClean (see Subheading 3.1.1, steps 2 and 3).

1. Test the proportion of sheared DNA by running 1–2 μL of the DNA solution on a 0.8% agarose gel. If more than 50% of the DNA fragments display the desired insert size proceed with Subheading 3.1.5.

2. Assemble the nebulizer as indicated by the manufacturer.

3. Add 10 μg environmental DNA to 750 μL of shearing buffer and transfer into the bottom of the nebulizer (see Note 3).
4. Screw on cap of the nebulizer and place on ice to keep the DNA cold.

5. Connect the nebulizer to the compressed gas or air source and shear the DNA by applying 9–10 psi for approximately 10–15 s to obtain DNA fragments that are 3–8 kb in size. Check the DNA on a 0.8% agarose gel to ensure that more than 50% of the DNA fragments display the desired insert size. To vary the size of the DNA fragments either change the applied pressure or vary the time for shearing.

6. Transfer the DNA to two sterile microcentrifuge tubes.

7. Precipitate DNA by adding one of ten volume of 3 M sodium acetate, pH 5, and 2.5 volumes of 96% ethanol. Mix gently. Leave the DNA on ice for 20 min, then centrifuge in a microcentrifuge at top speed for 30 min at 4°C.

8. Discard supernatant. Subsequently, wash the pellet twice with cold 70% ethanol. After the second washing step, carefully invert the tube and allow the pellet to air-dry for 5–10 min.

9. Gently resuspend the DNA in 36 μL H₂O.

3.1.5. End-Repair of Insert DNA

1. Add the following reagents to the resuspended DNA from Subheading 3.1.4, step 9: 5 μL Buffer O (10×), 1 μL 10 mM dNTP Mix, 1 μL T4 DNA polymerase (5 U/μL), and 1 μL DNA polymerase I (10 U/μL). Add H₂O to a final volume of 50 μL (see Note 4).

2. Incubate the reaction mixture for 3 h at room temperature.

3. Inactivate the enzymes for 10 min at 75°C.

3.1.6. Size Fractionation of the Insert DNA

1. Run the blunt-ended DNA on a 1% LMP agarose gel prepared with 1× TAE buffer and a DNA size marker at each of the outside lanes of the gel. Do not include ethidium bromide in the gel.

2. Following electrophoresis, cut off the outer lanes of the gel containing the DNA ladder and stain with ethidium bromide. Visualize the DNA ladder with UV light and mark the position of the desired fragment sizes on both DNA ladders. After removing the gel slices from the UV light, reassemble the gel and cut out a gel slice containing DNA with the desired fragment size.

3. Weigh the gel slice in a tared tube.

4. Exchange the electrophoresis buffer in the gel slice with 1× GELase buffer by adding 3 μL of 1× GELase buffer per mg of gel. Incubate at room temperature for 1 h and subsequently remove the buffer (see Note 5).
5. Melt the LMP gel by incubation at 70°C for 3 min for each 200 mg of gel. If required, continue incubating at 70°C for a few more minutes.

6. Transfer the molten agarose to 45°C and equilibrate 2 min for each 200 mg of gel. Temperatures higher than 45°C will inactivate the GELase enzyme.

7. Add 1 U of GELase enzyme for each 600 mg of gel. Keep the digested agarose solution at 45°C and gently mix. Incubate for at least 1 h.

8. Transfer the reaction mixture to 70°C to inactivate the enzyme for 10 min.

9. Chill tube on ice for 5 min. Centrifuge in a microcentrifuge at top speed for 20 min to pellet any insoluble oligosaccharides. Carefully remove the supernatant and transfer to a new tube.

10. Precipitate the DNA by adding one volume of 5 M NH₄OAc, pH 7, to the molten agarose and 4 volumes of 96% ethanol (see Note 6). In the following, proceed as described in Subheading 3.1.4, steps 7 and 8.

11. Gently resuspend the DNA in 50 μL H₂O.

1. Add the following reagents to the resuspended DNA from Subheading 3.1.6, step 11: 7 μL Taq DNA polymerase buffer (10×), 6 μL 25 mM MgCl₂, 1 μL 100 mM dATP, and 1 μL Taq DNA polymerase (5 U/μL). Add H₂O to a final volume of 70 μL.

2. Incubate at 72°C for 30 min.

3. Purify DNA by using SureClean (see Subheading 3.1.1, step 2).

4. Resuspend DNA pellet in 30 μL H₂O (see Note 7).

1. Prepare a reaction mixture containing the following ingredients: 12.5 μL prepared insert DNA (approximately 500 ng), 1.5 μL Antarctic phosphatase buffer (10×), 1 μL Antarctic phosphatase (5 U/μL).

2. Incubate for 15 min at 37°C.

3. Inactivate the enzyme at 65°C for 5 min.

1. Set up the following cloning reaction in a sterile microcentrifuge tube: 4 μL dephosphorylated insert DNA and 1 μL pCR®-XL-TOPO® vector.

2. Mix gently without pipetting the solution and incubate for 5 min at room temperature.
3. Add 1 μL of the TOPO® Cloning Stop Solution (6×) and mix gently.

4. Briefly centrifuge the tube and place on ice. The ligation mix may be stored for 24 h at 4°C.

5. Add 2 μL of the cloning reaction to one vial of Invitrogen's One Shot® electrocompetent *E. coli* cells and mix gently. Do not pipet.

6. Transfer cells and DNA to a prechilled 0.1 cm electroporation cuvette.

7. Electroporate the cells. We use a Bio-Rad Gene Pulser II with the following settings: 200 Ω, 25 μF, and 2.5 kV.

8. Immediately add 450 μL of room temperature S.O.C. medium (included in the Topo® XL PCR Cloning Kit) and mix well.

9. Transfer the solution to a 15 mL tube and shake horizontally for 1 h at 37°C and 150 rpm.

10. Spread 25 μL of the suspension on LB plates containing 50 μg/mL kanamycin, 48 μg/mL IPTG, and 40 μg/mL X-gal.

11. Incubate the plates overnight at 37°C.

12. Ensure that the plasmid library contains the desired insert size. Randomly pick several *E. coli* clones, grow each overnight in 5 mL LB broth supplemented with 50 μg/mL kanamycin, extract, digest, and analyze plasmid DNA by using standard techniques.

13. Count obtained clones and determine the blue/white ratio, which indicates the amount of insert-containing plasmids.

14. Extract total plasmid DNA by using standard techniques and store at −20°C.

**3.2. Generation of Large-Insert Metagenomic Libraries**

**3.2.1. Preparation of Host Cells**

1. Streak the *E. coli* EPI300-T1® cells on a LB plate. The cells are included in the CopyControl™ Fosmid Library Production Kit. Incubate overnight at 37°C. Seal the plate and store at 4°C.

2. The day before performing the lambda packaging reaction (see Subheading 3.2.6) inoculate 5 mL of LB broth with a single colony of EPI300-T1® cells and incubate overnight at 37°C and 150 rpm.

**3.2.2. Shearing of Metagenomic DNA**

(See Note 8)

1. Randomly shear the environmental DNA by passing it several times through a small bore pipette tip.

2. Load 1–2 μL of the DNA on an agarose gel and check if more than 50% of the DNA fragments display the desired insert size. If not, repeat step 1 until sufficiently sheared DNA is obtained.
1. Size-select the sheared metagenomic DNA as described in Subheading 3.1.6 with the following modifications.

2. Run the DNA on a 1% LMP agarose gel prepared with 1× TBE buffer using pulsed field gel electrophoresis. We use a Biometra Rotaphor with voltage and ramp times as recommended by the manufacturer. Load 100 ng of fosmid control DNA into each of the outside lanes of the gel with the environmental DNA.

3. Heat the GELase buffer (50×) (included in the CopyControl™ Fosmid Library Production Kit) to 45°C and melt the LMP agarose by incubating the tube at 70°C for 10–15 min. Transfer the tube to 45°C.

4. Add the preheated GELase buffer (50×) to 1× final concentration. Per 100 μL of molten agarose add 1 U of GELase and gently mix. Incubate for 1 h. Proceed with steps 8–11 in Subheading 3.1.6.

1. Add the following reagents, which are all included in the CopyControl™ Fosmid Library Production Kit, to the 50 μL resuspended size-fractioned DNA from Subheading 3.2.3, step 4: 8 μL end-repair buffer (10×), 8 μL 2.5 mM dNTP Mix, 8 μL 10 mM ATP, 4 μL end-repair enzyme mix. Add H₂O to a final volume of 80 μL.

2. Incubate at room temperature for 2 h.

3. Inactivate the enzyme mix at 70°C for 10 min.

4. Purify the blunt-ended DNA with SureClean (see Subheading 3.1.1, step 2).

5. Resuspend the DNA in 20–30 μL H₂O (see Note 7).

1. Add the following reagents, which are also included in the CopyControl™ Fosmid Library Production Kit, to the end-repaired insert DNA (Approximately 600 ng): 1 μL Fast-Link ligation buffer (10×), 1 μL 10 mM ATP, 1 μL CopyControl™ pCC1FOS Vector (0.5 μg/μL), 1 μL Fast-Link DNA ligase (2 U/μL). Add H₂O to a final volume of 10 μL.

2. Incubate overnight at 16°C.

3. Add 0.5 μL Fast-Link DNA ligase to the reaction mixture and incubate for another 1.5 h at room temperature.

4. Stop the reaction at 70°C for 10 min.

1. Inoculate 50 mL LB broth supplemented with 10 mM MgSO₄ with 5 mL of an overnight culture of the EPI300-T1® cells (see Subheading 3.2.1, step 2). Incubate the culture at 37°C and 150 rpm until an OD₆₀₀ of 0.8–1.0. Store the cells at 4°C for up to 72 h when required.
2. Thaw one tube of the MaxPlax Lambda Packaging Extracts (included in the CopyControl™ Fosmid Library Production Kit) on ice.

3. Immediately transfer 25 μL of the packaging extract to a new microcentrifuge tube on ice. Store the remaining 25 μL of the MaxPlax Packaging Extract to −70°C until use. Do not expose the packaging extracts to CO₂ sources such as dry ice.

4. Add the ligation reaction to the thawed packaging extracts on ice. Mix the solution without producing air bubbles. Briefly centrifuge the tube.

5. Incubate the reaction mixture for 90 min at 30°C.

6. Thaw the remaining packaging extract from step 3 and add it to the reaction mixture.

7. Incubate for an additional 90 min at 30°C.

8. Add phage dilution buffer to a final volume of 1 mL and mix gently. Add 25 μL chloroform and mix gently. Store at 4°C for up to 2 days.

1. Add 10, 20, 30, 40, and 50 μL of the packaged phage particles individually to 100 μL of the prepared EPI300-T1® cells from Subheading 3.2.6, step 1.

2. Incubate for 45 min at 37°C.

3. Spread the infected EPI300-T1® cells on an LB plate supplemented with 12.5 μg/mL chloramphenicol and incubate overnight at 37°C.

4. Count colonies and mix the remaining packaged phage particles with the host cells in the ratio, which yielded the highest amount of fosmid-containing E. coli clones.

5. Incubate for 45 min at 37°C.

6. Ensure that the fosmid library contains the desired insert size. For this purpose, pick randomly several E. coli clones, grow each in 5 mL LB broth supplemented with 12.5 μg/mL chloramphenicol overnight at 37°C and 150 rpm.

7. To induce a high copy number of the fosmids in the host cells combine 500 μL of the overnight culture from step 6, 5 μL of the CopyControl™ Induction Solution (1,000×), and 4.5 mL LB broth supplemented with 12.5 μg/mL chloramphenicol in a 15 mL tube.

8. Shake the tubes at 37°C horizontally for 5 h vigorously as aeration is critical for induction of a high copy number.

9. Extract, digest, and analyze the fosmid DNA by standard techniques to ensure that the fosmid library contains metagenomic DNA.
10. Store the fosmid library in microtiter plates containing LB broth supplemented with 12.5 μL chloramphenicol at −70°C.

4. Notes

1. If the DNA pellet is difficult to resuspend, add another 20 μL of H₂O and heat to 37°C for 30 min.
2. WGA of DNA results in a hyperbranched structure, which has to be resolved prior to cloning. By incubating the amplified DNA with phi29 polymerase without primers the density of branching junctions is reduced. Resulting 3’ single-stranded overhangs are removed by S1 nuclease treatment. Nicks in the resulting double-stranded DNA are removed by incubation with DNA polymerase I, which can be performed during end-repair of the insert DNA (see Subheading 3.1.5).
3. Shearing of the metagenomic DNA can be done either mechanically using a Nebulizer or a HydroShear® (Zinsser Analytic, Frankfurt, Germany), or by partial restriction endonuclease digestion using, e.g., Bsp143I (Fermentas, St. Leon-Rot, Germany). Note that restriction endonuclease digestion will lead to more biased libraries than mechanical shearing of DNA.
4. If environmental DNA was not subjected to WGA, instead of DNA polymerase I the Klenow fragment should be added to the reaction mixture. DNA polymerase I exhibits not only polymerase and proofreading activity, but also 5’–3’ exonuclease activity, which is important for removal of nicks, which originate from the S1 nuclease treatment described in Subheading 3.1.3.
5. Size fractionation of the insert DNA can also be done by gel extraction via columns, e.g., by using the QIAquick Gel Extraction Kit (Qiagen, Hilden, Germany). Gel purification via columns is less time-consuming but may result in breaking of the prepared DNA strands.
6. The oligosaccharides produced by GELase digestion are more soluble in ethanol in the presence of ammonium. When other salts are used for precipitation, coprecipitation of oligosaccharides may occur.
7. If the DNA concentration is too low after complete resuspension of the DNA pellet, the DNA solution can be concentrated by freeze-drying. We use a Savant SpeedVac Plus SC110A (Thermo Fisher Scientific, Waltham, MA).
8. In some cases, this step can be omitted as DNA extraction from environmental samples frequently results in sufficiently sheared DNA. Therefore, prior to cloning, the molecular
weight of the isolated DNA should be checked by agarose gel electrophoresis.

9. Alternatively, if only a small amount of environmental DNA is available, the size fractionation step can be omitted. Only DNA fragments of approximately 40 kb will be packaged. However, without size fractionation chimeras may form. Size fractionation of the insert DNA is recommended when large contiguous DNA fragments are needed.

References

Metagenomics
Methods and Protocols
Streit, W.; Daniel, R. (Eds.)
2010, XI, 341 p., Hardcover
A product of Humana Press