Published in final edited form as:

Methods Mol Biol. 2015 ; 1327: 1–9. doi:10.1007/978-1-4939-2842-2_1.

Library Construction for Mutation Identification by Whole-Genome Sequencing

Harold E. Smith¹

¹National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD, 20892, USA. smithhe2@niddk.nih.gov.

Abstract

Next-generation sequencing provides a rapid and powerful method for mutation identification. Herein is described a workflow for sample preparation to allow the simultaneous mapping and identification of candidate mutations by whole-genome sequencing in *Caenorhabditis elegans*. The protocol is designed for small numbers of worms to accommodate classes of mutations, such as lethal and sterile alleles, that are difficult to identify by traditional means.

Keywords

Sequence variant detection; Polymorphism mapping; NGS library construction

1 Introduction

For more than a century, researchers have used classical genetics (random mutagenesis and phenotypic screening) in model species to identify the genes involved in biological pathways. The availability of both self and outcross modes of reproduction makes *Caenorhabditis elegans* ideally suited for such genetic screens. Historically, the ability to generate interesting mutations in worms has greatly exceeded the ability to identify causative sequence variants. Dozens or even hundreds of alleles might be recovered in a single mutant screen (e.g., [1, 2]). The analysis of such a large number of strains involves a daunting (but tractable) number of crosses—to discriminate recessive, dominant, and multigenic traits, to remove extraneous mutations through backcrossing, and to define complementation groups. However, identifying even a single gene of interest by the traditional techniques of positional mapping and transgene rescue is considerably more labor-intensive and time-consuming. As a practical consequence, the molecular identity of many alleles remains unknown.

In recent years, the application of next-generation sequencing technology to mutation identification in *C. elegans* has largely eliminated this bottleneck [3–5]. Whole-genome sequencing provides a catalog of all sequence variants within the strain of interest. By crossing the strain into a highly polymorphic genetic background, both positional information and novel mutations can be obtained [6, 7]. Software pipelines specifically designed for *C. elegans* greatly facilitate the identification of candidate genes [8, 9], which can be validated by secondary criteria such as RNAi [10].

The workflow for sample preparation for whole-genome sequencing consists of the following steps: (1) crossing the mutation-bearing strain to a polymorphic strain and picking homozygous F2 progeny; (2) isolating and shearing the genomic DNA (gDNA); and (3) constructing a gDNA library from the sheared sample. The indicated protocols produce sequence-ready libraries, and are designed to accommodate phenotypes (e.g., nonconditional lethality or sterility) that are typically difficult to analyze. The only limitation is that the allele of interest be recessive so that homozygous and heterozygous segregants can be distinguished.

Each step in this workflow is subject to considerable variation. The picking of F2 progeny (step 1) will be determined by the mutant phenotype in question. The gDNA isolation (step 2) is intended for small numbers of worms, but can be used for bulk samples as well. The library construction protocol (step 3) is specific to the Illumina sequencing platform (which, as of January 2015, was the most cost-effective), but alternative platforms are possible. In all cases, the reader is referred to Subheading 4 for relevant parameters to consider when using alternative methods.

2 Materials

2.1 Worm Growth and Mating

1. Dissecting microscope.
2. Worm pick (*see Note 1*).
3. NGM plates: In 2 L flask, mix 3 g NaCl, 17 g agar, 2.5 g peptone, and 975 mL H₂O. Autoclave to sterilize. Cool to 55 °C. Add 1 mL cholesterol (5 mg/mL in ethanol; do not autoclave), 1 mL 1 M CaCl₂, 1 mL 1 M MgSO₄, and 25 mL 1 M KPO₄, pH 6.0 (108.3 g KH₂PO₄, 35.6 g K₂HPO₄, H₂O, 1 L H₂O) (*see Note 2*). Dispense 14 mL into 6 cm petri plates. Cool plates to room temperature. Seed plates with ~50 µL OP50 culture (grown overnight in LB medium @ 37 °C). Incubate seeded plates overnight @ room temperature (*see Note 3*).
4. *E. coli* strain OP50 (*see Note 4*).
5. *C. elegans* strain CB4856 (*see Note 4*).

2.2 gDNA Isolation

1. M9 buffer: Mix 3 g KH₂PO₄, 6 g Na₂HPO₄, 5 g NaCl, and 1 mL 1 M MgSO₄ in 1 L H₂O; autoclave to sterilize.
2. TE buffer: Mix 10 mL 1 M Tris–HCl, pH 8.0 and 2 mL 0.5 M EDTA, pH 8.0 in 988 mL H₂O; autoclave to sterilize.
3. Worm lysis buffer: Mix 100 µL 1 M Tris–HCl (pH 8.0), 20 µL 5 M NaCl, 100 µL 0.5 M EDTA, 125 µL 10 % (v/v) SDS, and 655 µL H₂O.
4. Proteinase K, 10 mg/mL concentration.
5. RNase A, 10 mg/mL concentration.

6. Bioruptor sonicating water bath (Diagenode) or similar.
7. MinElute PCR purification kit (Qiagen) or similar.

2.3 Library Construction

1. NEBNext Ultra DNA library prep kit for Illumina (New England Biolabs).
2. EB: 10 mM Tris-HCl, pH 8.0 or 8.5.
3. AMPure XP beads (Agencourt).
4. Magnetic tube rack.
5. 80 % (v/v) ethanol; prepare fresh immediately before use.
6. Adapter oligonucleotides (see **Note 5**): Prepare @ 100 μ M concentration in TE + 50 mM NaCl. Mix 25 μ L each universal and index adapter oligonucleotide. Incubate in a thermal cycler 2 min @ 95 °C, decrease 1 °C each minute until reaching 25 °C (70 min), then hold @ 4 °C. Dilute to the indicated concentration immediately before use.
7. PCR primers (see **Note 6**): Prepare @ 25 μ M concentration in TE.

3 Methods

3.1 Mating and F2 Selection (See Note 7)

1. Using a dissecting microscope and worm pick, transfer two to three young adult hermaphrodites containing the mutation of interest to a NGM plate seeded with OP50. Pick 10–12 adult males of strain CB4856 to the same plate. Allow to mate overnight @ room temperature (see **Note 8**).
2. Transfer each mated hermaphrodite to an individual fresh-seeded NGM plate. Incubate @ room temperature until the F1 progeny begin to reach the L4/young adult stages (~4 days) (see **Note 9**).
3. Pick 10–12 F1 L4/young adult hermaphrodites to a fresh-seeded NGM plate. Incubate @ room temperature until F2 progeny can be scored for the mutant phenotype (see **Note 10**).

3.2 gDNA Isolation

1. Pick homozygous F2 progeny into a 1.5 mL centrifuge tube containing 500 μ L M9 (see **Note 11**).
2. Vortex briefly (3–5 s), then spin 60 s @ 1300 RCF.
3. Remove most of the M9 by pipette, taking care to avoid the pellet.
4. Resuspend in 500 μ L M9 and repeat the wash at least four times (see **Note 12**).
5. Perform a final wash with 500 μ L TE; spin 1 min at top speed; remove TE, leaving ~100 μ L (see **Note 13**).
6. Add 400 μ L worm lysis buffer to the worm sample; mix briefly.

7. Sonicate with the BioRuptor using the following settings: high power; 30 s on/30 s off; 2 × 15 min sonication time (*see Note 14*).
8. Add 50 µL proteinase K; mix well; incubate 1 h @ 65 °C, vortexing briefly at 10–15 min intervals to maintain suspension (*see Note 15*).
9. Add 20 µL RNase A; incubate 30 min @ 37 °C.
10. Purify sheared gDNA using a MinElute column (Qiagen) per the manufacturer's protocol; the final elution volume is 10 µL (*see Notes 16 and 17*).

3.3 Library Construction

3.3.1 End Repair and A-Tailing

1. Mix in 0.5 mL tube:
10 µL sheared gDNA (2–10 ng)
6.5 µL 10× end repair reaction buffer 3 µL end prep enzyme mix
45.5 µL H₂O
2. Incubate in a thermal cycler:
30 min @ 20 °C
30 min @ 65 °C
Hold @ 4 °C

3.3.2 Adapter Ligation

1. Dilute the adapter to 1.5 µM in EB immediately before use (*see Note 18*).
2. Add the following and mix:
15 µL blunt/TA ligase master mix
2.5 µL adapter
1 µL ligation enhancer
3. Incubate in a thermal cycler:
15 min @ 20 °C
Hold @ 4 °C

3.3.3 AMPure XP Bead Clean-Up (See Note 19)

1. Remove the reaction tube from the thermal cycler; add 83.5 µL beads; mix well by vortexing (*see Note 20*).
2. Incubate the reaction tube 5 min in a tube rack.
3. Place the reaction tube in the magnetic stand; incubate 5 min; remove and discard the supernatant (*see Note 21*).

4. Leave the reaction tube in the magnetic stand; add 200 μ L freshly prepared 80 % (v/v) ethanol; incubate 30 s; carefully remove and discard the supernatant.
5. Repeat **step 4** twice, for a total of three washes.
6. Air-dry the beads in the open tube for 5 min on the magnetic stand.
7. Add 28 μ L EB to elute the DNA; pipette or vortex to resuspend the bead pellet; incubate for 2 min in a tube rack; return to the magnetic stand for 5 min; transfer 23 μ L of supernatant (containing the DNA) to a new 0.5 mL tube.

3.3.4 PCR Amplification (See Note 22)

1. Add the following and mix:
1 μ L primer TruSeq-1 @ 25 μ M
1 μ L primer TruSeq-2 @ 25 μ M
25 μ L NEBNext Q5 Hot Start HiFi PCR Master Mix
2. Incubate in a thermal cycler:
One cycle of:
30 s @ 98 °C
12 cycles of:
10 s @ 98 °C
75 s @ 65 °C
One cycle of:
5 min @ 65 °C
Hold @ 4 °C

3.3.5 Final AMPure XP Bead Clean-Up (See Note 23)—Follow the protocol above with the following differences:

Step 1. Add 50 μ L beads.

Step 5. Repeat **step 4** once, for a total of two washes.

Step 7. Add 25 μ L EB; transfer 20 μ L.

The library is now ready for quantification, using a method suitable for DNA concentrations in the low ng/ μ L range (e.g., dye fluorometry or qPCR). UV absorbance is relatively insensitive and therefore not recommended. Library quality can be assessed using the Agilent Bioanalyzer with a high-sensitivity DNA chip to detect the size distribution. If adapter dimer contamination (a discrete band of 121 bp) is observed, it can be removed by size fractionation via gel isolation or AMPure bead size selection. Libraries with different adapter indices can be pooled, or multiplexed. The recommended minimum sequencing depth for mutation identification is 20-fold genome coverage, or 2×10^9 bp. The amount of

data one obtains is determined by the sequencing capacity of the instrument and the degree of multiplexing; consult your sequencing service provider for guidance. The data can be analyzed using any variety of variant-calling software. For ease of use, we recommend CloudMap [9] on the public Galaxy server (available at <https://usegalaxy.org>) [11–13].

4 Notes

1. Available commercially from www.wormstuff.com. Also, instructions for making worm picks are available online (e.g., http://openwetware.org/wiki/BISC_219/F10:_Gene_Linkage#Making_a_Worm_Pick or <http://www.wormbook.org/wbg/articles/volume-19-number-1/a-better-worm-pick-handle>).
2. NGM medium lacks antibiotics, so sterile technique is essential to prevent contamination. The cholesterol solution is flammable and does not require sterilization. The remaining solutions can be sterilized by autoclave or filtration. Airborne contaminants can be avoided by working in a laminar flow hood.
3. Plates can be stored seeded or unseeded in airtight containers @ 4 °C for several weeks.
4. Strains are available from the Caenorhabditis Genetics Center (<http://www.cgc.cbs.umn.edu>).
5. All oligonucleotide sequences are derived from the TruSeq sample prep kit (©2007–2012 Illumina, Inc. All rights reserved), and generated by annealing the universal adapter to the index adapter. Each index adapter contains a unique 6mer; replace ‘NNNNNN’ with the indicated sequence in Table 1.

Libraries with different indices can be pooled (multiplexed) for sequencing.

Universal adapter (sequence 5′–3′; * = phosphorothioate bond):

AATGATACGGCGACCACCGAGATCTTACACTCTTTCCC
TACACGACGCTCTTCCGATC*T

Index adapter (5′–3′; p = 5′ phosphate; * = phosphorothioate bond):

p ~
GATCGGAAGAGCACACGTCTGAACTCCAGTCACNNNNNNATCTCGTAT
GCCGTCTTCTGCTT*G
6. Primer sequences (5′–3′):

TruSeq-1: AATGATACGGCGACCACCGAG
TruSeq-2: CAAGCAGAAGACGGCATACGAG
7. For strains that do not require mapping, begin with gDNA isolation (Subheading 3.2).
8. The mutation-bearing strain may be homozygous (preferable) or heterozygous (for phenotypes that preclude mating, such as nonconditional sterility or

lethality). Conditional alleles that require maintenance @ 15 °C should be allowed to mate for 24 h. Mutations that impair mating efficiency may require more hermaphrodites to insure success.

9. Successful mating produces equal numbers of male and hermaphrodite outcross progeny, which are first distinguishable at the L4 stage. Use only those plates with successful mating, especially when the starting strain is heterozygous.
10. For heterozygous starting strains, only half of the F1 progeny will contain the mutation; therefore, the number of picked F1s should be doubled. Temperature-sensitive alleles should be incubated at the nonconditional temperature to allow discrimination of homozygous F2 mutants.
11. The number of animals to pick is determined by the stage of development: 200 mid-stage embryos or 50 adult hermaphrodites is the recommended minimum. For larger samples, wash with 10× volumes of TE per volume of packed worms.
12. Washing is critical to remove as much of the OP50 bacteria as possible and minimize the amount of contaminating DNA in the sequencing library. Viable animals contain OP50 in the intestine. To remove, incubate the animals in M9 for 30 min with gentle shaking followed by multiple washes.
13. Samples can be frozen at –80 °C at this point if desired. To continue, thaw in ice bath and proceed with **step 6**.
14. Sonication disrupts the sample and shears the genomic DNA. If using a different instrument, parameters should be optimized to produce sheared gDNA in the 100–300 bp size range. Care should be taken to minimize sample heating and foaming.
15. The suspension should be clear by the end of incubation, indicating complete digestion.
16. Column purification kits from other vendors may be substituted. Be sure that the protocol is compatible with samples containing ~1 % SDS.
17. The amount of sheared gDNA should be 2–10 ng, which can be quantified by dye fluorometry (e.g., Qubit) or qPCR. Remember to increase the elution volume accordingly, leaving a final volume of 10 µL after quantification.
18. Excess adapters can ligate and produce dimers, which are effective competitors for PCR amplification and sequencing. At the lower limit of input (2 ng), the adapters should be diluted fivefold to 0.3 µM to minimize dimer formation.
19. Resuspend beads by vortexing immediately before pipetting; perform all incubations at room temperature.
20. Depending upon your magnetic stand, it may be necessary to transfer the sample to a tube of different size (e.g., 1.5 mL).

21. The solution will clear as beads adhere to the side of the tube adjacent to the magnet. When removing the supernatant, pipette slowly and carefully to avoid the bead pellet and bound gDNA.
22. At the lower limit of input gDNA (2 ng), the number of amplification cycles should be increased from 12 to 15.
23. Alternatively, a MinElute column can be used for the final clean-up.

Acknowledgments

I would like to thank Sevinc Ercan for sharing the small-scale gDNA isolation protocol. This work was supported by the Intramural Research Program of the National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases, and is subject to the NIH Public Access Policy.

References

1. Brenner S (1974) The genetics of *Caenorhabditis elegans*. *Genetics* 77:71–94 [PubMed: 4366476]
2. Hirsh D, Vanderslice R (1976) Temperature-sensitive developmental mutants of *Caenorhabditis elegans*. *Dev Biol* 49:220–235 [PubMed: 943345]
3. Sarin S, Prabhu S, O'Meara MM et al. (2008) *Caenorhabditis elegans* mutant allele identification by whole-genome sequencing. *Nat Methods* 5:865–867 [PubMed: 18677319]
4. Sarin S, Bertrand V, Bigelow H et al. (2010) Analysis of multiple ethyl methanesulfonate-mutagenized *Caenorhabditis elegans* strains by whole-genome sequencing. *Genetics* 185: 417–430 [PubMed: 20439776]
5. Zuryn S, Le Gras S, Jamet K et al. (2010) A strategy for direct mapping and identification of mutations by whole-genome sequencing. *Genetics* 186:427–430 [PubMed: 20610404]
6. Wicks SR, Yeh RT, Gish WR et al. (2001) Rapid gene mapping in *Caenorhabditis elegans* using a high density polymorphism map. *Nat Genet* 28:160–164 [PubMed: 11381264]
7. Doitsidou M, Poole RJ, Sarin S et al. (2010) *C. elegans* mutant identification with a one-step whole-genome-sequencing and SNP mapping strategy. *PLoS One* 5:e15435 [PubMed: 21079745]
8. Bigelow H, Doitsidou M, Sarin S et al. (2009) MAQGene: software to facilitate *C. elegans* mutant genome sequence analysis. *Nat Methods* 6:549 [PubMed: 19620971]
9. Minevich G, Park DS, Blankenberg D et al. (2012) CloudMap: a cloud-based pipeline for analysis of mutant genome sequences. *Genetics* 192:1249–1269 [PubMed: 23051646]
10. Fire A, Xu S, Montgomery MK et al. (1998) Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* 391:806–811 [PubMed: 9486653]
11. Giardine B, Riemer C, Hardison RC et al. (2005) Galaxy: a platform for interactive large-scale genome analysis. *Genome Res* 15:1451–1455 [PubMed: 16169926]
12. Blankenberg D, von Kuster G, Coraor N et al. (2010) Galaxy: a web-based genome analysis tool for experimentalists. *Curr Prot Mol Biol Unit* 19.10.1–21
13. Goecks J, Netrutenko A, Taylor J et al. (2010) Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol* 11:R86 [PubMed: 20738864]

Table 1

Index sequences

1 ATCACG	5 ACAGTG	9 GATCAG
2 CGATGT	6 GCCAAT	10 TAGCTT
3 TTAGGC	7 CAGATC	11 GGCTAC
4 TGACCA	8 ACTTGA	12 CTTGTA