Published in final edited form as:

Curr Protoc Mol Biol. ; 117: 7.29.1–7.29.12. doi:10.1002/cpmb.27.

Next-generation sequencing for identification of EMS-induced mutations in *Caenorhabditis elegans*

Nicolas J. Lehrbach^{1,2}, Fei Ji^{1,2}, and Ruslan Sadreyev^{1,3}

¹Department of Molecular Biology, Massachusetts General Hospital, Boston, MA USA

²Department of Genetics, Harvard Medical School, Boston, MA USA

³Department of Pathology, Massachusetts General Hospital & Harvard Medical School, Boston, MA USA

Abstract

Forward genetic analysis using chemical mutagenesis in model organisms is a powerful tool for investigation of molecular mechanisms in biological systems. In the nematode *Caenorhabditis elegans* mutagenesis screens using ethyl methanesulfonate (EMS) have led to important insights into genetic control of animal development and physiology. A major bottleneck to this approach is identification of the causative mutation underlying a phenotype of interest. In the past, this has required time-consuming genetic mapping experiments. More recently, next-generation sequencing technologies have allowed development of new methods for rapid mapping and identification of EMS-induced lesions. In this unit we describe a protocol to map and identify EMS-induced mutations in *C. elegans*.

Keywords

Next-generation sequencing; *C. elegans*; EMS mutagenesis

Basic Protocol

This unit describes a protocol to identify a recessive mutation underlying a phenotype of interest in *C. elegans*. In this approach, a mutant that has been identified from a generalized genome-wide random chemical mutagen based mutagenesis screen is outcrossed to the parental wild-type strain, and animals displaying the same mutant phenotype are isolated in the F2 generation. Thus many segments of the mutagenized genome randomly assort in this backcross, but the region bearing the recessive mutation that causes the phenotype is necessarily homozygous in any one animal with the mutant phenotype. A pooled DNA sample derived from these mutant F2 animals is extracted and analyzed using next-generation sequencing technology. By aligning next generation sequencing data to the reference genome, EMS-induced variants in the mutant F2 animals can be identified, and their putative effect on gene function can be predicted based on genome annotations. Standard EMS mutagenesis conditions used in *C. elegans* typically generate about 500

background mutations in addition to the mutation that causes the mutant phenotype. Mutations not linked (nearby on the same chromosome) to the causative allele are unlikely to be inherited in all mutant F2s after outcrossing, so this analysis should generate a short list of chromosomally linked candidate causative alleles for the mutant strain. Finally, we provide a method to prioritize candidate causative alleles, exploiting the fact that large-scale screens often isolate multiple independent alleles of genes that are disrupted to give the desired phenotype. Similar methods for mapping and identifying *C. elegans* mutants by whole genome sequencing have been described elsewhere, and should be consulted in combination with this protocol (Doitsidou et al., 2010; Hu, 2014; Jaramillo-Lambert et al., 2015; Minevich et al., 2012; Schneeberger, 2014; Zuryn and Jarriault, 2014).

Materials

5 cm Nematode Growth Medium (NGM) agar plates seeded with *E. coli* strain OP50, see (Stiernagle, 2006).

Puregene Cell and Tissue Kit (Qiagen)

Isopropanol

Ethanol

NEBNext DNA library prep kit for Illumina (NEB)

NEBNext Multiplex oligos for Illumina (NEB)

Zymo DNA Clean and Concentrator Kit (Zymo Research)

Covaris S2 Sonicator

microTUBE AFA Fiber Crimp-Cap 6×16mm (Covaris)

Covaris microTUBE holder S 5001144

AMPure XP Beads (Agencourt)

Magnetic stand

55°C water bath

37°C incubator or water bath

NanoDrop

Qbit Fluorometer (Life Technologies)

Qubit dsDNA HS Assay kit (Life Technologies)

Benchtop centrifuge

Thermal cycler

Heat block

Collection of mutant DNA samples

1. After outcrossing a mutant of interest to a non-mutagenized parental strain, pick at least 10, and ideally ~50 single mutant F2 animals to fresh 5 cm NGM plates seeded with *E. coli* OP50. Allow the F2 animals to self fertilize and re-check progeny for the desired phenotype. Discard plates that show any sign of non-mutant animals.
2. To harvest animals for genomic DNA extraction, wait until the plates have “starved” (all of the *E. coli* OP50 has been consumed), ideally collecting animals within 1-2 days after the culture has exhausted its food supply. At least three starved 5 cm plates are required to collect sufficient material for library preparation by this protocol.
3. Wash animals from plates with distilled water into single a 15 mL Falcon tube, creating a pooled sample containing material from self progeny of all the mutant F2s. Centrifuge at ~400 g for 45 sec at room temperature to pellet the animals, and aspirate the supernatant.
4. Wash the worm pellet twice by addition of fresh distilled water. After the final aspiration remove as much of the water as possible. Note the approximate volume of the pellet, then store at -80°C.

Genomic DNA extraction

Genomic DNA extractions are performed with the Puregene Cell and Tissue Kit (Qiagen). The following is adapted from the manufacturer's instructions, and can be used for extraction of genomic DNA from a ~100 µl worm pellet. For larger pellets the protocol can be scaled up as follows: ~150-400 µl, double all volumes; >400 µl, triple all volumes.

5. Thaw worm pellet on ice and add 1 ml Cell Lysis Solution, and 5 µl Proteinase K. Mix well, and place tubes in a water bath at 55-65°C and incubate for at least 3 hours, mixing every ~30 min to ensure full digestion of the sample.
6. Allow sample to cool to room temperature. Add 5 µl RNase A Solution, mix well and incubate at 37°C with shaking for 1-2 hours.
7. Incubate sample on ice for 3 min.
8. Add 330 µl Protein Precipitation solution and vortex sample at high speed for 20 sec. Incubate sample on ice for a further 5 min.
9. Centrifuge sample for 10 min at 2000 g at room temperature, and decant the supernatant to a fresh 15 ml tube.
10. Add 1 ml isopropanol and mix sample by inverting several times. If DNA does not begin to precipitate immediately, the sample can be stored at -20°C for 1 hour or overnight to increase yield. Storage overnight at 20°C is recommended for samples with small amount of starting material.

11. Centrifuge 10 min at 2000 g at room temperature to pellet the DNA. Discard the supernatant. A white DNA pellet should be visible.
12. Add 1 ml 70% Ethanol and invert several times to thoroughly wash the DNA pellet.
13. Centrifuge 5 min at 2000 g at room temperature. Discard the supernatant.
14. Air-dry the DNA pellet for 5-10 min. A Kimwipe can be used to carefully absorb excess liquid if necessary.
15. Add 100 µl DNA Hydration Solution to the DNA pellet. This volume may be adjusted depending on the size of the pellet.
16. Incubate at 65°C for 1 hour, or overnight at 4°C, to dissolve the DNA.
17. Transfer DNA solution to a fresh 1.5 ml eppendorf tube. Determine DNA concentration with a NanoDrop microvolume spectrophotometer.

Shearing of genomic DNA

18. Fill the tank with deionized water and turn on the Covaris S2 sonicator, and water cooler. Ensure the water cooler is set to 4°C. Degas for at least 30 min before beginning sonication.
19. Dilute 1.5 µg of genomic DNA to a final volume of 55 µl using TE.
20. Transfer the 55 µl DNA sample to a Covaris 6×16 mm microTUBE with AFA fibre and crimp cap. Remove any bubbles by pipetting up and down.
21. Mount the sample tube into a microtube holder S 500114, and place the holder in the water tank.
22. Sonicate the sample with the following settings:
 - Intensity: 5
 - Duty Cycle: 10%
 - Cycles/burst: 200
 - Time (sec): 120
 - Number of cycles: 2.
 - Power mode: Frequency sweeping.
 - Degassing mode: Continuous.
23. Remove the holder and retrieve the tube. Transfer the sonicated DNA solution to a fresh 1.5 ml eppendorf. Store the sonicated DNA at -20°C.

Library preparation and validation

The following protocol is for use with the NEBNext DNA library prep kit for Illumina. Alternative reagents can be used. The oligos for adaptor ligations and PCR enrichment of ligated fragments can be purchased from NEB, or from a separate oligo supplier.

End repair

24. To 50 µl of sheared DNA add the following:

35 µl sterile water

10 µl End repair reaction buffer (10×)

5 µl End repair enzyme mix

25. Incubate at 20°C for 30 minutes.

26. Purify the DNA fragments with Zymo DNA clean and concentrator kit, and elute with 42 µl of sterile water.

A tail end repaired DNA

27. To the end-repaired DNA add the following:

5 µl dA-tailing reaction buffer (10×)

3 µl Klenow Fragment (3' → 5' exo-)

28. Incubate at 37°C for 30 minutes.

29. Purify the A-tailed DNA fragments with Zymo DNA clean and concentrator kit, and elute with 25 µl of sterile water

Ligate adaptors

30. To the A-tailed DNA fragments, add the following:

10 µl Quick Ligation reaction buffer (5×)

10 µl Adaptor

5 µl Quick T4 DNA ligase

31. Incubate at 20°C for 15 min.

32. Add 3 µl of USER enzyme and mix by pipetting up and down

33. Incubate at 37°C for 15 min.

Size selection using AMPure XP beads

34. Add 50 µl sterile water to the ligated DNA fragments to increase the total volume of the sample to 100 µl.

35. Add 70 µl resuspended AMPure XP beads to the ligated DNA fragments. Mix well by with a vortex mixer or by pipetting up and down at least 10 times. The volume of AMPure

XP beads added at this step can be adjusted to change the size of the DNA fragments selected if necessary (an increased volume of AMPure XP beads results in selection of larger DNA fragments).

36. Incubate for 5 minutes at room temperature.

37. Place the tubes on a magnetic stand. Allow the sample to stand until the solution is clear (at least 5 minutes). Carefully transfer the supernatant to a fresh tube, avoiding transfer of any beads. Discard the beads (do not discard supernatant).

38. Add 20 μ l resuspended AMPure XP beads to the supernatant. Mix well by with a vortex mixer or by pipetting up and down at least 10 times.

39. Incubate for 5 minutes at room temperature.

40. Place the tubes on a magnetic stand. Allow the sample to stand until the solution is clear (at least 5 minutes). Carefully remove the supernatant, avoiding removal of any beads. Discard the supernatant (do not discard beads).

41. Keeping the tube on the magnetic stand, add 200 μ l freshly prepared 80% ethanol to the beads, incubate for 30 sec at room temperature and then carefully remove and discard the supernatant.

42. Repeat step 41 once.

43. Keeping the tube on the magnetic stand, air dry the bead pellet for 10 min.

44. To elute the DNA fragments, add 25 μ l sterile water to the beads. Mix well on a vortex mixer or by pipetting up and down at least 10 times. Return the sample to the magnetic stand and allow the sample to stand until the solution is clear (at least 1 min).

45. Carefully transfer 20 μ l of the supernatant to a fresh PCR tube.

PCR enrichment of adaptor ligated DNA fragments

46. To the size-selected ligated DNA fragments add the following:

2.5 μ l Universal PCR primer (25 μ M)

2.5 μ l Index PCR primer

25 μ l NEBNext High fidelity 2 $\times$ PCR master mix.

47. Amplify using the following PCR program:

1 cycle: 30 sec 98°C

7 cycles: 10 sec 98°C

30 sec 65°C

30 sec 72°C

1 cycle: 5 min 72°C

Purify amplified fragments with AMPure XP beads

48. Add 50 µl resuspended AMPure XP beads to the completed PCR reaction. Mix well by with a vortex mixer or by pipetting up and down at least 10 times.
49. Incubate for 5 minutes at room temperature.
50. Place the tubes on a magnetic stand. Allow the sample to stand until the beads have separated from the solution, and the solution is clear (at least 5 minutes). Carefully remove the supernatant, avoiding removal of any beads. Discard the supernatant. Do not discard beads!
51. Keeping the tube on the magnetic stand, add 200 µl freshly prepared 80% ethanol to the beads, incubate for 30 sec at and then carefully remove and discard the supernatant.
52. Repeat step 52 once.
53. Keeping the tube on the magnetic stand, air dry the bead pellet for 10 min.
54. To elute the DNA fragments, add 30 µl TE to the beads. Mix well on a vortex mixer or by pipetting up and down at least 10 times. Return the sample to the magnetic stand and allow the sample to stand until the solution is clear (at least 1 min).
55. Carefully transfer 25 µl of the supernatant to a fresh tube.

Library validation and quantification

In this protocol libraries are validated and quantified by gel electrophoresis and Qubit assay before pooling. More accurate estimation of size and concentration of each library can be performed using the Agilent Bioanalyzer and a quantitative PCR assay, but is not essential at this step.

56. Analyze 2 µl of the purified library by electrophoresis on a 2.5% agarose gel. The library should run as a tight smear with most fragments 300-350 bp in size. Estimate the average size of each library by comparison to an appropriate DNA ladder.
57. Quantify the DNA in 1 µl of the purified library using the Qubit dsDNA HS assay kit.

Pooling libraries

58. Using the estimated average library size, calculate the approximate molar mass of the library with the following formula:

$$\text{Library molar mass (g/mol)} = \text{Library average size (bp)} \times 660 (\text{g bp}^{-1} \text{mol}^{-1})$$

59. Calculate the molar concentration of each sample using the following formula:

Library concentration(nM)=Library concentration(pg/ml)/Library molar mass g/mol)

60. Pool libraries for multiplexed sequencing. Mix an appropriate volume of each sample and dilute with TE to generate a solution containing an equimolar concentration of each library, a total DNA concentration of 20 nM, in a total volume of 20-100 µl.

The number of libraries to be pooled depends on the sequencing technology being used, and the desired depth of sequence coverage. Take care to ensure that each library in the pool has been generated using a different index primer during PCR enrichment to allow demultiplexing.

Accurate quantification of the pooled sample with Agilent Bioanalyzer and quantitative PCR assays is recommended before sequencing. Check whether these are performed as part of the sequencing service you are using.

61. Submit pooled sample for high throughput sequencing.

Analyze next-generation sequencing data

In steps 62-70 we use SnPMAP, our computational pipeline for bulk segregant analysis. This pipeline is run from the command line and requires access to sufficient computational resources to run the analysis steps. The CloudMap pipeline, which is run through a website interface using the Public Galaxy Server cloud computing resource may be preferable in some circumstances (Afgan et al., 2016; Minevich et al., 2012).

62. Download and install the required tools (Table 1); download *C. elegans* reference genome WBcel235 from http://www.ncbi.nlm.nih.gov/assembly/GCF_000002985.6

63. Create the alignment index of the WBcel235 reference genome using the bwa index utility as described: <http://bio-bwa.sourceforge.net/bwa.shtml>.

64. Edit the tab-delimited configuration file provided with the bulk segregant analysis (snpmmap) program (see Table 1). In this file, column 1 contains the names of five required programs (installed in step 62). In column 2, supply the complete location paths to these tools on your machine.

65. (OPTIONAL) Test the installation with sample test data included in the download package. These data include a small input mutant *C. elegans* strain whole-genome sequencing dataset (Doitsidou et al., PLoS One 2010) and the expected output that can be used to confirm that the pipeline works correctly.

Run the bulk segregation analysis pipeline

Bulk segregation analysis runs in two steps. The first phase reports all protein coding missense polymorphisms between input FASTQ reads and the reference genome. Then it creates a histogram of distribution of identical variants across different samples. A SNP identified in multiple samples is more likely to represent the background polymorphism

between the experimental strain and reference strain, rather than a mutation induced by EMS. In the second phase, all such background SNPs are identified and filtered out, and candidate genes that have different SNPs detected in multiple samples are reported.

66. Use the following command for running the initial phase:

```
snpmap.pl -f align -in < read 1 FASTQ file > <read 2 FASTQ file> -t <reference genome  
FASTA file in directory containing bwa index> -config <snpmap config file>
```

Run the above command; make sure that the command is properly formatted and all file paths are valid. The execution process may take several hours, depending on the amount of input data and available computing power. This is an example of the command line to align the sample FASTQ file *sample/test.fastq* against *C. elegans* reference genome WS241 (CE10):

```
snpmap.pl -f align -in sample/test.fastq -t sample/CE10.fa -config sample/snpmap.config 2>  
snpmap.initial.err &
```

67. Once the process is complete, the run folder should contain several.snp.txt files. Each contains SNP data for one input FASTQ. The columns of this file are described in Table 2.

68. Use the following command for calculating frequencies of occurrence of candidate genes in the list of SNPs detected in multiple samples:

```
snpmap.pl -f freq -d <directory containing.snp.txt files from the previous step>
```

Run the above command; make sure that the command is properly formatted and all file paths are valid. This is an example command line, where *sample/snp* is the directory containing sample.snp.txt files from step 66:

```
snpmap.pl -f freq -d sample/snp
```

The final output table containing frequencies of SNP occurrence in candidate genes is written in gene.freq.txt in the same folder. In addition to the columns in table 2, gene.freq.txt contains a column of SNP frequency in mutant strains.

69. Using the output table with frequencies of SNP occurrence in candidate genes, filter out candidate genes mutated in fewer than 3 samples and sort the remaining candidate genes based on mutation frequency using the following command:

```
awk '{FS="\t"} $2>5{print}' gene.freq.txt | sort -k2nr
```

70. Carefully examine each candidate variant and interpret the data in the context of the particular biological question. Subsequent laboratory validation of potential SNPs is essential.

Commentary

Background information

Forward genetic analysis in *C. elegans* has provided important insights into mechanisms of animal development and physiology. Many of these insights were gained by molecular cloning of mutations isolated in EMS mutagenesis screens. Before the advent of next generation sequencing technologies, identification of EMS-induced mutant alleles required extensive mapping followed by testing candidate loci. Mapping involved crossing the mutant of interest to a mapping strain containing markers of known genetic linkage (Fay, 2006). These could either be classical alleles with easily scored visual phenotypes, or single nucleotide polymorphisms (Davis et al., 2005). More recently a collection of precisely mapped single copy transgenes has been generated that can also be used for this purpose (Frøkjær-Jensen et al., 2014). Although effective, this approach can be time consuming and often requires extensive strain construction to generate appropriate mapping strains for the phenotype of interest. Subsequent to mapping a mutation to a satisfactorily small interval, further experiments are required to identify the locus affected by the EMS-induced lesion, usually by individually testing candidate loci. Candidates may be prioritized based on phenotypes induced by RNAi or other alleles, or based on the function of orthologues in other species, if known. Testing of candidates can be performed by complementation, rescue, and sequencing. These conventional techniques of identifying EMS-induced mutations are time consuming, laborious, and often boring. The advent of next generation sequencing technology means that mapping and identification of candidate causative EMS-induced lesions can be combined in a single experiment that can be performed much more quickly than using conventional genetic mapping approaches. What in the past was at best a few months and usually almost a year of deferred gratification to learn the molecular identity of a single mutated gene, the geneticist can now learn the identity of 20 or more different mutants within weeks of performing the genetic screen from which they emerged.

Critical Parameters and Troubleshooting

Robustness of the phenotypic assay used to identify mutant animals—An important factor in the success of this protocol is robust phenotypic identification of mutant animals after outcrossing of the EMS-induced mutant. The presence of any non-mutant animals in the starting material for DNA extraction can result in failure to identify the causative allele. As such it may be important to re-test mutant F2 stocks (in the F3 or subsequent generations) depending on the reliability of the phenotypic assay used. In cases where the phenotypic assay used to identify mutant animals is labor intensive it may be necessary to reduce the number of F2s pooled to generate the DNA sample for analysis. In some cases, direct sequencing of a non-outcrossed mutant strain may be performed. This saves time spent outcrossing the mutant, and eliminates the risk of including non-mutant DNA, but will not generate any chromosomal linkage information. Direct sequencing of a mutant strain without pooling outcrossed F2 animals has been used successfully to identify mutations where independent mapping information is available (Chu et al., 2014; Sarin et al., 2008). This approach may also prove useful even in the absence of mapping data if a large number of mutants from a single screen are under analysis, allowing the screen to

approach saturation. In this case, causative alleles may be identified when multiple phenotypically similar mutant strains contain independent lesions in the same gene.

Outcrossing and mutant F2 population size—Assuming that a robust and rapid phenotypic assay is used to identify mutants, picking a large number of mutant F2s after outcross is desirable. Typical EMS mutagenesis protocols used in *C. elegans* introduce approximately 400 single nucleotide changes per genome, of which on average 22% potentially disrupt gene function by altering coding potential or mRNA splicing (Thompson et al., 2013). Thus identifying candidate causative alleles without linkage information can be difficult. Pooling DNA from 10 mutant F2 animals is sufficient to generate a list of candidate mutations strongly enriched for the chromosome harboring the causative allele, allowing chromosomal linkage to be inferred. However, the set of chromosomally linked candidate variants is likely to be large (>10) and cover a large interval on the linked chromosome. Pooling DNA from 50 or more mutant F2 animals is therefore desirable. This will generate a shorter list of linked candidate variants, over a smaller chromosomal interval, and in some cases will uniquely identify the causative allele. Picking large numbers of mutant F2 animals after outcrossing is the most time-consuming step of this protocol, so it may be useful to make a trade-off between number of mutants analyzed and the number of F2 animals picked per mutant, particularly as analyzing a large number of mutants aids identification of causative alleles by recovery of independent alleles of the same genes.

Library quality and sequencing coverage—Constructing libraries with a high proportion of fragments that align to the *C. elegans* genome is required to make efficient use of high throughput sequencing approaches. Collecting animals from starved plates avoids potential contamination with bacterial DNA that will reduce the proportion of reads that align to the *C. elegans* genome. If it is difficult to obtain starved plates, for example if a mutant has a strong growth defect, animals can be washed from unstarved plates, washed extensively and then incubated rotating in sterile water for 3-5 hours to allow digestion of remaining bacteria.

Genomic DNA inserts in the library must be at least twice as long as the read length (assuming paired end reads, see below) to ensure that sequencing capacity is not wasted by sequencing part of each insert twice. If library preparation generates inserts of smaller than the anticipated size, it may be necessary to optimize the DNA shearing protocol and/or volume of AMPure XP beads used in the size selection step.

We have found that 5-10× coverage is largely sufficient for detection of SNPs and deletions. A recent large scale study using similar method estimated a false positive rate of less than 1% and a false negative rate of 7% in SNP detection from 15× coverage Illumina sequencing of EMS mutagenized *C. elegans* strains (Thompson et al., 2013). As such it is worth bearing in mind that in a minority of cases the causative allele will not be detected by this protocol; we find it more cost effective to repeat sequencing of ‘difficult’ mutants, than to initially sequence all mutants at greater depth. In a routine sequencing run we analyze 24 multiplexed libraries (representing 24 different EMS-induced mutations) in a single lane of 50 bp paired end sequencing on an Illumina HiSeq instrument. This yields on average 8 million paired end reads per sample and approximately 8× coverage. The number of libraries

to be pooled can be varied depending on the sequencing technology to be used, and the desired depth of sequencing coverage.

Computational analysis of sequencing data—Here we describe SnPMAP, our newly developed computational pipeline for bulk segregant analysis that is dependent on BWA, GATK, Picard, and SNPeff. Using this pipeline requires some familiarity with the command line and access to sufficient computational resources to run the analysis steps. In cases where access to either or both of these may be limited, the CloudMap pipeline, which is run through a website interface using the Public Galaxy Server cloud computing resource may be preferable (Afgan et al., 2016; Minevich et al., 2012).

Subtraction of background mutations—Due to genetic drift and mutations induced during strain construction (for example transgenesis or previous mutagenesis screens) most *C. elegans* strains used in EMS mutagenesis screens will contain a large number of polymorphisms when compared to the reference genome. These variants are not candidates to underlie the phenotype of interest, but will be detected when reads from the sample are aligned to the reference genome. A large number of background variants hampers analysis of candidate causative alleles. To avoid this problem the parental strain used in the mutagenesis screen may be sequenced to identify the variants present in this strain before mutagenesis. Alternatively, any variant found in multiple independent mutants can be subtracted from the list of candidate causative alleles. This approach is effective and saves sequencing capacity for mutant samples, and is incorporated into this protocol. It is important to note that in some cases this approach might mask genuine causative alleles, if multiple independent mutants with an identical causative mutation are isolated. This can occur in highly saturated screens, or in *en masse* screens in which isolation of siblings bearing the same causative allele is possible. To mitigate this problem, it may be necessary to apply a less stringent filter when removing non-unique variants.

Confirmation of candidate causative alleles—Although the list of candidate lesions identified by this protocol is often short (usually fewer than 10 candidates, when ~50 F2s are pooled for sequencing), we recommend further analysis to confirm the causative allele, especially if a mutant will be used for detailed phenotypic and genetic analysis in the future. A number of approaches may be used to provide additional evidence that a particular lesion is causative. Recovery of multiple mutant alleles of the same gene is highly suggestive that these mutations are causative, and can be confirmed by complementation testing. Otherwise, obtaining a phenocopy using RNAi and/or independently derived alleles can serve to support the hypothesis that a particular lesion is causative. Transformation rescue experiments can provide definitive proof that a lesion in a given gene is causative.

Anticipated results

This protocol is expected to generate a short list of candidate causative alleles for each mutant strain analyzed, the genomic locations of which should be sufficient to infer linkage of the causative allele to a chromosome, or a small interval within a chromosome. This protocol will also generate a list of genes that are disrupted in multiple mutant strains that

aids identification of causative alleles, especially in cases where a large number of mutants is under analysis.

Time considerations

This experimental procedure spans roughly 2 months. Approximately 2 weeks are required to outcross and grow sufficient numbers of animals for genomic DNA preparation. This may take longer in cases where identification of the desired phenotype requires a more time consuming assay, or if mutant animals are slow growing. 1-2 weeks are required for DNA extraction and library preparation. Sequencing may take 1 week to ~1 month, depending on sequencer availability. The computational analysis takes 1-2 weeks.

Acknowledgments

Human Frontiers Science Foundation Long-Term Fellowship to Nicolas Lehrbach. NL is also supported by an NIH grant to Gary Ruvkun (R01 AG016636).

References

- Afgan E, van den Baker D, Beek M, Blankenberg D, Bouvier D, Chilton J, Clements D, Coraor N, Eberhard C, et al. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. *Nucleic Acids Res* gkw343. 2016
- Chu JSC, Chua SY, Wong K, Davison A, Johnsen R, Baillie DL, Rose AM. High-throughput capturing and characterization of mutations in essential genes of *Caenorhabditis elegans*. *BMC Genomics*. 2014; 15:361. [PubMed: 24884423]
- Davis M, Hammarlund M, Harrach T, Hullett P, Olsen S, Jorgensen E. Rapid single nucleotide polymorphism mapping in *C. elegans*. *BMC Genomics*. 2005; 6:118. [PubMed: 16156901]
- Doitsidou M, Poole RJ, Sarin S, Bigelow H, Hobert O. *C. elegans* mutant identification with a one-step whole-genome-sequencing and SNP mapping strategy. *PLoS ONE*. 2010; 5:e15435. [PubMed: 21079745]
- Fay, D. Genetic mapping and manipulation: Chapter 1-Introduction and basics. *WormBook*. 2006. http://www.wormbook.org/chapters/www_introandbasics/introandbasics.html
- Frøkjær-Jensen C, Davis MW, Sarov M, Taylor J, Flibotte S, LaBella M, Pozniakovsky A, Moerman DG, Jorgensen EM. Random and targeted transgene insertion in *Caenorhabditis elegans* using a modified *Mos1* transposon. *Nat Methods*. 2014; 11:529–534. [PubMed: 24820376]
- Hu PJ. Whole genome sequencing and the transformation of *C. elegans* forward genetics. *Methods*. 2014; 68:437–440. [PubMed: 24874788]
- Jaramillo-Lambert A, Fuchsman AS, Fabritius AS, Smith HE, Golden A. Rapid and Efficient Identification of *Caenorhabditis elegans* Legacy Mutations Using Hawaiian SNP-Based Mapping and Whole-Genome Sequencing. *G3 (Bethesda)*. 2015; 5:1007–1019. [PubMed: 25740937]
- Minevich G, Park DS, Blankenberg D, Poole RJ, Hobert O. CloudMap: A Cloud-Based Pipeline for Analysis of Mutant Genome Sequences. *Genetics*. 2012; 192:1249–1269. [PubMed: 23051646]
- Sarin S, Prabhu S, O'Meara MM, Pe'er I, Hobert O. *Caenorhabditis elegans* mutant allele identification by whole-genome sequencing. *Nat Methods*. 2008; 5:865–867. [PubMed: 18677319]
- Schneeberger K. Using next-generation sequencing to isolate mutant genes from forward genetic screens. *Nat Rev Genet*. 2014; 15:662–676. [PubMed: 25139187]
- Stiernagle, T. Maintenance of *C. elegans*. *WormBook*. 2006. http://www.wormbook.org/chapters/www_strainmaintain/strainmaintain.html
- Thompson O, Edgley M, Strasbourger P, Flibotte S, Ewing B, Adair R, Au V, Chaudhry I, Fernando L, Hutter H, et al. The million mutation project: A new approach to genetics in *Caenorhabditis elegans*. *Genome Res*. 2013; 23:1749–1762. [PubMed: 23800452]

Zuryn S, Jarriault S. Deep sequencing strategies for mapping and identifying mutations from genetic screens. *Worm*. 2014; 2:e25081.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 1
Required Bioinformatics Tools

All tools are publicly available for download and should be installed prior to the analysis. SnpMap is our newly developed computational pipeline for bulk segregant analysis that is dependent on BWA, GATK, Picard, and SNPeff.

Program	Source	Purpose
BWA	http://bio-bwa.sourceforge.net	Read alignment
GATK	https://www.broadinstitute.org/gatk/	Variant Calling
Picard	http://broadinstitute.github.io/picard/	Variant Calling
SNPeff	http://snpeff.sourceforge.net	Variant Calling
SnpMap	https://github.com/MolBioBioinformatics/SnpMap	Pipeline wrapper

Table 2
Contents of the Candidate SNP list

Column	Contents
1	Chromosome ID
2	Chromosome position
3	Reference allele
4	Alternate allele
5	SNP quality
6	SNP sequencing depth
7	Annotation
8	Wormbase Gene ID
9	Transcript ID
10	Gene Name