Rapid mapping and identification of mutations in *C. elegans* by RAD mapping and genomic interval pull-down sequencing.


Institute of Molecular Biology,
University of Oregon
Eugene, OR 97403
Running title

RAD mapping and pull-down sequencing

Keywords

mapping, RAD, genome interval pull-down, elegans, Illumina sequencing, GIPS

Corresponding author:

Bruce Bowerman
Institute of Molecular Biology,
University of Oregon
Eugene, OR 97403 USA
phone: 541-346-0853
FAX: 541-346-5891
e-mail: bbowerman@molbio.uoregon.edu
Abstract

Forward genetic screens provide a powerful approach for inferring gene function based on the phenotypes associated with mutated genes. However, determining the causal mutation by traditional mapping and candidate gene sequencing is often the rate-limiting step, especially when analyzing many mutants. We report two genomic approaches for more rapidly determining the identity of the affected genes in *C. elegans* mutants. First, we report our use of restriction site associated DNA polymorphism (RAD) markers for rapidly mapping mutations after chemical mutagenesis and mutant isolation. Second, we describe our use of genomic interval pull-down sequencing (GIPS) to selectively capture and sequence megabase-sized portions of a mutant genome. Together, these two methods provide a rapid and cost-effective approach for positional cloning of *C. elegans* mutant loci, and are also applicable to other genetic model systems.
INTRODUCTION

Determining mutant gene identity is a key step for understanding gene function in forward genetic screens following mutagenesis and phenotype-based mutant isolation. In some organisms such as fungi and bacteria, a recessive mutant allele can be complemented with a plasmid-borne wild-type gene to establish gene identification. In organisms that lack robust DNA transformation methods, mapping with visible or selected single nucleotide polymorphism (SNP) markers to progressively finer genomic intervals is the traditional route to ascertain identity of the mutant gene. Now whole genome sequencing (WGS) methods can significantly reduce the time required to identify the causal mutation. For example, WGS can simply be used to determine all of the sequence alterations present in a mutant strain (Blumenstiel et al. 2009; Irvine et al. 2009; Sarin et al. 2008b; Smith et al. 2008; Srivatsan et al. 2008). However, some mapping data is still required to differentiate the background mutational load from the causal mutation. More recently, WGS has been performed on outcrossed mutant progeny to combine mapping and sequencing for pinpointing the position of the causal mutation (Doitsidou et al. 2010; Zuryn et al. 2010).

While resequencing a genome to identify mutant alleles is being used more frequently, in some cases it is more efficient to sequence only a portion of a genome. For example, sequencing of a single chromosome, a defined genomic interval, exonic sequences, or a single locus can be more
cost effective when there is evidence that a mutation resides within a specific genome feature. There have been several throughput-enhancing advances in capturing targeted regions of a genome using DNA annealing since the first reported use of this methodology whereby individual microarray spots were physically scraped from the substrate (Ksiazek et al. 2003; Rota et al. 2003; Wang et al. 2003). For example, genomic DNA can be annealed to microarrays printed with oligonucleotides covering the region to be targeted, washed and then eluted for sequencing (Albert et al. 2007; Hodges et al. 2007; Okou et al. 2007). Alternatively, oligonucleotides can be used to capture homologous genomic DNA in solution (Gnirke et al. 2009). While these approaches are extremely high throughput, they also can be prohibitively expensive.

We have developed two Illumina-based sequencing methods in C. elegans that offer an alternative pipeline for mutation detection. First, we have performed restriction site associated DNA polymorphism (RAD) mapping to position the causal mutation to a relatively small region of the genome. Second, we have used genome interval pull-down sequencing (GIPS) to sequence a defined genomic interval. Genome intervals are captured by annealing sheared genomic DNA to sheared fosmids containing wild-type C. elegans DNA, eliminating the need for customized microarray or oligonucleotide production. Because multiple RAD mapping and genome interval sequencing samples can
be combined in a single Illumina lane, it is possible to positionally clone and identify the mutant loci rapidly and cost-effectively without performing WGS.
MATERIALS AND METHODS

C. elegans strains and culture
Strains were grown under standard laboratory conditions (BRENNER 1974). The temperature sensitive mutants were maintained in a 15° incubator and shifted to a 26° incubator to perform temperature upshifts for determining embryonic lethality. Mutants were isolated in a lin-2(e1309) background, as previously described (ENCALADA et al. 2000).

Genetic crosses for RAD mapping
To map or1167ts, we crossed the polymorphic C. elegans strain CB4856 into the original mutagenized background [or1167ts/or1167ts; lin-2(e1309)/lin-2(e1309)]. After self-fertilization of the heterozygous F1 outcross, we pooled 200 of the 1/16 of the F2 progeny that were again or1167ts/or1167ts; lin-2(e1309)/lin-2(e1309), taking advantage of the lin-2 egg-laying defect to identify with a stereomicroscope within F1 self-progeny or1167ts/or1167ts; lin-2(e1309)/lin-2(e1309) F2s filling up with dead embryos (avoiding laboriously singling out hundreds of F2s to look for production of dead embryos by egg-laying lin-2(+/+ or +/e1309) F2 progeny). Similarly, for mapping unc-13, we also performed a cross to CB4856 but selected ~200 Unc F2 progeny for the genomic DNA preparations. For mapping or1089ts, we crossed the original mutant to CB4856 males and isolated 800 F2 hermaphrodites that were tested for embryonic lethality. Approximately 200
homozygous \textit{or1089}ts animals were recovered and used for the RAD mapping procedure.

**Illumina sequencing for RAD mapping**

Genomic DNA was isolated from pools of \(~200\) homozygous \textit{unc-13} F2s, and \(~200\) \textit{or1054}ts; \textit{lin-2(e1309)} F2s as well as the N2 and CB4856 parental strains using the Qiagen DNeasy kit. 150 ng of each sample was digested with EcoRI and processed into barcoded RAD libraries as previously described (BAIRD et al. 2008) with the minor modification of using the paired end P2 adapter (HOHENLOHE et al. 2010). Briefly, each sample was individually digested with EcoRI and a P1 adaptor (with a 4 base pair barcode; see below) was ligated to the overhangs. After this step multiple samples were multiplexed. Next, the DNA was sheared and gel-extracted to obtain approximately 400 bp fragments and the Illumina P2 adaptor was ligated. Samples were then run on the Illumina flow cell. For RAD mapping, there is no need to use an Illumina kit (for full protocol see: BAIRD et al. 2008). The RAD library from the mutant pool was sequenced at \(>30x\) coverage in an Illumina Genome Analyzer IIX machine. With SNPs present at about every 1000 base pairs in the polymorphic CB4856 strain, and sequencing reads of about 75 base pairs from each EcoRI site, we anticipated detecting a SNP near 1 in 10 EcoRI sites, or about one every 50,000 base pairs, which was close to the observed value of one SNP in
64,000 base pairs achieved, on average. The RAD sequences were aligned to
the reference Bristol N2 genome using the Bowtie software package
(Langmead et al. 2009). The Bowtie output was then exported to SAMtools (Li et al. 2009) and converted into BAM format. We then produced a pileup file
which we applied the samtools.pl script “varFilter” command (using default
options) to identify snps. The varFilter results were then saved as a tab-
delimited file for use with graphing software (Microsoft Excel and Adobe
Illustrator). As an alternative method for identifying N2/CB4856 snps, one
could use the MAQGene program (Bigelow et al. 2009) which may be more
accessible to non-bioinformaticians (as used by Doitsidou et al. 2010).

**Illumina sequencing of genomic intervals**

To pull down intervals of genomic DNA to which or195ts, or600sd,ts and
or683ts were mapped, we used magnetic bead pull-downs. 5 µg of genomic
DNA was purified from each mutant strain using a DNeasy Blood & Tissue Kit
(Qiagen) and sheared to an average size of 500 bp by sonication in a
Bioruptor (Diagenode). The ends of the sheared DNA fragments were
blunted using a QuickBlunt kit (New England Biolabs), and the fragments
purified with a PCR purification kit (Qiagen). A-overhangs were added to the
genomic fragments by incubation of the purified, blunted DNA with 150 units
of Klenow DNA polymerase exo- (New England Biolabs) and dATP at 37° for
30 minutes. The modified fragments were purified with a mini-elute PCR purification kit (Qiagen). 7µL of 1 µM modified Illumina sequencing adapters (Top strand: 5’ AACTCTTTCCCTACACGACGCTCTTCCGATCATxxxx*T 3’
Bottom strand: 5’ phosphate xxxxGATCGGAAGACGGTTGCTACCAGGAATGCGGAG 3’ (where x indicates the barcode bases, *-phosphorothioate bond),
were ligated to the sheared genomic fragments at 16° for 2 hours using 2000 units of T4 DNA ligase (New England Biolabs). The ligation reaction was size separated by agarose gel electrophoresis, and fragments between 150-500 bp in size were purified from the gel using a Gel Extraction kit (Qiagen). The purified ligation products were PCR amplified using Phusion high fidelity DNA polymerase (New England Biolabs) and the Illumina amplification primers
5’
AATGATACGGCGACCACCGAGATCTACACTTTCCCTACACGACGCTCTTGCGATCT 3’ and
5’
CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCTGCTGAACCGCTCTTCGATCGATCT 3’.
The following cycling conditions were used for PCR: 98° for 2 minutes, 15 cycles of 98° for 10 seconds, 65° for 30 sec, 72° for 15 seconds. Following amplification, samples were size separated by agarose
gel electrophoresis, and fragments between 150 and 500 bp were purified with a Gel extraction kit (Qiagen).

**Biotinylated probe preparation.** DNA preps homologous to the targeted genomic interval were prepared from genomic fosmids, using a nearly genome-wide fosmid library for *C. elegans* that was developed by the Genome Sciences Centre in Vancouver, B.C, Canada. 100 ng of the fosmid DNA mixtures were combined with 20 µL of 2.5x random octamer solution (Life Technologies) and heated to 100° for 5 minutes. The mixture was rapidly cooled in an ice/water bath, following which 5 µL of biotin dNTP mixture (1mM biotin-14-dCTP, 1mM dCTP, 2mM dATP, 2mM dGTP, 2 mM dGTP, in 10 mM Tris-HCl (pH 7.5), 1mM Na₂EDTA) (Life Technologies) was added, along with 1 µL of Klenow fragment DNA polymerase (Life Technologies) and ultra pure water to bring the reaction volume to 50 µL. The reaction was then incubated at 37° for 1 hour, following which, the products were size-separated on an agarose gel and the predominant 100 bp product was purified with a Gel extraction kit (Qiagen).

**Streptavidin bead preparation.** 50 µL of M270 streptavidin Dynabeads (Life Technologies) were washed 3 times with 100 µL of 6x SSC and resuspended in 100 µL of bead block buffer (2% I-Block (Tropix), 0.5% SDS, 1x PBS) (105). Beads were incubated at room temperature for 30 minutes with occasional mixing, and were then magnetically captured, and washed 3x with 6x SSC.
Hybridization, immobilization, elution and sequencing. 5 μg of adapted, purified genomic DNA was combined with 150 ng of purified biotinylated probe in 300 μL of hybridization buffer (54% formamide, 1x SSC, 1% SDS, 5.4x Denhardt’s solution (Sigma), 1mg/ml Salmon sperm DNA (Life Technologies). The mixture was heated to 100° for 2 minutes, then transferred to a 42° incubator, where it was incubated with mixing overnight. Following overnight incubation, biotinylated probe/genomic DNA fragment hybrids were immobilized by binding to prepared blocked and washed streptavidin beads by combining the hybridization mixture (300 μL) with the bead/SSC mixture (100 μL), and incubating at room temperature for 15 minutes with occasional mixing. Beads were then magnetically captured, and washed 3x with wash solution 1 (1x SSC, 0.15% SDS), 3x with wash solution 2 (0.2x SSC), and 3x with wash solution 3 (0.05x SSC). After the final wash step, the beads were resuspended in 200 μL of ultra pure water, and heated to 100° for 2 minutes, and quickly magnetically captured. The supernatant was carefully collected and concentrated to a volume of 20 μL in a speedvac concentrator (Savant). 10 μL of the concentrated supernatant was then used as template for a PCR reaction utilizing Illumina amplification primers and Phusion high fidelity DNA polymerase (New England Biolabs) (2 minutes 98°, 24 cycles of 98° for 10 seconds, 65° for 30 sec, 72° for 15 seconds). The PCR products were
purified with a PCR cleanup kit (Qiagen), quantified, and submitted for Illumina sequencing on an Illumina Genome Analyzer II.
RESULTS

RAD mapping of *C. elegans* mutations.

To rapidly map *C. elegans* mutations, we have used an Illumina sequencing-based genome-wide single nucleotide polymorphism mapping procedure called Restriction site Associated DNA (RAD) polymorphism mapping (BAIRD et al. 2008; LEWIS et al. 2007; MILLER et al. 2007). RAD markers are single nucleotide polymorphisms (SNPs) adjacent to restriction enzyme recognition sequences in the genomes of divergent strains. In our case we used the N2 background (isolated in Bristol, UK) to isolate mutants and subsequently crossed them to the polymorphic Hawaiian CB4856 strain for mapping. The N2 and CB4856 genome sequences have diverged substantially but their hybrid progeny are fertile. On average, there is a SNP approximately every 1 kb, allowing physical mapping using a large number of markers.

To experimentally identify RAD tags, we crossed wild-type N2 hermaphrodites to CB4856 males. F1 hybrid progeny were isolated and genomic DNA was digested with EcoRI. After ligation of Illumina adaptors and selective amplification of the RAD tags (Figure 1), Illumina sequencing was performed with an Illumina Genome Analyzer IIX system. Selective amplification was carried out by using a “Y” adapter for the P2 adaptor.
which prevents fragments which lack a P1 adapter from being amplified after first round synthesis initiated from the P1 site (as described in Baird et al. 2008; Coyne et al. 2004). We detected 3462 SNPs with an average distance between them being 29 kb. Most SNPs (95%) were separated from an adjacent SNP by less than 100 kb (Figure 2). The largest distance separating adjacent SNPs occurred near the center of chromosome V (515 kb). Most of the SNPs we identified could be predicted in silico from the sequence of the CB4856 strain (not shown). Because the sequencing is done with purified RAD tags instead of total genomic DNA, multiple samples can be multiplexed on a single lane in an Illumina sequencer, with each sample containing unique barcodes for subsequent sequence data deconvolution. The barcodes used for RAD mapping are 6 base pair sequences added to the P1 adaptor primer. In one test we used one Illumina Genome Analyzer IIx lane to process 13 RAD mapping crosses. We used single-end sequencing with 80 bp reads, to achieve 50 million reads yielding ~40x coverage. We tested the applicability of RAD mapping coupled with Illumina sequencing using three different approaches.

First, we mapped a known mutant, unc-13(e450). We crossed the unc-13 mutant to the polymorphic C. elegans strain CB4856 and pooled 200 F2 progeny that were homozygous for the unc-13 mutation. We chose 200 recombinants as a goal because it afforded a relatively large number of independent recombination events, although it is possible that using fewer
recombinant F2s would also yield sufficient resolution. After producing a RAD library from the F2 genomic DNA sample, we performed sequencing on the Illumina machine to detect SNPs across the genome (see Methods). We used graphing software (Microsoft Excel and Adobe Illustrator) to plot the ratio of CB4856/Bristol SNPs across the *C. elegans* genomic sequence, indicating the fraction of samples for any one SNP that correspond to the polymorphic CB4856 sequence (Figure 3A). In this test we identified 683 RAD tags throughout the genome. The ratio of CB4856 to N2 SNPs was approximately 0.5 across the genome, except for chromosome I, where a large trough was present. The center of the trough on chromosome I is within ~800 kb of the known location of *unc-13* (Figure 3B). We conclude that Illumina-based RAD mapping can quickly provide the approximate physical position mutant loci.

We similarly applied RAD mapping to *or1089*ts, a mutant of unknown molecular identity with severe defects in mitotic spindle assembly in one-cell stage embryos (Figure 4A). In this case, we identified 3134 RAD markers in the F2 RAD library. Chromosome I was highly enriched for N2 DNA (Figure 4B), with the trough of N2 DNA centered at 9.25 Mb on chromosome I (Figure 4C), positioning the *or1089*ts mutation near the center of chromosome I. In a 1 Mb region centered on the trough, we found one candidate gene in online databases *spd-2*, that when reduced in function using RNAi results in defects that closely resembles the *or1089*ts mutant phenotype (*HARRIS et al. 2010; SONNICHSEN et al. 2005). The center of the
reduced CB4856 ratio is 276 kb to the left of the known location of spd-2 (Fig. 4C). We Sanger sequenced the spd-2 gene in genomic DNA from or1129ts mutants after amplifying the locus using gene-specific primers. We found a single nucleotide mutation that changes a phenylalanine codon to an isoleucine codon at amino acid 544 (Figure 4D). We also found that or1089ts failed to complement spd-2(or293ts) (data not shown). We conclude that or1089ts is a new spd-2 allele.

We have screened for temperature-sensitive, embryonic-lethal C. elegans mutants using an egg laying-defective (Egl) lin-2(e1309) mutant background that enables one to screen mutagenized populations for animals that produce inviable embryos without singling out individual worms (see also: Encalada et al. 2000; Jorgensen and Mango 2002; Kemphues et al. 1988; O'Rourke et al. 2011). To take further advantage of this Egl background for RAD mapping, we crossed or1167ts; lin-2(e1309) hermaphrodites to CB4856 males and obtained F2 animals that were mutant for both lin-2(e1309) and or1167ts. Instead of testing embryonic lethality of ~800 F2 animals individually on plates, as was done for spd-2(or1089ts), we were able to isolate the ~200 or1167ts; lin-2(e1309) animals that accumulated mostly dead embryos from a mixed F2 population after shifting them to the restrictive temperature as L4 larvae. After preparing and sequencing the RAD library, we identified >3400 RAD tags. The ratios of CB4856 to N2 DNA were plotted and we found two troughs that correspond
to an enrichment of N2 DNA (Figure 5). The trough on the X chromosome is 550 kb from the known location of lin-2, while the trough on the right arm of chromosome IV presumably correlates with the location of or1167ts. The fact that the lin-2 trough does not reach zero likely relates to inadvertently picking some worms that were in fact not lin-2, as even wild-type worms sometimes hold their embryos and can appear Egl. We found one candidate gene, sas-6, positioned 250 kb left of the center of the chromosome IV trough. Like animals depleted for sas-6, or1167ts animals shifted to the restrictive temperature as L4 larvae produce embryos that appear to assemble monopolar mitotic spindles in early embryonic cells (not shown). We sequenced the sas-6 locus with the Sanger method after amplification of the region by PCR. We found a single mis-sense mutation that changes an aspartic acid to a valine in the ninth amino acid of SAS-6 (Figure 5D). As or1167ts also failed to complement a known sas-6 mutant (not shown), we conclude that or1167ts is a sas-6 allele. This example demonstrates that RAD mapping can be easily applied in the context of the lin-2(e1309) marker to minimize the effort required to isolate F2 animals that are homozygous for embryonic lethal mutations. We continue to explore the use of RAD methodologies to rapidly map temperature-sensitive embryonic lethal C. elegans mutants and note that this approach can be used to map virtually any locus that can be assayed in N2/CB4856 F2 animals.
Illumina-based Genome Interval Pull-down Sequencing (GIPS).

To quickly and more cost-effectively identify causal mutations in mutant strains, we have also applied Illumina DNA sequencing to defined genomic intervals, rather than sequencing entire mutant genomes (as has been done to identify some mutant loci in *C. elegans* (SARIN et al. 2008a; ZURYN et al. 2010). Sequencing an entire genome involves more cost than our procedure because we multiplexed multiple barcoded sequencing experiments on a single Illumina Genome Analyzer IIx flow cell. Briefly, we used wild-type genomic DNA from defined genomic intervals, linked to magnetic beads, to partially purify regions of mutant genomic DNA (see Figure 6A and Methods). We first tested the feasibility of using interval pull-downs and Illumina sequencing by re-sequencing a previously identified mutation present in the *dhc-1* locus of *dhc-1(or195ts)* mutants worms (O'ROURKE et al. 2007). We then identified the mutations responsible for conditional lethality in two previously reported mutants [*tbb-2(or600sd,ts)* and *plk-1(or683ts)* (O'ROURKE et al. 2011)], after sequencing genomic intervals of 1.8 and 1.3 megabases, respectively (Figure 7).

To test this methodology, we first selected a fosmid that includes the entire *dhc-1* locus, available from Source BioScience (http://lifesciences.sourcebioscience.com). We purified the fosmid, sheared
it and linked the fragments to biotinylated beads as described in the Materials and Methods. After using these beads to isolate sheared dhc-1(or195ts) genomic DNA, the mutant genomic DNA was eluted and prepared for Illumina sequencing (see Methods). We aligned 66,853 30-base reads to the 30.1 kb dhc-1-containing fosmid on chromosome I and the average coverage for each position in the fosmid was 66x. For comparison, we also identified 570,311 reads that could be aligned to the rest of the genome, yielding an average read coverage of 0.17x for each nucleotide position. Therefore, we achieved a 388-fold enrichment for reads in the targeted region using our interval pull-down sequencing method. We identified the previously-sequenced C to T dhc-1(or195ts) mutation in a total of 66 reads, and no other mutations were detected in the dhc-1 locus (Figure 7A). We conclude that GIPS can readily identify the mutations present in relatively large regions of the genome.

Next, we used GIPS to identify the mutation responsible for the early embryonic cell division defects caused by the semi-dominant, temperature-sensitive mutation or600sd,ts. We defined the or600sd,ts interval using standard mapping crosses with visible morphological and behavioral markers. We localized the mutation to chromosome III between positions 3,618,381 and 5,447,436. We used the Wormbase genome browser to identify a minimal tiling path using genomic DNA from fosmid clones available from Source BioScience (http://lifesciences.sourcebioscience.com).
We identified 65 fosmids that spanned the region with 7 gaps that totaled less than 45 Kb (~2.5% of the region). We purified, sheared, and linked the fosmid fragments to biotinylated beads. After using these beads to isolate sheared or600sd,ts genomic DNA, the mutant genomic DNA was eluted and prepared for Illumina sequencing (see Methods). We found 1,596,403 48-base reads that could be aligned to the region giving an average coverage for each nucleotide in the interval of 42x. The total number of reads corresponding to the C. elegans genome was 6,034,221.

We used software (SAMtools) to output a text file that lists the mutations identified in the interval. One can also view the sequence reads aligned to the reference wild-type genome sequence with SAMtools (Figure 6B). There were 45 mutations in the 1.8 Mb or600sd,ts interval (Figure 7B). We found 14 extragenic changes, 23 intronic changes, one mutation in a pseudogene, 2 mutations that cause the transcript to go out of frame, 3 potential annotation errors, and 2 mis-sense mutations (Table 1). As over 90% of our identified temperature sensitive mutations are caused by mis-sense mutations (the remaining are due to premature stop codons, small deletions, and mutations in splice site boundaries (O'Rourke et al. 2011)), we narrowed our analysis of mutations in exons and intron splice sites. Three mutations that appeared to cause exonic changes are likely not the causal mutation because they were present in wild-type DNA: expressed sequence tags show the same mutations, perhaps indicating errors in the
reference sequence (we have labeled these as “annotation error?” in Table 1). None of the intronic changes occurred at intron boundaries and thus are unlikely to interfere with RNA splicing. The four remaining exonic mutations occur in the *tbb-2, ras-2, his-70,* and *clec-154* genes. Single nucleotide deletion and insertion mutations in the *his-70* and *clec-154* loci, respectively, encode proteins with altered C-termini, while the *ras-2* and *tbb-2* mutations are mis-sense. Of these four genes, only RNAi that targets the *tbb-2* locus phenocopies the *or600*sd,ts early embryonic phenotype (note that *tbb-2(RNAi)* also depletes the paralogous redundant *tbb-1* gene product). Depletion of the other three genes by RNAi does not result in any lethal phenotypes (Wormbase). Based on sequence data, the embryonic phenotype and genetic interactions with a previously isolated *tbb-2* allele, *or362*sd,ts (O’ROURKE et al. 2011), we conclude that *or600*sd,ts is a *tbb-2* allele.

We also identified a recessive temperature-sensitive embryonic lethal mutant, *or683*ts, that mapped between 6862157-8214712 on chromosome III, based on mapping crosses with visible markers and some PCR-based SNP mapping (not shown). To identify all of the mutations in the interval, we again performed GIPS, as detailed for *dhc-1(or195*ts) and *tbb-2(or600*sd,ts). There were 197,752 48-base reads (from a total of 5,616,109 reads) that aligned to the region giving an average coverage of 7x for the interval. Even though the sequencing coverage was greatly reduced compared to the *dhc-1*(or195*ts) and *or600*sd,ts results, we were still able to identify 29 mutations
in the interval after performing Illumina sequencing (Figure 7C). We found 9 intergenic changes, 8 mutations in introns, 2 mutations in 3’ untranslated regions, one mutation in a transposon, 6 insertions or deletions that caused frame shifting, 2 potential annotation errors, and one mis-sense allele (Table 2). The annotation errors may represent the wild-type sequence because expressed sequence tags show this same alteration. The single mis-sense mutation changes a methionine codon to a lysine codon in the \textit{plk-1} gene. Since \textit{or683ts} contains a mutation in the \textit{plk-1} gene and fails to complement a known allele of \textit{plk-1} (O’ROURKE et al. 2011), we conclude that \textit{or683ts} is a novel \textit{plk-1} allele.

\textbf{DISCUSSION}

The utility of \textit{C. elegans} as an animal model in which one can readily isolate temperature-sensitive mutations in essential genes, and the power of next generation DNA sequencing for greatly reducing the time required to positionally clone mutant loci, now make it possible to much more rapidly isolate experimentally useful conditional mutations in essential genes. Our two new Illumina-based sequencing methods should allow for increased throughput when analyzing large numbers of mutants.
Restriction site associated DNA polymorphism (RAD) mapping has been used successfully to map genes in Drosophila (Miller et al. 2007), threespine stickleback fish (Miller et al. 2007), Neurospora (Lewis et al. 2007), diamondback moths (Baxter et al. 2011), barley (Chutimanitsakun et al. 2011) and now C. elegans. So long as a hybrid strain is available to generate F2 progeny, the methodology should be feasible in any organism. RAD mapping makes it possible to simultaneously and rapidly determine the approximate location of large numbers of mutations isolated after mutagenesis. We have used EcoRI to cut the genomic DNA because it provides relatively good resolution between RAD markers (Figure 1). However, any other restriction enzyme could also be used, or multiple enzymes could be used to gain increased mapping resolution. RAD mapping provides an ideal procedure to identify mutant loci when the number of gene candidates is limited.

We explored the use of RAD mapping in three different contexts. First we have performed a proof of principle experiment where we mapped the position of unc-13, a known mutant (Figure 3). While the numbers of RAD markers obtained was relatively low in this example (683, presumably because we experienced some sample loss), the center of the trough on chromosome I is still within 800 kb of the known position of unc-13. In the second approach, we used RAD mapping to clone the or1089ts mutant. After picking 800 F2 animals from an or1089ts/CB4856 cross, we identified those that were homozygous for the or1089ts mutation. We found a substantial enrichment of the N2 DNA on the center of chromosome I that was positioned within 276 kb from the known position of the spd-2 locus. We
performed Sanger DNA sequencing on PCR products derived from the \textit{spd-2} locus and identified one sequence alteration that causes a mis-sense mutation (Figure 4D). Thus, RAD mapping can rapidly identify candidate genes that can be further investigated by sequencing candidate genes, GIPS, or complementation tests with existing mutants. We also tested the feasibility of performing RAD mapping on F2 animals that were doubly mutant for an embryonic lethal mutation and an egg-laying defective mutant, \textit{lin-2} (present in the original mutagenized strain). Since homozygous \textit{lin-2} animals hold their embryos, it was possible to more easily and rapidly select F2 progeny homozygous for the mutation being mapped from an \textit{or1167ts}; \textit{lin-2} and CB4856 cross (Figure 5). As expected, we found two regions of the genome that were enriched for N2 DNA: one corresponds to the \textit{lin-2} locus while the second corresponds to the \textit{or1167ts} mutation in the \textit{sas-6} locus. Finally, we are currently exploring the use of RAD mapping by crossing temperature-sensitive mutants to CB4856 males and allowing the progeny to reproduce at the nonpermissive temperature for many generations. This method may significantly reduce the labor in isolating the homozygosed F2 progeny, and would show an exclusion of N2 DNA corresponding to the lethal locus.

In a second use of Illumina sequence technology, we have developed a genome interval pull-down method to sequence defined portions of the genome. By sequencing only intervals that contain the causal mutation, one can reduce the expenses associated with whole genome sequencing. We have successfully applied this technology to positionally clone two new
mutants so far (Figure 8), as well as the control *dhc-1(or195ts)* mutant. We identified 45 sequence alterations in the 1.83 Mb *or600sd,ts* interval and 29 mutations in the 1.3 Mb *or683ts* interval versus the Wormbase reference sequence. Thus, if we had sequenced the entire genome of these mutants, we would have found many mutations. Therefore, it is clearly important to have some mapping data to narrow the search for the causal mutation, and RAD mapping fills this role well. As costs continue to come down, whole genome sequencing using either recombinant F2 animals (DOITSIDOU et al. 2010) or backcrossed mutants (ZURYN et al. 2010), two techniques that simultaneously map and sequence mutations, may become more cost-effective than the relatively more labor-intensive GIPS. Nevertheless, RAD mapping may continue to prove useful for analyzing large numbers of mutants, as many mutants can be identified by sequencing only candidate genes in the vicinity, or by complementation tests with previously identified alleles in the region. In fact, a large number of non-conditional mutants exist that can be used for performing complementation tests (for example, CLARK and BAILLIE 1992; JOHNSEN and BAILLIE 1991; JOHNSEN et al. 2000; STEWART et al. 1998). GIPS should also remain useful for sequencing candidate genes that are too large to easily amplify with PCR for Sanger sequencing.

As of August 2011, the cost to sequence the entire *C. elegans* genome at 30x coverage is about $600 (U.S.) on the HiSeq2000 platform. With this many reads, one could perform 50 RAD mapping experiments (at 30x
coverage) or about 50 GIPS procedures using 2 Mb pull-down regions. Both of these sequencing techniques can also be run on Illumina runs with other samples (with the use of barcoded adaptors). Depending on the type of mutant being sequenced, using WGS strategies will be more straightforward (and more time- and cost-effective), for a recent review, see Doitsidou et al. (submitted to Wormbook). For example, if the mutant locus is resistant to RNAi, or if large-scale RNAi screens have not assayed the phenotype represented by the mutant, then WGS is likely the best approach. However, if the mutant phenotype is likely to be recapitulated by RNAi, such as early embryonic lethality (as we are studying), then RAD mapping will reveal a limited number of candidate genes. Sanger sequencing, complementation tests with existing mutants, or GIPS could then identify the causal mutation, although the time required to perform this two-step approach is longer than using a mapping/WGS approach. Thus, RAD mapping may be a viable alternative to WGS when large numbers of mutants are being investigated. Similarly, GIPS can be useful for sequencing loci that are known to be defective in many different mutants. For example, suppressor screens often identify many intragenic suppressor alleles (for example, see: GREENWALD and HORVITZ 1980) which, depending on size, can be very time consuming to Sanger sequence, yet performing WGS may be too expensive with many alleles to sequence. GIPS fills this gap and allows the simultaneous sequencing of many different mutant loci on the same Illumina lane with the
use of barcoded samples. In conclusion, we offer two new strategies for mutant identification in *C. elegans* that can fill roles not currently provided by WGS for certain applications.
Author contributions

SMO analyzed RAD mapping and genome interval pull-down data, performed alignments, performed Sanger sequencing, and wrote the manuscript. JY, AC, and JL isolated mutants and performed crosses for RAD mapping. DWT performed Illumina sequencing and purified the DNA samples for genome interval mapping. NS, MRM, and NKH provided bioinformatics analysis, NKH also prepared DNA for RAD mapping. LC and MHP isolated mutants and prepared DNA for genome interval pull-downs. EAJ conceived of the genome interval pull-down method. BB designed the mutant isolation and helped to draft the manuscript. All of the authors have read and approved the final manuscript.
Description of additional data files

The following additional data are available with the online version of this paper. Additional data file 1 is a table listing the genomic positions and sequence of the EcoRI-associated RAD tags present in the N2 and CB4856 strains using the WS190 referential version of Wormbase.


**Acknowledgments**

We thank members of the Bowerman and Johnson labs for help in isolating conditional mutants and developing mapping and sequencing protocols. This work was supported by a Leukemia and Lymphoma Society of America fellowship to SOR, and NIH grants GM050817 and GM049869 to BB.
References

ALBERT, T. J., M. N. MOLLA, D. M. MUZNY, L. NAZARETH, D. WHEELER et al., 2007
Direct selection of human genomic loci by microarray hybridization.
Nat Methods 4: 903-905.

BAIRD, N. A., P. D. ETTER, T. S. ATWOOD, M. C. CURREY, A. L. SHIVER et al., 2008
Rapid SNP discovery and genetic mapping using sequenced RAD

BAXTER, S. W., J. W. DAVEY, J. S. JOHNSTON, A. M. SHELTON, D. G. HECKEL et al.,
2011 Linkage Mapping and Comparative Genomics Using Next-
Generation RAD Sequencing of a Non-Model Organism. PLoS ONE 6:
e19315.

BIGELOW, H., M. DOITSIDOU, S. SARIN and O. HOBERT, 2009 MAQGene: software
to facilitate C. elegans mutant genome sequence analysis. Nat
Methods 6: 549.

BLUMENSTIEL, J. P., A. C. NOLL, J. A. GRIFFITHS, A. G. PERERA, K. N. WALTON et al.,


CHUTIMANITSAKUN, Y., R. W. NIPPER, A. CUESTA-MARCOS, L. CISTUE, A. COREY et al.,
2011 Construction and application for QTL analysis of a Restriction Site Associated DNA (RAD) linkage map in barley. BMC Genomics 12:
4.

CLARK, D. V., and D. L. BAILLIE, 1992 Genetic analysis and complementation
by germ-line transformation of lethal mutations in the unc-22 IV

COYNE, K. J., J. M. BURKHOLDER, R. A. FELDMAN, D. A. HUTCHINS and S. C. CARY,
2004 Modified serial analysis of gene expression method for
construction of gene expression profiles of microbial eukaryotic

DOITSIDOU, M., R. J. POOLE, S. SARIN, H. BIGELOW and O. HOBERT, 2010 C.
elegans mutant identification with a one-step whole-genome-
sequencing and SNP mapping strategy. PLoS ONE 5: e15435.


Figure Legends

Figure 1. Restriction site Associated DNA (RAD) mapping schematic.
Genomic DNA was isolated from 200 pooled F2 progeny in crosses between
the N2 mutants and the polymorphic CB4856 (Hawaiian) strain. Genomic
DNA was digested with EcoRI and P1 Illumina adaptors were ligated to the
fragments. This DNA was mechanically sheared and Illumina P2 adaptors
were ligated to the fragment ends. Next, RAD tags were selectively amplified
and sequenced from the Illumina sequencing primer site encoded on the P2
adaptor on an Illumina genome analyzer IIx machine.

Figure 2. EcoRI-associated RAD tag locations.
A. RAD mapping results from an N2/CB4856 cross. Vertical lines represent
EcoRI-associated RAD markers on each of the *C. elegans* chromosomes. The
total chromosome sizes are listed on the right. For a list of the RAD markers,
sequences, and positions, see Supplemental Dataset 1. B. Distance between
adjacent RAD markers. The plot represents the number of RAD marker pairs
at the indicated distances, in kilobase pairs.

Figure 3. RAD mapping results for *unc-13*.
A. Genome-wide RAD mapping results for *unc-13* after crossing it to
CB4856. 683 snps across the genome in F2 progeny were detected and the
ratio of the CB4856 snps plotted along the chromosomes. The vertical axis represents the percentage of CB4856 SNPs in the F2 population. B. Magnification of chromosome I. The trough on chromosome I correlates with the known location of unc-13.

**Figure 4. Phenotype and RAD-mapping of spd-2(or1089ts).**
A. Defective mitotic spindle formation and cytokinesis failure in an early or1089ts embryo produced from a worm shifted to 26° for six hours (lower panels) compared to a wild-type embryo (upper panels). B. Genome-wide RAD-mapping of or1089ts. A reduction of CB4856 DNA on chromosome I results from the selection of the mutant homozygotes. C. The center of the trough is near 9 Mb and is positioned near the causal mutation (see text). D. DNA sequence analysis of or1089ts identifies a mutation in spd-2 open reading frame [GenBank: AY340594.1]. The mutation causes a phenylalanine to isoleucine change in codon 544, is shown relative to the changes of or183 and or188, two known temperature-sensitive alleles of spd-2.

**Figure 5. RAD mapping results for a sas-6(or1167ts); lin-2 double mutant.**
or1167ts; lin-2 mutants were crossed to CB4856 males and F2 progeny homozygous for both the or1167ts embryonic lethal mutation and the lin-2
mutation were isolated. A. Genome-wide RAD mapping results show enrichment for N2 DNA on chromosomes IV and X. B. The sas-6 gene is located near the chromosome IV trough. C. The trough on the X chromosome lies near the known position of the lin-2 locus. D. The sas-6 locus contains a mis-sense mutation that changes an aspartic acid codon to a valine [NCBI: NP_502660.1]. An alignment was performed with the C. elegans (Cel) wild-type SAS-6, C. elegans SAS-6(or1167ts), C. remanei (Cre) SAS-6, and C. briggsae (Cbr) SAS-6.

**Figure 6. Genome Interval Pull-down Sequencing (GIPS) using the Illumina platform.**

A. Schematic overview of the interval pull-down sequencing method. First, fosmids of wild-type DNA covering a region of the genome are purified, sheared, and ligated to biotinylated adaptors. Next, mutant genomic DNA is sheared and annealed to the biotinylated fosmids. After purification and release of the mutant DNA using magnetic beads, the fragments are subjected to sequencing on an Illumina machine. Finally, the reads are assembled onto the genome scaffold and polymorphisms are identified. B. Example output of mutant genome assembly. Shown is a small region of the tbb-2 locus with portions of reads aligned beneath the reference sequence. In this case each read shows that a cytosine in wild type has been changed to a thymidine in tbb-2(or600sd,ts). Also shown is one apparent sequencing
error where an A>G change was called in one of the reads. Nucleotides are color-coded, capital letters and periods in the sequence reads represent identity with the reference sequence, while reads containing small case letters and commas represent sequence data obtained from the reverse complement strand.

**Figure 7. Analysis of the genome interval pull-down sequencing.**
The protocol shown in Figure 7 was applied to three mutants. The positions of nucleotide alterations versus the reference sequence is shown in graphical form. Each line represents a single change, diagonal lines attached to the top of the vertical lines represent multiple changes located very close together. Lines are color-codes as shown in the figure to show various types of mutations. The “^” in the black bars point to the causative mutations. A. Sequencing results for the 15 kb dhc-1(or195ts) locus purified using a single genomic fosmid clone. The previously identified mis-sense mutation was found. B. Sequencing results for a 1.82 Mb region on chromosome III in the or600sd,ts genome. We identified 45 total mutations in the region including the mis-sense mutation in tbb-2 (see Table 1 for details). C. Sequencing results for a 1.3 Mb region on chromosome III in the or683ts genome. We identified 29 total mutations in the region including the mis-sense mutation in plc-1 (see Table 2 for details).
Table 1. Mutations present in the or600sd,ts interval.

<table>
<thead>
<tr>
<th>Position 1</th>
<th>Gene</th>
<th>WT</th>
<th>Alteration</th>
<th>Notes</th>
<th>Reads 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>3645769</td>
<td>unc-93</td>
<td>C</td>
<td>A</td>
<td>intron</td>
<td>20</td>
</tr>
<tr>
<td>3645821</td>
<td>unc-93</td>
<td>T</td>
<td>G</td>
<td>intron</td>
<td>58</td>
</tr>
<tr>
<td>3716562</td>
<td>C14B1.6</td>
<td>*</td>
<td>C insertion</td>
<td>intron</td>
<td>22</td>
</tr>
<tr>
<td>3805251</td>
<td>toh-1</td>
<td>C</td>
<td>single base deletion</td>
<td>intron</td>
<td>591</td>
</tr>
<tr>
<td>3830219</td>
<td>intergenic</td>
<td>*</td>
<td>G insertion</td>
<td></td>
<td>23</td>
</tr>
<tr>
<td>3885443</td>
<td>tir-1</td>
<td>*</td>
<td>G insertion</td>
<td>intron</td>
<td>30</td>
</tr>
<tr>
<td>3971283</td>
<td>msi-1</td>
<td>T</td>
<td>A</td>
<td>intron</td>
<td>45</td>
</tr>
<tr>
<td>3997832</td>
<td>intergenic</td>
<td>A</td>
<td>T</td>
<td></td>
<td>142</td>
</tr>
<tr>
<td>4007014</td>
<td>pdx-1</td>
<td>*</td>
<td>C insertion</td>
<td>intron</td>
<td>22</td>
</tr>
<tr>
<td>4017108</td>
<td>tbb-2</td>
<td>C</td>
<td>T</td>
<td>mis-sense G&gt;E, E</td>
<td>614</td>
</tr>
<tr>
<td>4019053</td>
<td>intergenic</td>
<td>*</td>
<td>C insertion</td>
<td></td>
<td>920</td>
</tr>
<tr>
<td>4019157</td>
<td>intergenic</td>
<td>*</td>
<td>G insertion</td>
<td></td>
<td>1104</td>
</tr>
<tr>
<td>4032835</td>
<td>hum-5</td>
<td>*</td>
<td>C insertion</td>
<td>intron</td>
<td>72</td>
</tr>
<tr>
<td>4056056</td>
<td>rcq-5</td>
<td>A</td>
<td>T</td>
<td>intron</td>
<td>28</td>
</tr>
<tr>
<td>4056057</td>
<td>rcq-5</td>
<td>T</td>
<td>A</td>
<td>intron</td>
<td>28</td>
</tr>
<tr>
<td>4056058</td>
<td>rcq-5</td>
<td>A</td>
<td>C</td>
<td>intron</td>
<td>29</td>
</tr>
<tr>
<td>4059721</td>
<td>his-70</td>
<td>T</td>
<td>single base deletion</td>
<td>altered C terminus, NE</td>
<td>112</td>
</tr>
<tr>
<td>40741</td>
<td>unc-79</td>
<td>T</td>
<td>single base deletion</td>
<td>intron</td>
<td>341</td>
</tr>
<tr>
<td>4091848</td>
<td>C03C10.2</td>
<td>*</td>
<td>C insertion</td>
<td>intron</td>
<td>122</td>
</tr>
<tr>
<td>4209222</td>
<td>intergenic</td>
<td>C</td>
<td>A</td>
<td></td>
<td>1959</td>
</tr>
<tr>
<td>4209223</td>
<td>intergenic</td>
<td>A</td>
<td>C</td>
<td></td>
<td>2105</td>
</tr>
<tr>
<td>4209244</td>
<td>intergenic</td>
<td>C</td>
<td>A</td>
<td></td>
<td>2694</td>
</tr>
<tr>
<td>4209245</td>
<td>intergenic</td>
<td>G</td>
<td>C</td>
<td></td>
<td>2546</td>
</tr>
<tr>
<td>4210819</td>
<td>F43C1.1</td>
<td>*</td>
<td>G insertion</td>
<td>intron</td>
<td>716</td>
</tr>
<tr>
<td>4286570</td>
<td>R10E4.3</td>
<td>C</td>
<td>T</td>
<td>intron</td>
<td>30</td>
</tr>
<tr>
<td>4445015</td>
<td>intergenic</td>
<td>*</td>
<td>G insertion</td>
<td></td>
<td>24</td>
</tr>
<tr>
<td>4447068</td>
<td>ceh-43</td>
<td>*</td>
<td>G insertion</td>
<td>intron</td>
<td>81</td>
</tr>
<tr>
<td>4447641</td>
<td>ceh-43</td>
<td>*</td>
<td>C insertion</td>
<td>intron</td>
<td>83</td>
</tr>
<tr>
<td>4502805</td>
<td>cct-5</td>
<td>*</td>
<td>G insertion</td>
<td>intron</td>
<td>1140</td>
</tr>
<tr>
<td>4530454</td>
<td>cdh-4</td>
<td>A</td>
<td>single base deletion</td>
<td>intron</td>
<td>55</td>
</tr>
<tr>
<td>4605001</td>
<td>sel-2</td>
<td>T</td>
<td>single base deletion</td>
<td>intron</td>
<td>25</td>
</tr>
<tr>
<td>4612783</td>
<td>intergenic</td>
<td>*</td>
<td>C insertion</td>
<td></td>
<td>411</td>
</tr>
<tr>
<td>4640496</td>
<td>clec-154</td>
<td>*</td>
<td>G insertion</td>
<td>altered C terminus, NE</td>
<td>195</td>
</tr>
<tr>
<td>4684393</td>
<td>T04A8.3</td>
<td>G</td>
<td>single base deletion</td>
<td>pseudogene</td>
<td>45</td>
</tr>
<tr>
<td>4687673</td>
<td>intergenic</td>
<td>*</td>
<td>C insertion</td>
<td></td>
<td>82</td>
</tr>
<tr>
<td>4689556</td>
<td>T04A8.5</td>
<td>*</td>
<td>C insertion</td>
<td>intron</td>
<td>64</td>
</tr>
<tr>
<td>4708949</td>
<td>intergenic</td>
<td>*</td>
<td>C insertion</td>
<td></td>
<td>780</td>
</tr>
<tr>
<td>4712953</td>
<td>T04A8.15</td>
<td>C</td>
<td>G^6</td>
<td>annotation error?</td>
<td>626</td>
</tr>
<tr>
<td>4738437</td>
<td>ras-2</td>
<td>C</td>
<td>T</td>
<td>mis-sense G&gt;R, NE</td>
<td>53</td>
</tr>
<tr>
<td>4808804</td>
<td>intergenic</td>
<td>*</td>
<td>C insertion</td>
<td></td>
<td>91</td>
</tr>
<tr>
<td>4947903</td>
<td>intergenic</td>
<td>*</td>
<td>C insertion</td>
<td></td>
<td>59</td>
</tr>
<tr>
<td>4962073</td>
<td>C27F2.5.1</td>
<td>T</td>
<td>single base deletion</td>
<td>intron</td>
<td>44</td>
</tr>
<tr>
<td>5105425</td>
<td>F54D8.6</td>
<td>*</td>
<td>C insertion</td>
<td>intron</td>
<td>280</td>
</tr>
<tr>
<td>5105939</td>
<td>F54D8.6</td>
<td>*</td>
<td>G insertion^6</td>
<td>annotation error?</td>
<td>133</td>
</tr>
<tr>
<td>5333260</td>
<td>F52C9.1</td>
<td>*</td>
<td>C insertion^6</td>
<td>annotation error?</td>
<td>977</td>
</tr>
</tbody>
</table>

^1Chromosome III position from the WS200 reference version of Wormbase.
The wild-type sequence.
The changes found in the *or600* genome. For insertions, the alteration takes place immediately following the indicated position.
For mutations that cause protein coding changes, we have classified the loci as “NE” for not essential, or “E” for essential.
The total number of Illumina reads for the position. Only mutations with 20 or greater reads were considered and furthermore, more than 75% of the reads had to show the alteration presented in column 4.
These apparent mutations may represent the wild-type sequence in some backgrounds because expressed sequence tags also show these alterations.
Table 2. Mutations present in the or683ts interval.

<table>
<thead>
<tr>
<th>Position</th>
<th>Gene</th>
<th>WT²</th>
<th>alteration³</th>
<th>Notes⁴</th>
<th>Reads⁵</th>
</tr>
</thead>
<tbody>
<tr>
<td>7029617</td>
<td>intergenic</td>
<td>*</td>
<td>A insertion</td>
<td></td>
<td>11</td>
</tr>
<tr>
<td>7047283</td>
<td>intergenic</td>
<td>*</td>
<td>C insertion</td>
<td></td>
<td>20</td>
</tr>
<tr>
<td>7063069</td>
<td>R01H2.4</td>
<td>*</td>
<td>C insertion</td>
<td>altered C terminus, NE</td>
<td>15</td>
</tr>
<tr>
<td>7214105</td>
<td>R151.2</td>
<td>A</td>
<td>C⁶ insertion</td>
<td>annotation error?</td>
<td>27</td>
</tr>
<tr>
<td>7228289</td>
<td>R151.8</td>
<td>AAC</td>
<td>deletion</td>
<td>intron</td>
<td>11</td>
</tr>
<tr>
<td>7272182</td>
<td>psd-1</td>
<td>*</td>
<td>G insertion</td>
<td>intron</td>
<td>26</td>
</tr>
<tr>
<td>7297779</td>
<td>B0361.2</td>
<td>*</td>
<td>AA insertion</td>
<td>intron</td>
<td>10</td>
</tr>
<tr>
<td>7318365</td>
<td>F56C9.8</td>
<td>*</td>
<td>TT insertion</td>
<td>intron</td>
<td>11</td>
</tr>
<tr>
<td>7587395</td>
<td>cup-5</td>
<td>G</td>
<td>deletion</td>
<td>intron</td>
<td>12</td>
</tr>
<tr>
<td>7628358</td>
<td>fbn-1</td>
<td>TG</td>
<td>deletion</td>
<td>altered C terminus, E</td>
<td>11</td>
</tr>
<tr>
<td>7686828</td>
<td>C18H2.3</td>
<td>*</td>
<td>TA insertion</td>
<td>3' UTR</td>
<td>10</td>
</tr>
<tr>
<td>7696424</td>
<td>C18H2.4</td>
<td>TA</td>
<td>deletion</td>
<td>altered C terminus, NE</td>
<td>12</td>
</tr>
<tr>
<td>7712933</td>
<td>intergenic</td>
<td>*</td>
<td>G insertion</td>
<td></td>
<td>17</td>
</tr>
<tr>
<td>7781836</td>
<td>mab-5</td>
<td>AG</td>
<td>deletion</td>
<td>intron</td>
<td>11</td>
</tr>
<tr>
<td>7864173</td>
<td>C02C2.4</td>
<td>C</td>
<td>deletion</td>
<td>intron</td>
<td>23</td>
</tr>
<tr>
<td>7919577</td>
<td>tag-250</td>
<td>*</td>
<td>GA insertion</td>
<td>altered C terminus, NE</td>
<td>12</td>
</tr>
<tr>
<td>7937298</td>
<td>ran-2</td>
<td>TC</td>
<td>deletion</td>
<td>3' UTR</td>
<td>11</td>
</tr>
<tr>
<td>7958664</td>
<td>F54H12.3</td>
<td>C</td>
<td>T</td>
<td>transposon</td>
<td>15</td>
</tr>
<tr>
<td>8042705</td>
<td>K12H4.6</td>
<td>*</td>
<td>G insertion</td>
<td>altered C terminus, NE</td>
<td>32</td>
</tr>
<tr>
<td>8053797</td>
<td>intergenic</td>
<td>C</td>
<td>deletion</td>
<td></td>
<td>10</td>
</tr>
<tr>
<td>8059766</td>
<td>intergenic</td>
<td>C</td>
<td>T</td>
<td></td>
<td>21</td>
</tr>
<tr>
<td>8064605</td>
<td>intergenic</td>
<td>C</td>
<td>T</td>
<td></td>
<td>17</td>
</tr>
<tr>
<td>8068402</td>
<td>ceh-26</td>
<td>*</td>
<td>G insertion⁶</td>
<td>annotation error?</td>
<td>11</td>
</tr>
<tr>
<td>8070697</td>
<td>ceh-26</td>
<td>*</td>
<td>G insertion</td>
<td>intron</td>
<td>25</td>
</tr>
<tr>
<td>8075189</td>
<td>dcr-1</td>
<td>*</td>
<td>GG insertion</td>
<td>altered C terminus, E</td>
<td>19</td>
</tr>
<tr>
<td>8103295</td>
<td>plk-1</td>
<td>T</td>
<td>A</td>
<td>mis-sense M&gt;K, E</td>
<td>12</td>
</tr>
<tr>
<td>8127908</td>
<td>intergenic</td>
<td>*</td>
<td>T insertion</td>
<td></td>
<td>12</td>
</tr>
<tr>
<td>8128979</td>
<td>intergenic</td>
<td>*</td>
<td>C insertion</td>
<td></td>
<td>30</td>
</tr>
<tr>
<td>8154518</td>
<td>intergenic</td>
<td>T</td>
<td>A</td>
<td></td>
<td>10</td>
</tr>
</tbody>
</table>

¹Chromosome III position from the WS200 reference version of Wormbase.
²The wild-type sequence.
³The changes found in the or683 genome. For insertions, the alteration takes place immediately following the indicated position.
⁴For mutations that cause protein coding changes, we have classified the loci as “NE” for not essential, or “E” for essential.
⁵The total number of Illumina reads for the position. Only mutations with 10 or greater reads were considered and furthermore, more than 90% of the reads had to show the alteration presented in column 4.
⁶These apparent mutations may represent the wild-type sequence in some backgrounds because expressed sequence tags available from Wormbase also show these alterations.
Figure 1

- Digest hybrid genomic DNA (EcoRI)
- Ligate P1 adaptor & Shear
- Ligate P2 adaptor
- Selectively amplify RAD tags & Illumina sequence
Figure 2

A

<table>
<thead>
<tr>
<th>Chromosome</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>15.1 Mb</td>
</tr>
<tr>
<td>II</td>
<td>15.3 Mb</td>
</tr>
<tr>
<td>III</td>
<td>13.8 Mb</td>
</tr>
<tr>
<td>IV</td>
<td>17.4 Mb</td>
</tr>
<tr>
<td>V</td>
<td>20.1 Mb</td>
</tr>
<tr>
<td>X</td>
<td>17.7 Mb</td>
</tr>
</tbody>
</table>

B

Distance between adjacent SNPs (kb)

Number of SNPs

- <25: 2285
- 25-50: 606
- 50-75: 257
- 75-100: 145
- 100-200: 121
- 200-300: 33
- 300-400: 6
- 400-500: 1
- 500-600: 1
Figure 3

O’Rourke et al

A

Fraction of CB4856 SNPs

Chromosomes

B

Fraction of CB4856 SNPs

Chromosome I (Mb)

unc-13 (7.42 Mb)
Fraction of CB4856 SNPs

Chromosome I (Mb)

0 4 8 12 16

0.0

0.1

0.2

0.3

0.4

spd-2 (8.97 Mb)

Figure 4

O’Rourke et al

A

wild type

-220 s 0 s 235 s 590 s 1155 s

or1089

-234 s 0 s 314 s 764 s 1014 s

B

Fraction of CB4856 SNPs

I II III IV V X

0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0

Chromosome

C

Fraction of CB4856 SNPs

Chromosome I (Mb)

0 4 8 12 16

spd-2 (8.97 Mb)

D

I or1089

543 SFSPSSVGYQVIMSIEVPANFIHKPMWNGGIAKFVPTSPDLQQTINQSEYAMCTSCAKRISFKISNSAGT 616

S or183

E or188
**Figure 6**

**A**
- Genomic DNA
- Target region
- Fosmids
- Sheared genomic DNA
- Biotinylated fosmid fragments
- Purification
- Magnetic bead
- Illumina sequencing

**B**
- Chromosome III coordinates
- Reference Sequence
- Mutant Consensus
- Portions of Individual Sequence Reads

C>T mutation in *tbb-2*(or600)
Figure 7

A

single nucleotide change

Chr I

4.39 4.40

15.1 Mb

B

single nucleotide change
insertion
deletion

Chr III

3.62 5.45

13.8 Mb

C

single nucleotide change
insertion
deletion

Chr III

6.86 8.21

13.8 Mb