Towards a mutation in every gene in Caenorhabditis elegans

Donald G. Moerman and Robert J. Barstead

Advance Access publication date 16 April 2008

Abstract

The combined efforts of the Caenorhabditis elegans Knockout Consortium and individuals within the worm community are moving us closer to the goal of identifying mutations in every gene in the nematode C. elegans. At present, we count about 7000 deletion alleles that fall within 5500 genes. The principal method used to detect deletion mutations in the nematode utilizes polymerase chain reaction (PCR). More recently, the Moerman group has incorporated array comparative genome hybridization (aCGH) to detect deletions across the entire coding genome. Other methods used to detect mutant alleles in C. elegans include targeting induced local lesion in genomes (TILLING), transposon tagging, using either Tc1 or Mos1 and resequencing. These combined strategies have improved the overall throughput of the gene-knockout labs, and have broadened the types of mutations that we, and others, can identify. In this review, we will discuss these different approaches.

Keywords: C. elegans; mutations; transposon tagging; array comparative genome hybridization; resequencing

INTRODUCTION

In May 1998, a small group of scientists met at the Sanger Center in Hinxton, England to discuss the feasibility of mounting a large-scale Caenorhabditis elegans project to mutate every single gene in this organism. The impetus for this meeting was the impending publication of the complete sequence of this small roundworm [1]. As the first metazoan to reach this historic milestone, C. elegans offered an unprecedented opportunity for a comprehensive genetic dissection of the development and physiology of different tissues and cell types. Clearly, to fully exploit the potential of this model system would require mutations in all the genes.

The yeast community was already underway with a genome-wide gene-knockout project at the time of the Hinxton meeting. Making use of the facile homologous recombination and gene-disruption methods available for this system, in 1999 yeast workers reported the creation of deletion alleles for over 2000 genes [2]. This was followed 3 years later by a nearly complete deletion set for all genes in Saccharomyces cerevisiae [3]. At present, yeast is the only eukaryote with a set of deletions for all genes in the genome. As homologous recombination was not well developed for C. elegans (and this method remains elusive even today), it was clear at the Hinxton meeting that the nematode project would require the development of new, scalable methods if we wished to develop a similar resource for the worm community. In this review, we will describe the history of the development of new methods to knockout C. elegans genes, and we will report on the present status of others and our efforts to obtain mutations in every gene in C. elegans.

FORWARD GENETIC PRECURSORS

Though the methods differ, today’s genetic objectives are conceptually the same as those in the pre-genome era; that is, we seek to generate mutations and map them to molecules. In the pre-genome era, there were several ways to do this. Most methods began with forward genetic mutant hunts looking for visible phenotypes. The mutant hunts were then followed by efforts to identify the mutant
gene(s). Several strategies were used to traverse the path from mutant to cloned gene. These include (i) the use of DNA transposable elements and (ii) the development of an ordered set of overlapping DNA clones across the genome. Subsequently, we discuss how these methods and reagents were used to identify the DNA sequences associated with a given genetic locus.

Transposon tagging in Caenorhabditis elegans was modelled after methods that were first developed in the fruit fly, Drosophila melanogaster [4]. In C. elegans, the transposable element most widely exploited initially was Tc1 [5]. To tag a gene, one began with a hunt for spontaneous mutants in a C. elegans strain that was especially active in germline transposition and excision of Tc1 [6–8]. Then, using Southern blots and radioactively labelled probes for Tc1, one sought to identify new Tc1 insertion events and correlate those with the mutant phenotype. The challenge was to identify such events in strains where the presence of large numbers of Tc1 elements often hid newly transposed copies. Further, one had to contend with a relatively high level of background transposition. Nevertheless, the method was effective as shown by the discovery of the DNA sequences encoding the genetic loci lin-12 and unc-22 [9, 10].

Though effective in some cases Tc1 gene tagging is limited because this element does not transpose randomly throughout the genome; rather it has a decided gene and sequence composition bias [11–14]. Therefore, once the C. elegans genome was fully represented by an ordered collection of yeast artificial chromosomes and cosmid clones [15, 16] and DNA transformation methods were developed [17, 18], positional cloning became the method used most often to connect a genetic locus with DNA sequence. Positional cloning methods depend on the strength of the correlation between the physical and genetic maps [19]. Correlating these two maps is done using DNA sequence that can be followed in genetic crosses. Such markers include cloned genetic loci, restriction fragment length polymorphisms (RFLP) and single nucleotide polymorphisms (SNPs) [20–22]. Further, the beauty of positional cloning is that as each new mutant gene is cloned, one has yet another anchor point to connect the genetic and physical maps thus simplifying subsequent projects.

**REVERSE GENETICS**

The approaches described earlier to link a genetic locus to a DNA sequence are based fundamentally on forward genetic screens where one starts with animals that express a mutant phenotype and then correlates the responsible genetic loci with DNA sequence. Often the route from genetic locus to a molecularly defined gene is long and arduous. In 1989 Ballinger and Benzer [23] proposed an alternative approach, one that today we call ‘reverse genetics’. Reverse genetics begins with a cloned gene and proceeds to the derivation of a mutant. Using Drosophila and P transposable elements, Benzer and colleagues demonstrated that one could select for a P element insertion in a gene of choice and detect it molecularly in a mixed population of animals. This approach, using the polymerase chain reaction (PCR) with primers that match both the gene of interest and the terminal sequence of the P element, proved to be rapid, robust and convenient such that Ballinger and Benzer was compelled to write ‘[the method] could be utilized to detect insertion of a foreign element into any gene for which at least a partial sequence is known and could be extended to other organisms’ [23].

Philip Anderson and colleagues were the first to demonstrate that this approach could be applied to the worm when they isolated a strain where a C. elegans transposable element, Tc1, inserted into and disrupted the expression of a myosin light chain gene (mlc-1) [24]. This work, however, illustrated both the power and the challenge of reverse genetics; that is, even though the strain was clearly defective in the expression of the targeted gene, it did not exhibit any obvious visible phenotype. The gene, therefore, would not have appeared in any simple forward genetic mutant hunt. Moreover, even with the Tc1 allele, a genetic analysis of its function would be complex because such an analysis would depend on finding conditions (e.g. in combinations with other mutant genes) where the mutant allele led to an altered phenotype.

The work by Anderson and colleagues was followed by efforts to scale the method for a genome-wide project. Ronald Plasterk and colleagues constructed a frozen bank of clonal worm strains isolated from a so-called mutator strain. To demonstrate the effectiveness of this frozen bank, these workers described six mutant derivatives isolated using PCR to detect the mutant allele [25].

As Anderson, Plasterk and others worked to develop this method, however, a significant flaw was discovered. It was known that Tc1 had a very high bias for A/T rich sequences (as mentioned earlier),
and so many insertion events captured using a PCR assay resulted in the disruption of intron sequence, which often did not disrupt gene function. Even more fatal, however, one often found insertions into coding sequence that were not functionally disruptive because the Tc1 insert was spliced out of the pre-mRNA [26]. Plasterk and colleagues found a solution to this nearly fatal flaw that, again, was modelled after work done in *Drosophila*. Like other IS type transposons, *Tc1* when it excises leaves a double-strand gap that can lead to illegitimate repair and thus generate a deletion at the site [27–29]. This too could be assayed using PCR. For the most part, therefore, the use of *Tc1* to generate functional mutants required two steps: (i) the recovery of a *Tc1* insertion allele and (ii) the recovery of a derived deletion allele. The end result is that while the Plasterk modification did save the method, it introduced a level of complexity that was not scalable for a whole genome project.

Though *Tc1* tagging was impractical for a genome-wide project, the PCR methods developed to detect *Tc1* insertion events were applied to the development of the robust reverse genetic methods we use today [30]. As first described by Plasterk and colleagues, one could use PCR on DNAs from genetically heterogeneous populations to identify deletion alleles at given targets following chemical mutagenesis [31]. These workers reported data, which showed that two chemical mutagens, ethyl methane–sulphonate (EMS) and trimethylpsoralen (UV/TMP), which had already been shown to cause deletions [32], could be used successfully to generate deletion alleles in selected target genes. In similar studies, Liu and colleagues [33] reported the results of PCR–based screens following treatment with four chemical mutagens: EMS, UV/TMP, ethynitrosourea (ENU) and diepoxynitrobutane. Further, these workers increased the scale at which one could generate and screen mutant banks when they described methods to culture worms in 96-well microplates.

THE GENESIS OF A GENOME-SCALE REVERSE GENETIC CONSORTIUM

We point to several important milestones in route to the present day effort to generate deletions in all *C. elegans* genes. First, the Ballinger and Benzer work was seminal, as it became the foundation for present day post-genome efforts to recover mutations in all sequenced genes. Second, Rushforth and Anderson developed a PCR–based method to detect gene rearrangements in genetically heterogeneous populations. Third, Plasterk and colleagues and Liu et al. showed how to scale the Anderson methods to the level required for a genome-wide project, the focus of the Hinxton meeting.

Several of those in attendance at Hinxton had a great deal of reverse genetic expertise; indeed many in attendance had developed the protocols described above (notably Plasterk, Liu and Barstead as well as scientists from the companies DevGen, Nemapharm/Axis and Exelixis). What emerged from the meeting was the idea that the time was ripe to form a consortium of laboratories dedicated to providing the research community with gene knockouts (i.e. deletions or ko’s) upon request. The Consortium as it stands today consists of three laboratories: the Barstead group located at the Oklahoma Medical Research Foundation (OMRF) in Oklahoma USA, the Mitani group located at Tokyo Women’s University, Tokyo, Japan and the Moerman group located at the University of British Columbia (UBC) in Vancouver, Canada.

Though the details differ, the fundamental strategy for each consortium lab is the same; populations of animals are exposed to a mutagen and PCR is used to identify deletions at target loci. Our standard protocols have been described in detail elsewhere ([34]; http:// www.zoology.ubc.ca/%7Edgmweb/research1_pcr.htm) and various detailed derivatives of the protocol are also available [35], so the protocol itself will not be described in detail here (Figure 1 gives an overall view of the methodology). A fair assessment of our progress over the years, however, shows that although we have not significantly increased our conceptual understanding of the chemistry and genetics of mutagenesis, each consortium lab has contributed data to improve the operational aspects of the method. As examples, Genkyo-Ando and Mitani [36] explored UV/TMP mutagenesis in order to provide a more effective mutagenesis protocol. Second, the Barstead and Moerman labs developed methods to increase the sensitivity of the PCR assay for the detection of small deletions, the so-called ‘Poison Primer’ method [37].

Despite the overall methodological similarities in the approaches taken by each consortium group, the groups do differ in detail. For example, the Barstead group differs significantly from the Moerman and
Mitani groups in its use of EMS rather than UV/TMP as the primary mutagen. Further, the Barstead group has developed an extensive robotics platform. Second, the Mitani group uses a large frozen mutant bank, whereas the Barstead and Moerman groups use fresh mutant libraries.

The three deletion laboratories are largely, but not solely, request-driven. Requests are handled via the following two websites: http://celegankoconsortium.omrf.org/ or http://shigen.lab.nig.ac.jp/c.elegans/index.jsp. Prior to making a request one first should confirm that no deletion alleles are available for the gene of interest. The best way to do this is to visit either of the web sites above, or the websites of the Caenorhabditis Genetics Center (CGC) or WormBase, the primary data repository.

Figure 1: Summary of current deletion protocol. (A) We partially synchronize a large population of C. elegans. We then treat L4 hermaphrodites with UV/TMP (or ethyl methane–sulphonate, EMS). The F1 progeny of the mutagenized animals are harvested and distributed in microtitre plates as shown in (B). The worms are cultured 6–7 days. When the wells are cleared of food, we harvest 33% of each population and prepare the DNA for PCR. We do nested PCR with primers spaced about 3000 Bp apart as shown in (C, D) The results from a screen for deletion in the gene K07D4.7 are shown. The arrow points to a positive address that was then followed through the final sib selection stage.
for this model system. Each mutant allele generated by a consortium lab is numbered sequentially and a two letter prefix is appended to the number that serves to identify the originating lab. All mutant alleles, whatever their origin, are sequenced and the sequence data is deposited in WormBase. All strains produced in Japan are available directly from the Mitani group. In contrast, the UBC and OMRF groups do not distribute strains directly, but rather all strains from these two groups are deposited in the main C. elegans strain repository, the CGC in St. Paul, Minnesota. Further, to insure the stability of strains that carry alleles that are homozygous lethal, all strains produced at the OMRF and UBC are balanced before they are sent to the CGC.

This PCR method, while robust, is not without its problems. These include false positive PCR assays and the fact that one cannot control the deletion endpoints. As such, some deletion alleles are not molecular nulls, some fall entirely within large introns, some involve complex mutational events that are not easily resolved with the typical PCR assay, and some eliminate more than one gene. Nevertheless, the majority of the alleles produced using this method provide an adequate entrée to the genetic analysis of the target gene. As well, over the years we have developed strategies to deal with all of these complications.

In addition to outside user requests, the OMRF and UBC groups are targeting genes by functional group. At present, we are working on DNA binding transcription factors (934 genes) and kinases (438 genes) [38, 39]. To date, together the worm community and the Consortium laboratories have identified mutations in 602 (71%) of all transcription factors and over half of the kinases. We expect to complete these two projects in the next year.

Although the Consortium labs account for most of the gene knockouts done in this organism over the past half dozen years, there is a cottage industry of independent labs that have constructed frozen mutant banks which have also contributed to the effort to generate deletion alleles in all C. elegans genes. To date there are over 7000 alleles identified in 5500 genes. Alleles in about 4000 of these genes have been generated by one of the three consortium laboratories. Many more knockout strains are in progress in each of the screening laboratories. The three laboratories are on pace to bring the total number of knockouts to 10,000 genes by 2010. Although this is only 50% of the ~20,000 genes in the worm, given the relatively primitive methods available for the knockout of C. elegans genes, the worm community can be truly proud of this collective effort. Finally, the current pace of the work and the selection of target genes for most Consortium work would not have been possible if a sequenced genome work were not first available.

NEW REVERSE GENETIC METHODS

As of early 2008, the standard PCR/deletion protocol is still the major engine for delivering single gene deletion mutations. A number of other methods have been or will be used to complete the C. elegans project. Among these alternative methods we will discuss targeting induced local lesion in genomes (TILLING), Mos1 insertions, array comparative genome hybridization (aCGH), and resequencing, or deep sequencing, using next generation DNA sequencing machines.

TILLING is a technique developed in the laboratory of Steven Henikoff and first applied on a large scale to the plant Arabidopsis thaliana [40, 41]. In this method, PCR amplified target gene DNAs from control and mutagen-treated samples are mixed, denatured, and allowed to reanneal. Sequences that are not completely complementary due to a mutation(s) will be mismatched in the heteroduplex. Such mismatched sites are cleaved using a single strand specific nuclease (for example CEL 1). As the amplifying primers have different coloured fluorescent labels the cleavage fragments resolve into two labelled fragments with sizes that add up to the total size of the PCR amplicon. When analysed using a LI-COR Biosciences sequencing system, one can resolve the position of the mismatch cleavage site to within a few nucleotides. When one uses an efficient mutagen like EMS, and depending on the number of genomes sampled, one can recover multiple alleles within a given target gene. The best case scenario is that these multiple alleles will fall into an allelic series for the gene. A recent application of TILLING to C. elegans targeted 10 genes and obtained several mutations for each [42]. As might be expected, many of the alleles generated in this work were silent, some were likely missense alleles, and a few were putative nulls.

As described above, at one time the C. elegans transposable element Tc1 was a common tool to
both mutate and identify the molecular lesion in *C. elegans* genes. While identification and cloning of genes using Tc1 is now a relic of the past, the basic principals are still employed, but now substituting strains engineered with the fly transposable element Mos1. The transposon Mos1 was first identified and characterized in *Drosophila* [43, 44], and was later shown to be active in *C. elegans* [45–47]. In the genetically engineered *C. elegans* strains Mos1 is low to single copy, so it is a simpler molecular tag compared to Tc1 for tagging and recovering mutated genes.

A consortium of European laboratories has identified over 10,000 Mos1 inserts across the *C. elegans* genome [48] (NEMAGENETAG website at http://elegans.imbb.forth.gr/nemagenetag/; [49]; Laurent Segalat, personal communication; Mos1 insert data available at WormBase and the above website). A portion of these Mos1 insertions are in exons, but similar to the transposon Tc1, Mos1 prefers thymine and adenine (TA)-rich regions so many more are in introns and intergenic regions. Such insertion events may still be useful, however, because, as with Tc1, one can use them to derive strains where the transposable element has excised imprecisely, taking part of an adjacent gene with it. Nevertheless, many genes are now disrupted by Mos1 and for those genes that are recalcitrant to chemical mutagens, imprecise excision of a nearby Mos1 insert may be the only recourse. Perhaps most important, however, as suggested by studies done with P elements in flies [27, 50] and Tc1 in *C. elegans* [51], one can use Mos1 insertions to facilitate the replacement of the Mos1 allele with another engineered allele, so-called gene replacement, a method now known as MosTIC [52].

The newest way to identify mutant alleles in *C. elegans* is array aCGH [53]. The intellectual roots of this approach lay in fluorescent in situ hybridization (FISH). Instead of a fluorescently labelled probe hybridized to an entire chromosome, however, one hybridizes fluorescently labelled sheared DNA to oligo probes on a chip (Figure 2). For our deletion detection studies in the nematode we have used the Nimblegen platform which, at present, allows for a probe density of almost 370,000 oligos of 50 base length. The Nimblegen platform offers a high level of design flexibility and this has allowed us to test a wide variety of oligo and chip parameters. Though we have used chips where an entire region is represented by a tiled set of probes, we more typically screen chips where only the exons are represented. We have designed chips for the whole genome, specific chromosomes or specific regions within chromosomes, depending on the density of oligos required for the specific experiment.

In our initial aCGH screens we first used standard forward genetic methods to collect lethal or morphological mutants [53]. We then used genomic DNA isolated from the resulting strains to hybridize

---

**Figure 2**: Schematic of array Comparative Genome Hybridization protocol. See text for details.
DNA from each mutant strain was isolated, sheared and then labelled with Cy3. Control genomic DNA, typically from wild-type strains, was labelled with Cy5. Both labelled DNA preparations were hybridized to chips. The light intensity of each probe was read using standard methods. The intensity ratios were used to identify regions of DNA within the mutants that were either increased or reduced when compared to the control. Our work shows that homozygous deletions are relatively easy to detect (Figure 3). An even more exciting result is that we have demonstrated that one can detect deletions in balanced heterozygous lethal mutant strains as long as the noise around the ordinate is low (Figure 4, A).

To identify a deletion with confidence we demand that three or more consecutive oligo probes show a reduced hybridization on the chip. The size of deletion that can be detected is dependent on the density of oligos on the chip. We have detected deletions as big as 770 000 base pairs and as small as 30 base pairs. An additional advantage of aCGH is that one can detect whether a deletion is contiguous or discontinuous. Moreover, many deletions have duplications at their ends and this too can be detected by aCGH (Figure 3). The resolution of the deletion endpoints is also dependent on the density of the oligos on the chip. In one example, we were able to resolve a deletion breakpoint to <25 nt because of the fortuitous placement of two oligos that overlapped, with only one oligo being included within the deletion [53].

A novel and exciting new application of aCGH is the detection of SNPs. In preliminary experiments, we have been able to reliably detect single base alterations within 1 MB regions tiled with 50 mer probes on Nimblegen chips (J. Maydan, S. Flibotte, M. Edgley and D.M., unpublished results; an example is shown in Figure 4, B). There does not seem to be any bias in the types of alterations detected as we have observed a variety of transitions and transversions involving all possible nucleotide combinations.

One key to identifying SNPs is to have a very high-tilling density of oligos such that many oligo’s span the mismatch site. Further, hybridizations need to be very clean with a low level of noise as the signal is never more, (and often less), than that observed for heterozygous deletions. Even given our best conditions, however, not all SNPs are detected; at present the false negative rate is about 60%.

We have not yet determined the false positive rate. Finally, a new generation Nimblegen chip will allow for a 5-fold increase in probe density compared with current technologies (over 2 million sites as opposed to 370 k), which will result in greater resolution for deletions and their endpoints, and a much higher sensitivity for the detection of single-base mutations.

It seems only fitting to end a review dedicated to celebrating the 10th anniversary of the sequencing of the *C. elegans* genome with a discussion of resequencing genes to identify mutations. Edwin Cuppen and colleagues [54] recently presented data...
that demonstrated that resequencing may be a cost effective means to obtain nonsense and missense alleles of \textit{C. elegans} genes. Starting with a mutant library derived from over 6000 individual F1 animals after EMS mutagenesis (about 12 000 haploid genomes), PCR was used to isolate DNA for 32 target genes from each sample population. Over 1000 independent mutations were found using standard Sanger sequencing on each of the 192 000 samples (32 genes/C2/6000 samples). Based on these data, these workers estimated a mutation load of 1000 point mutations per genome after treatment with 0.05 M EMS. Similarly, our CHIP SNP study discussed above showed approximately six or more SNPs within each of several 1 MB intervals, supporting this resequencing data (J. Maydan, S. Flibotte, M. Edgley and D.M., unpublished results). These combined results should be quite sobering and we hope it will encourage geneticists to outcross and backcross strains even more diligently. At this measured mutation rate, therefore, one can calculate that individual animals derived from such a mutagenesis may contain mutations in 250 genes, more-or-less. Some number of these will either be silent third base DNA sequence changes or conservative amino acid changes in the protein. At present, the data does not allow for a thorough analysis of the likely number of useful mutational events, but Cuppen \textit{et al}. [54] showed that for one locus 10% of the derived mutations resulted in premature stop codons. Thus, mutagenized individuals from this study were likely to be heterozygous for as many as 25 loss-of-function mutations.

The Cuppen \textit{et al}. [54] study used standard Sanger sequencing methods. Newer, more cost effective next generation methods make resequencing even more attractive. For example, the Solexa/Illumina platform can generate sufficient sequence data to give 5-fold coverage of the \textit{C. elegans} genome with a single pass on one sequencing machine. As proof in principal of this approach a recent study compared an N2 Bristol isolate to the canonical sequence and then searched the strain CB4858 for SNP’s and small indels [55]. The results of this study are very encouraging as they obtained almost 100% coverage of the unique portions of the \textit{C. elegans} genome and they were able to identify and confirm many SNP variants. If high-sequence quality can be achieved consistently, Illumina deep sequencing, or an analogous approach, may be the best means to

**Figure 4:** Detection of a heterozygous deletions and an example of chip SNP. (A) Detection of the 1202 bp deletion in \textit{dab-1}(gk291) in a balanced heterozygous population. Normalized fluorescence for probe pairs targeting \textit{dab-1} is shown. Vertical dotted lines indicate the sequenced deletion breakpoints. (A from Maydan \textit{et al}. [53].) (B) A single base change in \textit{unc-112}(r367). The change is a C to T transition altering Threonine (85) to Isoleucine. For this chip the probes were piled densely around the nucleotide change site. Note also that a noisy chip, i.e. one with a lot of fluctuation around zero would make a SNP or a heterozygous deletion very difficult to identify.

**Key Points**

- There are now over 5500 genes with loss-of-function mutations in \textit{C. elegans}. Note that these are not all null alleles as the current screening methods for mutations are still blunt instruments and cannot be used to precisely eliminate the full activity of a gene in all cases.
- New methods exist, in particular tiling chips (aCGH) and resequencing, offering the promise that mutations may be obtained at an accelerated rate. These alternative methods also offer the promise of obtaining a wider spectrum of mutations affecting gene expression and function.
- Over the next 2 years half the genes in this organism will have mutations. Included in this group will be mutations in all the transcription factors and all the kinases, two very important gene families with major roles in all aspects of development and homeostasis.
obtain comprehensive mutation coverage of the C. elegans genome.

Acknowledgements
We thank Stephane Flibotte, our long-term collaborator on aCGH. We thank Laurent Segalat for communicating details of the Mosl project and Marco Marra for discussing the possibilities of next generation sequencing machines. We thank Jason Maydan and Alexandra Rogula for providing Figure 2 and Jason Maydan for Figure 4 panel B. We particularly thank members of our staff at UBC, especially Mark Edgley, and the OMRF and the laboratory of Shohei Mitani, all of whom contribute so much to making the Knockout Consortium a success. Research in the laboratory of DGM described in this review was supported by Genome Canada, Genome British Columbia, the Michael Smith Research Foundation and the Canadian Institute for Health Research. Research in the laboratory of RJB described in this review was supported by the United States National Institutes of Health (P41HG003652).

References


