FILE S2

Supporting Methods

Preparation of mutants: *C. elegans* were cultured using standard methods (BRENNER 1974). We conducted an EMS mutagenesis screen in N2 nematodes carrying a *cog-1::gfp* transgene (PS3662 strain) that localizes GFP expression to the PDA neuron and thus allows us to select mutants that lack PDA (J. Richard, S.Z. , N. Fischer, V. Pavet, N. Vaucamps & S.J., submitted). Three mutants belonging to 3 distinct complementation groups were chosen and backcrossed to the original PS3662 strain (6X for *fp6* and *fp12*, and 4X for *fp9*). We recommand that a mutagen introducing typical changes at a high frequency be used, as this facilitates the subsequent causal mutation identification. For example, damage induced by ethane methyl sulfonate (EMS) or *N*-ethyl-*N*-nitrosourea (ENU), two commonly used chemical mutagens, may be detected by the canonical nucleotide transitions they cause. Although the type of mutagen used depends on the organism to be mutagenised, one aspect that impacts on this strategy, alongside a predictable mutation type, is the mutation rate of the mutagen. In the centre of the linked regions for each mutant we sequenced, we observed G/C > A/T nucleotide changes at a frequency of 7-8 per Mb. This would correspond to a frequency of 1 mutation in every ~125 000-143 000 base pairs from a 50 mM EMS dose. In the un-linked regions of the genome, where backcrossing would have removed most, if not all, of the EMS-induced nucleotide changes, we observed between 0-2 typical EMS-induced changes per Mb. Even though a small proportion of these G/C > A/T nucleotide changes may have occurred spontaneously through genetic drift, this particular mutation load presented us with very obvious high-density variation peaks in which to concentrate our search for the causal mutation. However, a lower mutation rate may also suffice. ENU has been reported to induce 0.5-1 mutation in every 100 000 nucleotides in mice (BEIER 2000) and would thus be appropriate for use with our strategy.

Genetic drift, mixed-origin reference genome and backcrossing: A number of studies have suggested that natural genomic variations occur between strains of the same species, especially in laboratory conditions. For example, strains of *C. elegans* continuously grown for more than 2 years have been suggested to accumulate as many mutations as after EMS treatment (DENVER *et al.* 2004). The use of WGS has started to provide extensive molecular evidence of the existence of hundreds to thousands of differential variants between the N2 sequenced reference genome and N2 derived laboratory strains has been described (HILLIER *et al.* 2008; SARIN *et al.* 2008). The existence of mutation accumulation and genetic drift represents a challenge to mutation identification using strategies that involve comparison of the genome sequence of mutants to a reference genome, considering that, in addition to mutagenesis, the mutants may have accumulated other new variants spontaneously. In addition, a reference genome can have been made from an assembly of a number of different individuals (as is the case for the ongoing Zebrafish genome [www.sanger.ac.uk/Projects/D_rerio]). However, genetic drift or a mixed-origin reference genome do not represent an issue for mutation identification using our strategy, as the mutant genomes are compared between each other for identification and removal of background variants. This strategy vastly increases the overall robustness of mutation identification by minimizing false-positive and false-negative results caused by reference genome errors. We recommend backcrossing to the original strain that was subjected to mutagenesis. Doing so ensures that all mutant strains will share the same background variation, which can be subsequently subtracted. Outcrossing of the mutant strains to another wild-type isolate is also possible, but may yield more than one region of high density variation: one around the causal mutation and at least one around any other genomic locus that has to be kept in the mutant strain, such as the integration site of a reporter transgene. The number of high-density regions increases with the number of loci that need to be kept during outcrossing. We believe that the risk of having a mixed variants signature when another strain than the original strain is used to outcross is very low if outcrossing has been performed thoroughly (4X should be enough based on our experience of removing EMS-induced changes outside of linked genomic regions). Depending on the speed of genetic drift within a given species, backcrossing mutants from a mutagenesis screen in a timely manner with the original un-mutagenised strain will minimize drift. If backcrossing cannot be performed directly after the screen and if, like *C. elegans*, populations can be frozen and retrieved at a latter time, we recommand to keep a frozen aliquot of the original strain used for the screen and freeze the mutants retrieved directly after the screen. We also recommand to backcross multiple mutants in parallel if they are to be sequenced together.

Preparation of genomic DNA and genome sequencing: Genomic DNA was prepared from populations of each mutant using the Gentra Puregene Kit (Qiagen). 10 mg of genomic DNA was fragmented by nebulisation (according to Illumina instructions) to obtain fragments in the range of 500bp in size. Sequencing libraries were made according to the Illumina protocol. The three paired-end libraries were sequenced at the IGBMC sequencing platform on the Illumina GAII as 57-bp paired end reads, following the manufacturer's protocols. Each mutant was sequenced on two flow cell lanes producing coverage of 52.2-55.3X for the 3 mutants across the genome (SI table 2). Image analysis and base calling was performed with Illumina Pipeline version 1.6 with default parameters.

Analysis software: Sequences were mapped to *E. coli* strain 536 using Bowtie (LANGMEAD *et al.* 2009) version 0.12.0. Subsequently, short read alignment and variant calling were performed using MAQGene software (BIGELOW *et al.* 2009) revision 33. Mutant reads were aligned against the N2 reference genome (wormbase.org version WS201). MAQGene was used with default parameters except for the max distance between two paired reads which was set to 1000. Comparison of called variants between mutants was performed with custom Perl scripts. Before filtering out common nucleotide variants between our mutants we observed between 2 336 and 2 457 single-nucleotide differences between our mutants and the N2 reference genome. After this step, the number of variations was dramatically reduced to between 415 to 488 nucleotides, thus eliminating approximately 2 000 point mutations as potential candidates for our causal mutation. This result strongly

emphasizes the advantage of conducting WGS on two or more mutants side-by-side, as reference genomes may contain many nucleotide variations when compared to organisms sequenced from the laboratory, and as such would confound mutation identification. Alternatively, one may conduct WGS on the original starting strain used for mutagenesis, which could be used to eliminate strain specific variations. However, in the interests of reducing costs it would be more advantageous to directly sequence multiple mutants of the same background instead. Quality filtering involved selecting only those variants with a MAQgene mapping score of 63 with 0 wild-type reads. EMS may cause other genetic changes such as small insertions or deletions (indels); however, for mapping purposes, we concentrated on the most typical EMS-induced change, as these were sufficient to identify a region containing the causal mutation. After the genomic region is identified quality filtering may be removed to reveal all variants (of all quality) detected within the mapped region, thus avoiding false negatives. In our experience, an EMS-induced variant linkage region is still discernable without firstly performing quality filtering suggesting that this step is not absolutely required for our strategy.

Confirmation of *egl-5(fp6)* allele: Sanger-based sequencing on *fp6* animals was performed using the forward primer: CAAGCTTCTGCAAGGAATGCCT and the reverse primer: TTACGGTGGACACAACGGGTAT to generate an amplicon of 439bp containing the expected variant. Sequencing on both strands confirmed the presence of the variant. Genetic complementation analysis between *fp6* and *egl-5(n945)* mutants (recessive and null) yielded no complementation. RNAi of *egl-5* by the feeding method was performed as previously described (KAMATH *et al.* 2001) and was conducted in RNAi sensitive *rrf-3(pk1426);cog-1::gfp* animals.

Supporting literature cited

BEIER, D. R., 2000 Sequence-based analysis of mutagenized mice. Mamm Genome **11:** 594-597.

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

BRENNER, S., 1974 The genetics of Caenorhabditis elegans. Genetics **77:** 71-94.

DENVER, D. R., K. MORRIS, M. LYNCH and W. K. THOMAS, 2004 High mutation rate and predominance of insertions in the Caenorhabditis elegans nuclear genome. Nature **430:** 679-682.

HILLIER, L. W., G. T. MARTH, A. R. QUINLAN, D. DOOLING, G. FEWELL *et al.*, 2008 Whole-genome sequencing and variant discovery in C. elegans. Nat Methods **5:** 183-188.

KAMATH, R. S., M. MARTINEZ-CAMPOS, P. ZIPPERLEN, A. G. FRASER and J. AHRINGER, 2001 Effectiveness of specific RNA-mediated interference through ingested double-stranded RNA in Caenorhabditis elegans. Genome Biol **2:** RESEARCH0002.

LANGMEAD, B., C. TRAPNELL, M. POP and S. L. SALZBERG, 2009 Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol **10:** R25.

SARIN, S., S. PRABHU, M. M. O'MEARA, I. PE'ER and O. HOBERT, 2008 Caenorhabditis elegans mutant allele identification by whole-genome sequencing. Nat Methods **5:** 865-867.