

MAQGene: software to facilitate *C. elegans* mutant genome sequence analysis

To the Editor: Forward genetic screens for mutants in which specific biological processes are disrupted are a key strength of model systems like *Caenorhabditis elegans* or *Drosophila melanogaster*. However, the steps necessary to go from isolating a phenotype-causing mutant strain to identifying the molecular nature of the genetic change, most often merely a single point mutation, are cumbersome, traditionally involving time-consuming genetic mapping strategies. We have recently shown in a proof-of-principle study that conventional positional cloning can be shortcut through the use of whole-genome sequencing (WGS) with massively parallel, deep-sequencing technology^{1,2} (Supplementary Table 1). A similar proof-of-principle approach has proven successful for *D. melanogaster* mutant identification as well³.

The key challenge of the WGS approach is the mapping of millions of small (<100 base pair) reads, obtained from sequencing the mutant genome, to a wild-type reference genome. Various mapping tools are available for this purpose, including efficient large-scale alignment of nucleotide databases (ELAND) or mapping and assembly with quality (MAQ)⁴. A disadvantage of many of these tools is that implementation, use and data output formats may be non-intuitive for many biologists and may require outside bioinformatic support that is not always readily available.

To circumvent this problem and thereby help popularize the WGS approach, we developed a user-friendly, simple web browser interface, called MAQGene, that automatically launches the publicly available MAQ software and assembles a customized summary of the location and specific features of sequence variants of the mutant genome compared to a wild-type reference genome (Fig. 1). The MAQGene submission form allows the user to select specific parameters for aligning and interpreting WGS reads (Supplementary Note 1). Default parameters that we have used to analyze mutant *C. elegans* genomes are provided in the installation package and are easily reconfigured to suit individual preference. MAQGene may handle reads up to 127 bases long and map in both single-read or paired-end modes. The output file (Supplementary Note 1) is easily convertible to an Excel spreadsheet and allows easy browsing of sequence variants as well as comparisons of different genomes (which is, for example, helpful to subtract background variants).

Various measures are provided in the output file to allow the user to rapidly assess the degree of coverage for a given nucleotide position and the likelihood that a nucleotide variant is indeed real and of functional relevance. For example, provided the reference genome has all exons annotated, as is the case for the *C. elegans* genome, each variant is indicated as being intronic, intergenic, within a protein-coding gene (and if so whether the variant is silent, missense, splice site or nonsense) or within an annotated noncoding RNA. These features are sortable in the

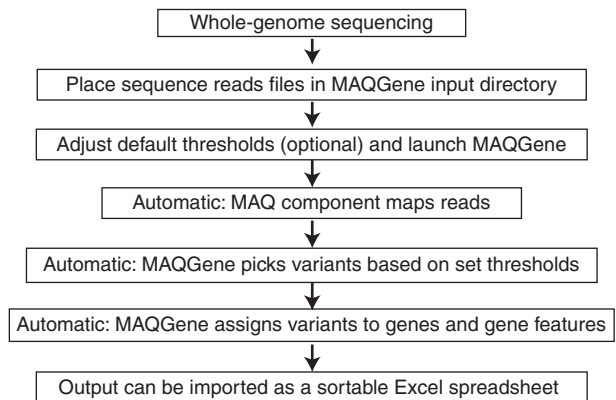


Figure 1 | Whole-genome analysis with MAQGene.

output file, allowing for the generation of a ‘priority list’ of variants which are to be chosen for validation by Sanger resequencing and for tests probing functional relevance. The output file can also be easily filtered so as to reveal variants present specifically in a genetically mapped interval.

Using data generated by an in-house Illumina Genome Analyzer II platform, we used MAQGene to identify sequence variants in more than six different *C. elegans* genomes compared to the wild-type *C. elegans* reference genome. In principle, MAQGene also provides the option to compare any input WGS reads (in fastq format) to any wild-type reference genome that is available in fasta format with GFF (general-feature format) annotation files, thereby easily allowing adaptation of MAQGene to analyze, for example, WGS data from *D. melanogaster* mutant strains.

Updated versions of MAQGene (Supplementary Software) are available at <http://maqweb.sourceforge.net>. Detailed descriptions of MAQGene, its installment and its use can be found in the package itself and in Supplementary Note 1.

Note: Supplementary information is available on the Nature Methods website.

ACKNOWLEDGMENTS

This work was supported by the Howard Hughes Medical Institute, the US National Institutes of Health (NIH) (R01NS039996-05; R01NS050266-03), a postdoctoral NIH training grant to H.B. (5T32HD055165-03) and a predoctoral NIH fellowship to S.S. (NS054540-01). We thank members of the Hobert lab for discussions.

Henry Bigelow, Maria Doitsidou, Sumeet Sarin & Oliver Hobert

Howard Hughes Medical Institute, Department of Biochemistry and Molecular Biophysics, Columbia University Medical Center, New York, New York, USA. e-mail: hbigelow@gmail.com or or38@columbia.edu

PUBLISHED ONLINE 20 JULY 2009; DOI:10.1038/NMETH.F.260

1. Sarin, S., Prabhu, S., O’Meara, M.M., Pe’er, I. & Hobert, O. *Nat. Methods* **5**, 865–867 (2008).
2. Shen, Y., Sarin, S., Liu, Y., Hobert, O. & Pe’er, I. *PLoS ONE* **3**, e4012 (2008).
3. Blumenstiel, J.P. *et al. Genetics* **182**, 25–32 (2009).
4. Li, H., Ruan, J. & Durbin, R. *Genome Res.* **18**, 1851–1858 (2008).