Classical genetics goes high-tech

David S Fay

A combination of automated screening and next-generation sequencing makes it possible to identify *Caenorhabditis elegans* mutants at unprecedented speed and scale.

Even the most ardent practitioners of classical genetics in metazoan organisms such as worms and flies will admit, in moments of weakness, that their fields have a well-deserved bad rap for harboring experimental bottlenecks. This is particularly true of the monumental efforts often involved in isolating mutants and uncovering their molecular identities. In this issue of *Nature Methods*, two papers from Hobert and colleagues outline technologically driven efficient alternatives to these rate-limiting steps in *C. elegans* forward genetics[1,2] (**Fig. 1**). Notably, these methods should in many cases be applicable to genetic approaches in other organisms.

Certain attributes of *C. elegans*, including its invariant cell lineage and optical transparency, have made it uniquely amenable to genetic screens that detect alterations in cell fate. Often such screens are carried out with the aid of a tissue- or cell type–specific *gfp* reporter that can be visualized in living animals. For example, mutants can be identified that display fewer GFP-positive cells as compared with the wild type, thereby indicating a failure to execute a normal program of cellular differentiation. Typically, such screens require mutagenized worms to be clonally picked to media plates over several generations and also entail labor-intensive visual inspection using a stereomicroscope.

Doitsidou *et al.*[1] describe an approach that bypasses both the picking and visual monitoring steps through the judicious use of a worm flow cytometer[3]. Specifically, they designed an automated screen to isolate mutations that show reduced fluorescence of a *gfp* reporter that is normally expressed in eight dopaminergic neurons[4,5]. To control for worm-to-worm variability in fluorescence intensities, which can result from developmental-stage and stochastic effects, they introduced a broadly expressed but nonoverlapping *rfp* reporter as well. In this way, they identified candidate mutations of interest based on GFP-to-RFP ratios that fell below the range observed in the parental, nonmutagenized population. This use of the *rfp* reporter should also eliminate classes of mutations that lead to a generalized reduction in transgene expression.

As part of their analysis, the authors compared the outcomes of the automated sorting screen to those obtained from a conventional manual screen carried out in parallel. Whereas the manual screen identified 10 mutants over a period of several months, the sorting screen identified 22 mutants over just several days. What is more, the sorting strategy led to the successful isolation of mutants that lacked expression of the *gfp* reporter in just one or two of the eight dopaminergic neurons, attesting to the sensitivity of the approach.

One intrinsic limitation to the sorting approach, as noted by the authors, is its inability to isolate mutations that cause worms to be inviable or sterile. In addition, mutations that lead to more qualitative changes, such as alterations in cell position, shape or organization, would likely be missed. Moreover, screens of this nature provide, at best, an indirect measurement of cellular function; a cell may correctly express a differentiation marker but lack normal function. Nevertheless, the estimated seven-or-greater-fold improvement in the overall efficiency of mutant isolation constitutes a considerable advantage to manual methods in cases in which viable and quantifiable phenotypes are anticipated. Furthermore, the sorting approach could likely be adapted to other types of studies such as screens for mutants that are defective in their response to specific environmental stresses, through the use of stress-activated *gfp* reporters.

Not merely content to isolate their mutants at relative light speed, the Hobert group, in a second paper, outlines a scheme for identifying molecular lesions in just a fraction of the time required using traditional methods[2]. Historically, molecular identification has been the single most protracted step in metazoan forward genetics. The current process in *C. elegans* typically entails an extensive analysis of recombinants, using genetic and polymorphism-based markers, and can take anywhere from months to years to complete. Definitive proof that one has identified the relevant gene requires, among other things, the detection of a molecular lesion within the implicated locus.

To identify the relevant gene altered in *C. elegans lsy-12* mutants, Sarin *et al.*[2] cut directly to the chase by using a rapid whole-genome resequencing approach. *lsy-12* mutants are defective at producing an asymmetric pair of left and right ASE chemosensory neurons[6]. After preliminary mapping that narrowed the genomic region encompassing *lsy-12* to a 4-Mb interval containing 1,142 genes (5% of all genes in the worm), they sequenced DNA from the mutant strain using paired-end Illumina technology[7]. A 1-week run using a single machine produced 4.35 gigabasses of sequence, which was then mapped to the published genomic sequence, resulting in an average coverage of ~28×. They then compared sequences from the mutant strain to the wild-type reference genome and analyzed 80 candidate variations that fell within the mapped region.

Traditional Sanger sequencing carried out in parallel revealed that 16 of the 80 candidates were the result of errors in the Illumina sequencing; these could most probably have

David S. Fay is in the department of Molecular Biology, College of Agriculture, University of Wyoming, Laramie, Wyoming 82071, USA.
e-mail: davidfay@uwyo.edu

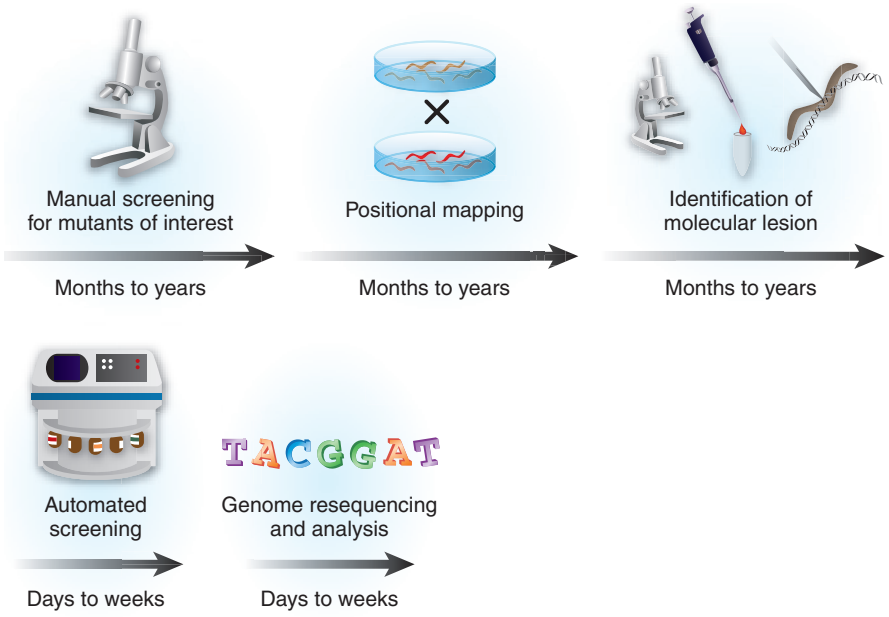

Figure 1 | New technologies for classical genetics in *C. elegans*. Traditional forward genetics in *C. elegans* (top) includes time- and labor-intensive steps in the isolation and identification of mutants. With the application of automated sorting and next-generation sequencing (bottom), screening time can be substantially reduced.

been eliminated *a priori* based on the quality of the reads in the initial dataset. Notably, more than half of the confirmed variants were present in the starting strain and could therefore not be responsible for the mutant phenotype. Because most known mutations in *C. elegans* lead to sequence changes in the corresponding peptide products[5,8], Sarin *et al.*[2] focused on the four nonsilent exonic variants that were unique to the *lsy-12* mutant background. Functional analysis of the corresponding four genes by RNA interference, along with sequencing of these genes in additional *lsy-12* alleles, identified a single uncharacterized locus, R07B5.9, as *lsy-12*.

How generally applicable will whole-genome sequencing be for identifying causal mutations in *C. elegans* and in other species? One practical issue for many investigators, at least in the short term, will revolve around the substantial financial cost of the undertaking, which incurs if in-house sequencers are unavailable. This issue is also related to the question of how much sequencing coverage is minimally necessary to consistently identify mutations of interest. Sarin *et al.*[2] address this point through a statistical analy-sis showing that for an average 4-Mb interval, as little as 0.8 gigabasses of sequence (eightfold coverage) is likely to be required.

As with the automated sorting screen, whole-genome sequencing analysis may be more difficult to execute if homozygous mutant strains are inviable. In addition, as pointed out by the authors, the follow-up analysis is greatly aided by the availability of multiple alleles, each representing an independent mutation at the locus. Furthermore, some preliminary mapping of the locus is clearly advantageous and possibly even necessary for this approach to be effective. Those issues aside, direct genome sequencing offers obvious advantages for the identification of mutants with subtle or weakly penetrant phenotypes or where complex genetic backgrounds are necessary to observe phenotypic effects. Moreover, mutants that have been recalcitrant to molecular identification by standard approaches may now be identified. Finally, as this approach can drastically reduce or even eliminate laborious mapping steps, it promises to shorten the time required to identify mutations in *C. elegans* by an order (or more) of magnitude.

Of course, *C. elegans* is not the only genetic system that will benefit from high-throughput sequencing methods. Flies, frogs, fish and mice, as well as any other creature that boasts an assembled genome, will also be amenable to this approach. In fact, it could be argued that given the historic and inherent advantages of *C. elegans* for use in classical genetics, other systems stand to gain relatively more from this advance. This could substantially level the playing field and may usher in a golden era of genetics in a diverse array of species.

At the very least, by minimizing the time required to identify *C. elegans* mutations, whole-genome sequencing should render the pain associated with relegating months or years of labor to a paltry few sentences in a published paper a thing of the past.

1. Doitsidou, M., Flames, N., Lee, A.C., Boyanov, A. & Hobert, O. *Nat. Methods* **5**, 869–872 (2008).
2. Sarin, S., Prabhu, S., O'Meara, M.M., Pe'er, I. & Hobert, O. *Nat. Methods* **5**, 865–867 (2008).
3. Pulak, R. *Methods Mol. Biol.* **351**, 275–286 (2006).
4. Nass, R. *et al. J. Neurochem.* **94**, 774–785 (2005).
5. Nass, R., Hall, D.H., Milerr, D.M., III & Balekely, R.D. *Proc. Natl. Acad. Sci. USA* **99**, 3264–3269 (2002).
6. Sarin, S. *et al. Genetics* **176**, 2109–2130 (2007).
7. Davis, M.W. *et al. BMC Genomics* **6**, 118 (2005).
8. Anderson, P. *Methods Cell Biol.* **48** 31–58 (1995).