Plasmodium. They thus present numerous evolutionary novelties, including differences in fundamental cell biology (their lack of an organelle called the apicoplast is one example), in their infectious cycle, and in their genome, which at around 3,950 genes is much smaller than that of other apicomplexans[5–7]. The *Cryptosporidium* genome contains several essential genes acquired by lateral transfer from other microorganisms[5–7], which perhaps reflects the parasite's intimacy with intestinal bacteria. Collectively, these features provide exciting opportunities for basic research as well as for identifying cellular pathways relevant to therapy — but both these tasks have been made difficult by a lack of genetic tools.

The true challenge, however, was not the molecular technology but the limitations of working with *Cryptosporidium*, which cannot be cultured long term *in vitro*. Instead, oocysts must be isolated from infected calves or purchased commercially. Cysts can be stored for months, but excysted parasites that are inoculated onto mammalian-cell monolayers for growth undergo one or two rounds of replication at most. This narrow time window has profoundly hindered experimental manipulation[2].

Vinayak *et al.*[1] have dramatically improved this state of affairs. They made a series of optimizations to existing genetic-modification techniques that establish the basic parameters for successful transient transfection of *Cryptosporidium* sporozoites. This procedure introduces a segment of DNA (in this case, a plasmid) encoding a gene of interest that is then expressed by the cell for a short time. The authors verified successful transfections using a marker gene that encodes the protein luciferase, which produces bioluminescence in the presence of the appropriate substrate. This marker is fused to a gene conferring resistance to neomycin-class antibiotics, which provides a means of selecting transfected cells.

Not content with achieving reproducible transient transfection, Vinayak *et al.* proceeded to overcome the narrow experimental window. During *in vitro* culture, *Cryptosporidium* does not generate the thick-walled cyst forms that survive in the faeces and the stomach, but the researchers bypassed this biological block by inoculating the manipulated sporozoites directly back into the intestines of immunodeficient mice, in which the parasites propagated and produced oocysts (Fig. 1).

For stable genomic modifications, in which the introduced DNA is incorporated into the genome, rather than relying on the parasite's own mechanisms for doing this, the authors turned to the genetic '*tool de jour*' — the CRISPR/Cas9 system, a genome-editing approach that has proved effective in almost all organisms tested, including protozoan parasites. Another series of clever optimizations established the functionality and utility of this system in *Cryptosporidium*. Eventually, transfection of sporozoites with both the luciferase–neomycin-resistance fusion gene and DNA encoding the CRISPR/Cas9 machinery, followed by infection of mice with the sporozoites and treatment with the neomycin analogue paromomycin, led to the recovery from mouse faeces of antibiotic-resistant parasites stably expressing an integrated luciferase gene.

This first demonstration of genetically engineered *Cryptosporidium* introduces a method that is primed for real-world applications, already enabling *in vitro* or *in vivo* assays for monitoring parasite survival after drug or other treatments. The authors further demonstrated the utility of CRISPR/Cas9 by using it in the sporozoites to ablate expression of thymidine kinase, one of the few enzymes used by *Cryptosporidium* to generate nucleotides[8]. These experiments showed that this enzyme's activity provides a bypass for the activity of another enzyme, dihydrofolate reductase, which accounts for the relative ineffectiveness of antifolate drugs against *Cryptosporidium* compared with other apicomplexan parasites.

The success of Vinayak and colleagues' study lies not so much in the novelty or insight of particular steps, but rather in the systematic and incisive integration of them all towards what had been considered an impossible goal. As such, this is a textbook study on how to tackle a previously intractable pathogen, and it will serve as a model for future attempts with other disease-causing organisms.

The approach is by no means perfect — it is cumbersome and time-consuming to generate genetically modified cell lines by passaging them through mice, and the parasites can be studied only following recovery of cysts from faeces. But one can imagine many advances and future directions, such as using CRISPR-based systems to generate and probe panels of mutated parasites simultaneously. Perhaps high on the list of priorities will be the generation of modified parasites that can replicate and differentiate indefinitely *in vitro*. A second challenge is that genes required for parasite survival inside host cells cannot be ablated in order to study their mechanism; however, the importation of RNA- or protein-based regulatory strategies from other apicomplexans should overcome this.

So, having found how to 'get there', the application of *Cryptosporidium* genetic modification will greatly increase our understanding of the pathogen's basic biology and virulence, and provide key information and validation for the development of improved vaccines and therapeutics. ■

Stephen M. Beverley *is in the Department of Molecular Microbiology, Washington University School of Medicine, St Louis, Missouri 63110, USA.*
e-mail: beverley@wusm.wustl.edu

1. Vinayak, S. *et al. Nature* **523,** 447–480 (2015).
2. Checkley, W. *et al. Lancet Infect. Dis.* **15,** 85–94 (2015).
3. Striepen, B. *Nature* **503,** 189–191 (2013).
4. MacKenzie, W. R. *et al. N. Engl. J. Med.* **331,** 161–167 (1994).
5. Abrahamsen, M. S. *et al. Science* **304,** 441–445 (2004).
6. Bouzid, M., Hunter, P. R., Chalmers, R. M. & Tyler, K. M. *Clin. Microbiol. Rev.* **26,** 115–134 (2013).
7. Xu, P. *et al. Nature* **431,** 1107–1112 (2004).
8. Sun, X. E. *et al. J. Biol. Chem.* **285,** 15916–15922 (2010).

This article was published online on 15 July 2015.

GENETICS

Feedforward loop for diversity

DNA-sequence analysis suggests that genetic mutations arise at elevated rates in genomes harbouring high levels of heterozygosity — the state in which the two copies of a genetic region contain sequence differences. SEE LETTER P.463

MICHAEL LYNCH

The rate at which genetic mutations arise is relevant to every area of biology. Evidence indicates that mutation rates vary almost 1,000-fold between species, from 10^{-11} mutations per nucleotide site per generation in some unicellular organisms to approximately 10^{-8} in primates[1]. These figures represent genome-wide averages, but mutation rates can vary between nucleotide sites[2–4] and between members of the same species[5]. Intraspecies differences have long been assumed to be a consequence of genetic variation at discrete regions, or loci, containing genes involved in genome-wide aspects of DNA replication and repair. But on page 463 of this issue, Yang *et al.*[6] suggest something quite different: that mutation rates are elevated in individuals with high genome-wide levels of heterozygosity (sequence variation between the two copies, called alleles, of each genetic locus).

Yang and colleagues' gold-standard analyses compared whole-genome sequences of parents and offspring for two plants and an insect. They found that mutation rates are elevated in individuals with higher overall heterozygosity, particularly in regions close to heterozygous sites and regions in which there are high rates of DNA exchange between chromosomes (recombination). The authors therefore propose a positive-feedback loop, whereby high levels of molecular variation in an individual facilitate the production of more variation.

It is accepted that recombination is mutagenic[7], but the implications of Yang and coworkers' results for population-level genetic analyses, which rely on measures of heterozygosity, could be substantial. For example, average levels of variation are often assumed to directly reflect recent population sizes — independent of the mutation rate — because large population sizes enhance the maintenance of variation. But such an assumption is compromised if a transient boost in heterozygosity, for whatever reason, also boosts the rate of mutational production of variation. Furthermore, a feedforward effect might help to explain the clustering of variation at adjacent sites[8], which may in turn relate to the fact that closely spaced sites have elevated levels of linkage disequilibrium (a measure of the statistical association between specific alleles at different genetic loci)[9].

Some forms of natural selection that favour the maintenance of variation — for example, to promote avoidance of specialized pathogens — might also be associated with elevated mutation rates[10]. As Yang and colleagues note, their results bear on this controversial idea. Whether natural selection is efficient enough to modulate gene-specific mutation rates is questionable[11]. But if loci under diversifying selection (which favours variation) passively acquire elevated mutation rates as variation grows, gene-specific modifiers of the mutation rate need not be invoked to explain this model.

Although the authors' results concerning the mutagenic effect of heterozygosity are surprising, the mutation rate that they calculate for inbred strains of the plant *Arabidopsis* is not greatly different from that reported previously[12], so the results do not seem to be artefactual. But what biological peculiarities could elevate mutation rates in heterozygotes? Much goes wrong in inbred organisms owing to an increase in homozygosity (in which the two alleles of a gene are identical), which increases the exposure of an organism to deleterious 'recessive' alleles[13]. One might therefore expect the mutation rate to be higher in inbred than outcrossed individuals — the opposite pattern to that observed by Yang and colleagues. However, outcrossing between distantly related strains can sometimes lead to outbreeding depression, in which offspring have lower fitness than those from intra-strain crosses.

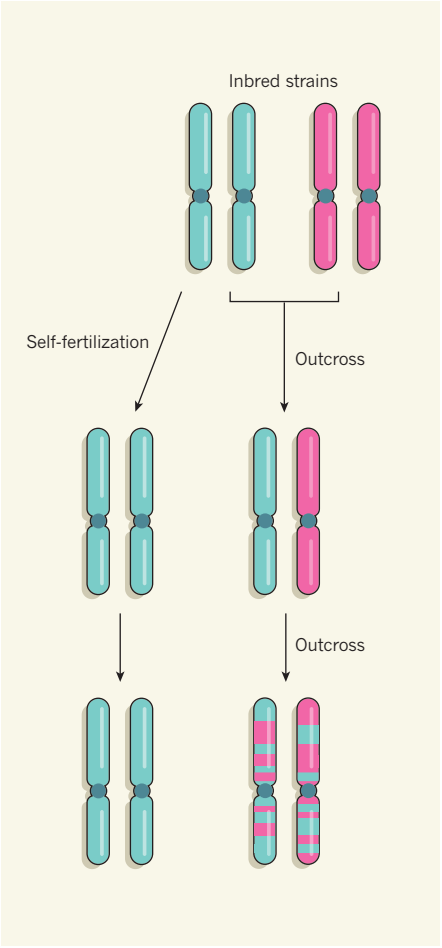

Figure 1 | Generating variation. This simplified schematic demonstrates the changes in diversity that arise in intercrosses of a diploid organism, which has two sets of chromosomes, one from each parent. In inbred organisms, most genetic regions are homozygous — they are identical on both chromosomes (completely homozygous chromosomes are depicted here for simplicity). When inbred plants self-fertilize, levels of homozygosity remain the same in offspring. But in the first generation of a cross between two inbred strains, the offspring have two different copies of each gene (heterozygosity). Further intercrossing of offspring leads to a decrease in levels of heterozygosity, because some regions become homozygous once again. Yang *et al.*[6] report that levels of heterozygosity correlate with the rate at which genetic mutations arise.

The parental strains used in this study might have been divergent enough to generate incompatibilities that influence the mutation rate. For instance, many proteins involved in DNA replication and damage repair operate as multimeric complexes, and the mixture of subunits from divergent strains might lead to malfunctioning complexes. Physiological effects on a cellular level, such as the production of free radicals that damage DNA, might also be a factor.

One argument against the involvement of outbreeding depression is the authors' observation that mutation rates are not uniformly elevated across the genomes of first-generation offspring from outcrossing, but are concentrated near heterozygous sites. However, the elevation in mutation rate near heterozygous sites is less than twofold, and an outbreeding-depression effect cannot be entirely ruled out. For example, when a heterozygous site is part of a locus that is involved in a recombination event, the 'mismatch-repair' pathway used to resolve the difference at the site also engages with the surrounding DNA. Because this pathway is relatively error-prone[14], if the repair complex is made up of a mixture of subunits from the different parents, this could specifically elevate the mutation rate near heterozygous sites.

The authors show that mutation rates decline in the third and fourth generation after outcrossing, consistent with expectations based on the associated decline in heterozygosity, but care must be taken with this interpretation. Immediately after outcrossing, each gene has an allele from each parental line, whereas in later descendent generations, offspring tend towards 50% mixtures of homozygous and heterozygous allele complements (Fig. 1). It then becomes difficult to determine whether a reduction in mutation rate is a direct consequence of the decline in heterozygosity, or whether changes in outbreeding depression or in its counterpart, outbreeding enhancement, are partially or wholly responsible[13].

It should be straightforward to test whether heterozygosity per se is a direct determinant of the mutation rate by focusing on species such as the honeybee, in which males contain only a set of chromosomes inherited from their mothers — if the authors' hypothesis is correct, mutation rates should be lower in males than in their heterozygous sisters. Moreover, if recombination magnifies the mutation rate, rates should be reduced on chromosomes that cannot recombine, such as the X and Y of human males and all the chromosomes of male fruit flies.

Under the authors' proposed scenario, might runaway magnification of both the mutation rate and population-level heterozygosity be possible? This would seem to require a rather implausible set of conditions, but there are reports of extraordinarily high levels of heterozygosity in organisms such as the urochordate *Ciona savignyi*[15] and the nematode *Caenorhabditis brenneri*[16]. Whether these taxa actually reflect stable alternative states of heterozygosity could be answered by evaluating whether individuals engineered to be more homozygous show reduced mutation rates.

Finally, it is worth considering how the approximately 3.5-fold difference in mutation rate between inbred and outbred strains found in the current study compares with variation among individuals in normal populations. The mutation rates in two inbred lines of fruit fly

differ by around 2.3-fold[5], and these rates are slightly higher than those of outbred flies[17]. Self-fertilizing organisms with exceptionally low heterozygosity do not have unusually low mutation rates compared with outcrossing species with similar genome sizes[1]. Furthermore, humans and chimpanzees, which are highly homozygous, have extremely high mutation rates[1,18]. Of course, there are many biological differences between these species, so caution must be taken not to overinterpret these observations.

Overall, this study raises several intriguing questions. Even if the results are eventually found to reflect outbreeding depression or simply natural variation in replication fidelity, Yang and colleagues have done us a service, encouraging a focus on variation in the process that itself generates variation. ■

Michael Lynch *is in the Department of Biology, Indiana University, Bloomington, Indiana 47405, USA.*
e-mail: milynch@indiana.edu

1. Sung, W. *et al. Proc. Natl Acad. Sci. USA* **109,** 19339–19344 (2012).
2. Lynch, M. *Proc. Natl Acad. Sci. USA* **107,** 961–968 (2010).
3. Foster, P. L. *Genes Genomes Genet.* **3,** 399–407 (2013).
4. Sung, W. *et al. Mol. Biol. Evol.* **32,** 1672–1683 (2015).
5. Schrider, D., Houle, D., Lynch, M. & Hahn, M. *Genetics* **194,** 937–954 (2013).
6. Yang, S. *et al. Nature* **523,** 463–467 (2015).
7. Arbeithuber, B., Betancourt, A. J., Ebner, T. & Tiemann-Boege, I. *Proc. Natl Acad. Sci. USA* **112,** 2109–2114 (2015).
8. Harris, K. & Nielsen, R. *Genome Res.* **24,** 1445–1454 (2014).
9. Lynch, M. *et al. Genetics* **198,** 269–281 (2014).
10. Amos, W. *BioEssays* **32,** 82–90 (2010).
11. Chen, X. & Zhang, J. *Mol. Biol. Evol.* **30,** 1559–1562 (2013).
12. Ossowski, S. *et al. Science* **327,** 92–94 (2010).
13. Lynch, M. & Walsh, J. B. *Genetics and Analysis of Quantitative Traits* (Sinauer, 1998).
14. Lynch, M. *Genome Biol. Evol.* **3,** 1107–1118 (2011).
15. Small, K. S., Brudno, M., Hill, M. M. & Sidow A. *Proc. Natl Acad. Sci. USA* **104,** 5698–5703 (2007).
16. Dey, A., Chan, C. K., Thomas, C. G. & Cutter, A. D. *Proc. Natl Acad. Sci. USA* **110,** 11056–11060 (2013).
17. Keightley, P. D., Ness, R. W., Halligan, D. L. & Haddrill, P. R. *Genetics* **196,** 313–320 (2014).
18. Venn, O. *et al. Science* **344,** 1272–1275 (2014).

This article was published online on 15 July 2015.

COMPUTATIONAL IMAGING

Machine learning for 3D microscopy

Artificial neural networks have been combined with microscopy to visualize the 3D structure of biological cells. This could lead to solutions for difficult imaging problems, such as the multiple scattering of light.

LAURA WALLER & LEI TIAN

How can researchers see inside an object without using invasive techniques, or recover 3D information by capturing only 2D images? This question was answered decades ago with the invention of tomography — a technique that computationally reconstructs 3D objects from a set of 2D images, usually captured from a range of projection angles. Tomography, which is used in magnetic resonance imaging and computerized tomography scanners for medical and other applications, conventionally provides an analytical solution to the 3D reconstruction problem. However, as the use of tomography expands to applications that involve complex scenarios, it is not always possible, or desirable, to devise analytical solutions. Now, machine-learning methods are turning optical tomography on its head with the use of algorithms borrowed from data science, which reconstruct the 3D refractive index of an object by solving a large-scale optimization problem. Writing in *Optica*, Kamilov *et al.*[1] demonstrate this experimentally using a holographic optical-phase microscope.

Tomography is the quintessential example of computational imaging, a discipline that transcends conventional imaging techniques by simultaneously designing both the optical system and the image-processing algorithms. Together, the optics and the algorithms can achieve things that neither could do alone. For example, Kamilov *et al.* recover the 3D 'phase' of a biological cell — the nanometre-scale distortions of a wavefront as it passes through an object — thus rendering transparent objects visible.

Kamilov *et al.* use machine-learning algorithms — computer programs that can learn from and make predictions based on input data — to give a boost to 3D phase imaging. By doing so, the authors bridge the fields of

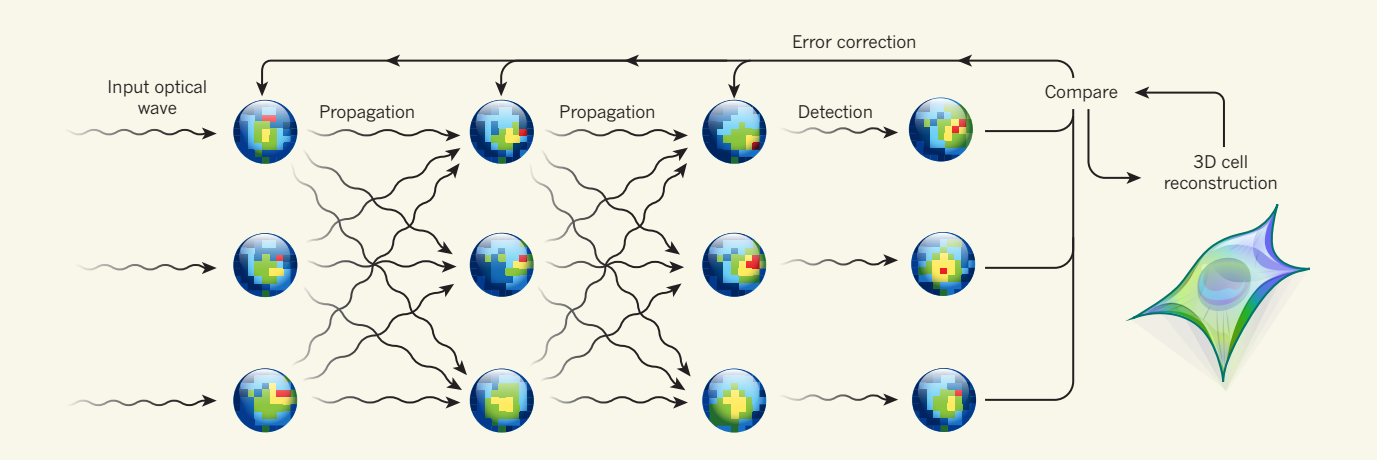

Figure 1 | 3D image reconstruction with artificial neural networks. Kamilov *et al.*[1] use an artificial neural network (ANN) algorithm to describe how the phase of optical light is modified as it propagates through a 3D biological sample (here, a cell). The sample is modelled as a series of layers. Each pixel (circles) of the 3D model corresponds to a node of the ANN. These are connected to nodes in the subsequent layer (arrows) to represent the scattering of the input wavefront in the direction of propagation. The algorithm incorporates error correction by comparing a detector's measurements with the model's output — a 3D reconstruction of a cell's refractive index — and minimizing the difference between the two. (Figure adapted from Fig. 2 of the paper[1].)