

Lessons in the genomic diversity of malaria

Genome-wide characterization of *Plasmodium falciparum* genetic diversity represents an important step in malaria genetics. The organizational and collaborative efforts to reach this goal and the results on linkage disequilibrium and genetic diversity among worldwide isolates provide general lessons.

The publication of the *Plasmodium falciparum* genome sequence (*Nature* **419**, 498–511; 2002) represented a landmark in the study of this eukaryotic human pathogen, accompanied by the sequence of the model rodent parasite *Plasmodium yoelii* (*Nature* **419**, 512–519; 2002) and the mosquito vector *Anopheles gambiae* (*Science* **298**, 129–149; 2002). The sequence of the less deadly human species *Plasmodium vivax* was also recently completed (*Trends Parasitol.* **19**, 227–231; 2003). These, together with a range of transcriptional and proteomic *Plasmodium* data sets, have provided valuable resources to the malaria community.

One recent opportunity to discuss the use of these resources and set priorities came at the ‘Next Steps in Malaria Research’ meeting held in April 2005 (*Trends Parasitol.* **22**, 1–4; 2006). There, representatives agreed on the importance of high-throughput approaches, combined with broad data release policies. Central goals for malaria research remain, including characterizing the factors involved in pathogenesis, virulence, transmission and drug resistance, in order to better understand the biology of the parasite and aid in defining drug and vaccine targets.

Several leading malaria genomics groups and sequencing centers have coordinated their efforts once again to characterize the genome-wide genetic diversity of *Plasmodium* (see News & Views by Jane Carlton, p. 5). These efforts exemplify progress in infectious disease genomics and provide several lessons for the broader genetics community.

The first lesson is the reward of international collaborations and study replication. Large-scale genomic studies are often found in pairs (such as the public and private human genome sequences, or the first phase of the International HapMap along with Perlegen Science’s haplotype map), with independent efforts providing momentum as well as replication. Ultimately, reconciliation of the results may be achieved by communication and collaboration among the research centers, as when Perlegen joined the International HapMap Project, or in the recent example of the MHC HapMap (*Nat. Genet.* **38**, 1166–1172; 2006).

Similarly, the complementary studies presented here make great progress in defining the genome-wide genetic diversity of *P. falciparum* and in discovering SNPs and microsatellite polymorphisms. Dyann Wirth and colleagues provide breadth and depth of sequencing, including partial genome sequencing of 16 new and geographically diverse isolates as well as targeted sequencing of an additional 54 worldwide isolates, and they provide insights into linkage dis-

equilibrium (LD) and global variation (Volkman *et al.*, p. 113). Xin-zhuan Su, Philip Awadalla and colleagues (Mu *et al.*, p. 126) focus on coding regions, manually sequencing some 3,500 genes in four *P. falciparum* isolates, representing about 19% of the genome, and they report new candidate targets for vaccines. In the third study, Manolis Dermitzakis, Matthew Berriman and colleagues take a comparative genomics and evolution approach, providing the first sequence of the sister species *P. reichenowi* (Jeffares *et al.*, p. 120) and of two *P. falciparum* isolates, including a new uncultured clinical isolate from Ghana, which may provide a better model than clones cultured in the laboratory.

A second lesson is the benefit of immediate data release. The community has developed its own sequence and reagent databases. The first is PlasmoDB (<http://www.plasmodb.org>), which was designed to host genome sequence information on *Plasmodium* and closely related species and accepts submission from both small and large-scale projects. Another resource, MR4 (the Malaria Research and Reference Reagent Resource Center), provides access to biological resources, including parasite samples and reagents (<http://www.mr4.org>).

The relatively high degree of variation, particularly in the African samples, makes this parasite remarkable. SNPs and small indels were found to make a similar contribution to sequence diversity, paralleling patterns in maize. While it becomes straightforward to identify regions that show long-range LD patterns (suggestive of positive natural selection) among the otherwise short-range LD observed, these studies may not initially provide a ‘malaria HapMap’ with which to facilitate association studies for the parasite, particularly in Africa. Their approach is still valid, but the studies suggest that a dense coverage of the SNPs discovered will be required in order to tag causal loci associated with phenotypes, such as virulence, transmissibility and drug resistance.

The final lesson from these studies is the narrowing of the gap between whole-genome genotyping and whole-genome sequencing. In one recent example from our pages, Bernhard Palsson and colleagues conducted whole-genome resequencing of *E. coli* strains undergoing selection, identifying mutations that became fixed during gradual adaptation (*Nat. Genet.* **38**, 1406–1412; 2006). While comparable whole-genome sequencing studies of large numbers of individual parasite isolates from the field would also be one of the next desirable steps for the malaria community, the foundations laid in the current set of papers reinforce the importance of cooperative efforts between large-scale and targeted sequencing. ■