GENOME WATCH

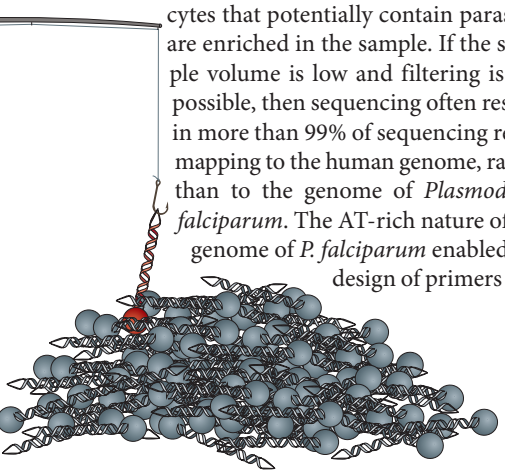
Finding the needle in the haystack

Gavin G. Rutledge and Cristina V. Ariani

This month's Genome Watch discusses the potential of selective whole-genome amplification for overcoming the challenges of whole-genome sequencing of malaria parasites in clinical samples in which they are low in abundance.

One of the main obstacles for carrying out whole-genome sequencing (WGS) of bacteria and protozoan parasites from the field is obtaining high-quality samples. As culturing these organisms may not be possible, large volumes of sample are often needed to overcome host contamination, especially for organisms that exist at low densities. Selective whole-genome amplification (sWGA) enables the selective enrichment of target DNA from a contaminating background and thus facilitates WGS of samples that would otherwise contain insufficient pure genomic DNA[1].

To amplify the target DNA, sWGA uses custom primers that are designed to bind to the target DNA more frequently than to contaminating DNA[1]. Originally developed for sequencing the genomes of bacteria, the method has been rapidly adopted by the malaria research community. Traditionally, sequencing clinical malaria samples required the filtration of large volumes of blood from infected individuals to remove contaminating uninfected white blood cells, so that erythrocytes that potentially contain parasites are enriched in the sample. If the sample volume is low and filtering is not possible, then sequencing often results in more than 99% of sequencing reads mapping to the human genome, rather than to the genome of *Plasmodium falciparum*. The AT-rich nature of the genome of *P. falciparum* enabled the design of primers that bind to the *P. falciparum* genome more frequently than to the human genome, which is GC neutral. Using this approach, more than 70% of reads could be mapped to the genome of *P. falciparum*, even when the sample volume was low[2].



Philip Patenall/Macmillan Publishers Limited

Recent studies using sWGA have demonstrated the potential of this new technology in WGS of clinical samples. Previously, it was not possible to extract sufficient *P. falciparum* DNA from dried blood spots (DBS) to carry out WGS, owing to the low sample volume and high levels of contamination with DNA from the host. However, a recent study[3] amplified the parasite DNA that was obtained from DBS of patients through the use of sWGA, and subsequent WGS revealed that drug resistance is an unlikely explanation for the recurrences of malaria that were observed in their clinical trial, and that these recurrences were probably due to the inherent properties of the antimalarial used in the trial[3]. The sequencing of DNA from DBS is an important milestone in monitoring malaria genetics and epidemiology, because it has enabled the acquisition of sequencing data from clinical samples, even when the sample volume is very low and the quality is limited. WGS of *P. falciparum* from DBS using sWGA has recently been extensively tested and standardized[2], and this integrative approach of using sWGA and WGS is likely to be used more frequently in epidemiological studies in the future.

Demonstrating the applications of sWGA, two studies recently used this method to investigate the evolutionary relationship between *Plasmodium* species from chimpanzees and their human counterparts[4,5]. Owing to very low parasite levels in the blood of the chimpanzees, traditional WGS of the parasite genome was not possible. Using the selective amplification strategy, the authors were able to increase the ratio of parasite to chimpanzee DNA by more than 100-fold, thereby enabling the assembly of the genomes of the different *Plasmodium* species that were identified. Moreover, the data revealed the presence of the *var* gene family, which is thought to be involved in immune evasion in other *Plasmodium* species[4] and an incidence of genetic exchange of a blood cell invasion gene from a chimpanzee-infective ancestor to the ancestor of *P. falciparum*[5]. However, the latter study also highlighted the limitations of using sWGA when genomes are not uniform in their GC-content distribution. For example, *Plasmodium* species have a substantially higher GC content in regions that are close to the telomeres than in the core regions, and, as the primers are designed for AT-rich regions, only the core regions of the genomes could be sequenced adequately[2,5].

Despite the limitations of sWGA, this new technology overcomes major obstacles that the collection and WGS of clinical malaria samples pose. The potential uses for sWGA are far-reaching; its simplicity and the benefits of sequencing low-abundance DNA in clinical samples that have high levels of host contamination will help aid future studies of parasites and other pathogens.

Gavin G. Rutledge and Cristina V. Ariani are at the Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK.
e-mail: microbes@sanger.ac.uk

doi:10.1038/nrmicro.2017.7
Published online 13 Feb 2017

1. Leichty, A. R. & Brisson, D. Selective whole genome amplification for resequencing target microbial species from complex natural samples. *Genetics* **198**, 473–481 (2014).
2. Oyola, S. O. *et al.* Whole genome sequencing of *Plasmodium falciparum* from dried blood spots using selective whole genome amplification. *Malar. J.* **15**, 597 (2016).
3. Guggisberg, A. M. *et al.* Whole-genome sequencing to evaluate the resistance landscape following antimalarial treatment failure with fosmidomycin-clindamycin. *J. Infect. Dis.* **214**, 1085–1091 (2016).
4. Larremore, D. B. *et al.* Ape parasite origins of human malaria virulence genes. *Nat. Commun.* **6**, 8368 (2015).
5. Sundararaman, S. A. *et al.* Genomes of cryptic chimpanzee *Plasmodium* species reveal key evolutionary events leading to human malaria. *Nat. Commun.* **7**, 11078 (2016).

Competing interests statement
The authors declare no competing interests.