# Plasmodium knowlesi Genome Sequences from Clinical Isolates Reveal Extensive Genomic Dimorphism

Miguel M. Pinheiro[1☯], Md Atique Ahmed[2☯], Scott B. Millar[1], Theo Sanderson[3], Thomas D. Otto[3], Woon Chan Lu[4], Sanjeev Krishna[5], Julian C. Rayner[3], Janet Cox-Singh[1]*

1 School of Medicine, University of St Andrews, Medical and Biological Sciences Building, North Haugh, St Andrews, United Kingdom, 2 Malaria Research Centre, University Malaysia Sarawak, Kuching, Sarawak, Malaysia, 3 Malaria Programme, Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, United Kingdom, 4 Sarikei Hospital, Sarikei, Sarawak, 96100, Malaysia, 5 Division of Clinical Sciences, St. George's, University of London, London, United Kingdom

☯ These authors contributed equally to this work.
* jcs26@st-andrews.ac.uk

## Abstract

*Plasmodium knowlesi* is a newly described zoonosis that causes malaria in the human population that can be severe and fatal. The study of *P. knowlesi* parasites from human clinical isolates is relatively new and, in order to obtain maximum information from patient sample collections, we explored the possibility of generating *P. knowlesi* genome sequences from archived clinical isolates. Our patient sample collection consisted of frozen whole blood samples that contained excessive human DNA contamination and, in that form, were not suitable for parasite genome sequencing. We developed a method to reduce the amount of human DNA in the thawed blood samples in preparation for high throughput parasite genome sequencing using Illumina HiSeq and MiSeq sequencing platforms. Seven of fifteen samples processed had sufficiently pure *P. knowlesi* DNA for whole genome sequencing. The reads were mapped to the *P. knowlesi* H strain reference genome and an average mapping of 90% was obtained. Genes with low coverage were removed leaving 4623 genes for subsequent analyses. Previously we identified a DNA sequence dimorphism on a small fragment of the *P. knowlesi normocyte binding protein xa* gene on chromosome 14. We used the genome data to assemble full-length *Pknbpxa* sequences and discovered that the dimorphism extended along the gene. An in-house algorithm was developed to detect SNP sites co-associating with the dimorphism. More than half of the *P. knowlesi* genome was dimorphic, involving genes on all chromosomes and suggesting that two distinct types of *P. knowlesi* infect the human population in Sarawak, Malaysian Borneo. We use *P. knowlesi* clinical samples to demonstrate that *Plasmodium* DNA from archived patient samples can produce high quality genome data. We show that analyses, of even small numbers of difficult clinical malaria isolates, can generate comprehensive genomic information that will improve our understanding of malaria parasite diversity and pathobiology.