Alistair Miles (corresponding author)
MRC Centre for Genomics and Global Health
Wellcome Trust Centre for Human Genetics
Roosevelt Drive
Oxford, OX3 7BN, UK
Email: alistair.miles@well.ox.ac.uk
Tel: +44 1865 287721
Fax: +44 1865 287501

Dominic Kwiatkowski
Wellcome Trust Sanger Institute
Hinxton, CB10 1SA

Hillary E. Sussman, Ph.D.
Executive Editor, Genome Research
Cold Spring Harbor Laboratory Press
One Bungtown Road
Cold Spring Harbor, NY 11724

22 December 2015

Dear Dr Sussman,

We would be grateful if you would consider this manuscript on genome variation and meiotic recombination in *Plasmodium falciparum* for publication in Genome Research.

P. falciparum is the most virulent species of malaria parasite and causes a massive global burden of disease. Understanding its genetic diversity is of practical relevance for fighting drug resistance and vaccine development. Although thousands of parasite isolates from around the world have now been sequenced, there remain many fundamental gaps in our knowledge, since the *P. falciparum* genome is much more difficult to analyse than most other species. This is a consequence of its remarkably high AT content (90% in non-coding regions), highly repetitive sequences and hypervariable genes that prevent accurate alignment to the reference genome.

In this study we have addressed the problem by sequencing the parents and progeny of three genetic crosses of *P. falciparum* performed in the laboratory. This rich dataset (with 15-35 progeny per cross) allows us to calculate Mendelian error rates, and thus to evaluate and optimise methods of variant discovery and genotype calling. This has allowed us to analyse the architecture of variation in the core *P. falciparum* genome to a much higher level of resolution and accuracy than has hitherto been possible, revealing an exceptionally high density of INDELs and tightly localised regions of intense SNP polymorphism.

This dataset also allows a detailed analysis of sexual recombination in this species, distinguishing crossover and non-crossover events, and showing how recombination can modify amplifications spanning known drug resistance genes. Taken together these data provide a reference resource for deeper studies of evolutionary processes in natural parasite populations.

The manuscript has been approved by all authors. Potential referees include Drs. Ken Vernick (Institut Pasteur, kenneth.vernick@pasteur.fr), Dan Neafsey (Broad Institute, neafsey@broadinstitute.org), Pradip Rathod (University of Washington, rathod@chem.washington.edu), Tim Anderson (Texas Biomedical, tanderso@txbiomedgenetics.org) and Mike Eberle (Illumina, meberle@illumina.com). Gene names used are *msp1*, *msp3, msp6, dblmsp, dblmsp2, ama1, surf1.2, surf4.1, surf4.2, surf8.2, surf13.1, surf14.1,* PF3D7_0113800, PF3D7_0104100, *mdr1* and *gch1*. All sequence read data have been submitted to the European Nucleotide Archive. All genome variation data are available from a public FTP site hosted by the Wellcome Trust Sanger Institute. Further details of data released are available from http://www.malariagen.net/data/pf-crosses-1.0.

Yours sincerely,

Alistair Miles and Dominic Kwiatkowski