

# PacBio assembly of a *Plasmodium knowlesi* genome sequence with Hi-C correction and manual annotation of the SICAvAr gene family

Juliana Assis<sup>1</sup>

*1 UFMG*

## Abstract

*Plasmodium knowlesi* has risen in importance as a zoonotic parasite that has been causing regular episodes of malaria throughout South East Asia. The *P. knowlesi* genome sequence generated in 2008 highlighted and confirmed many similarities and differences in *Plasmodium* species, including a global view of several multigene families, such as the large SICAvAr multigene family encoding the variant antigens known as the schizont-infected cell agglutination proteins. However, repetitive DNA sequences are the bane of any genome project, and this and other *Plasmodium* genome projects have not been immune to the gaps, rearrangements and other pitfalls created by these genomic features. Today, long-read PacBio and chromatin conformation technologies are overcoming such obstacles. Here, based on the use of these technologies, we present a highly refined de novo *P. knowlesi* genome sequence of the Pk1(A+) clone. This sequence and annotation, referred to as the 'MaHPIC Pk genome sequence', includes manual annotation of the SICAvAr gene family with 136 full-length members categorized as type I or II. This sequence provides a framework that will permit a better understanding of the SICAvAr repertoire, selective pressures acting on this gene family and mechanisms of antigenic variation in this species and other pathogens.

**Funding:** Federal funds from the National Institute of Allergy and Infectious Diseases; National Institutes of Health, Department of Health and Human Services (Contract No.HHSN272201200031C) and the National Center for Research Resources (ORIP/OD P51OD011132). This study was also financially supported by the National Institutes of Health (R01 AI06775-01) to KGLR, the University of California, Riverside (NIFA-Hatch-225935) to KGLR and Institute Leadership Funds from La Jolla Institute for Allergy and Immun