Computational comparison of six *Plasmodium* genomes identifies putative human virulence genes

Christian Frech[1] and Nansheng Chen[1]

[1]Department of Molecular Biology and Biochemistry, Simon Fraser University, 8888 University Drive, Burnaby, V5A 1S6, Canada

E-mail: chenn@sfu.ca

BACKGROUND: With about 500 million infections and over one million deaths each year, malaria remains a major global health burden. Recent genome sequencing of both human and non-human *Plasmodium* parasites provides an opportunity to use comparative genomics to identify novel parasite-encoded molecular factors contributing to human disease.

METHODS: We computationally compared six sequenced *Plasmodium* genomes with the aim to identify new species-specific protein-coding genes potentially linked to human virulence. In this study, we focus on non-subtelomeric regions where species-specific differences are less apparent and thus less well characterized. First, to identify genes possibly linked to severe human malaria, we used OrthoCluster, a program recently developed in our laboratory, to perform whole-genome synteny analysis comparing two human parasites, the highly virulent parasite *P. falciparum* and its less virulent relative *P. vivax*. Second, to identify genes possibly linked to endemic human malaria, we performed whole-genome synteny analysis between *P. vivax* and its closely related monkey parasite *P. knowlesi*, which causes zoonotic disease but, other than *P. vivax,* is not endemic in human populations. Third, to identify genes possibly linked to human pathogenicity, we performed stringent BLAST-based proteome comparisons of the three aforementioned human disease-causing parasites with partial genome sequences from three rodent parasites, *P. chabaudi*, *P. berghei*, and *P. yoelii*, which do not cause malaria in humans.

RESULTS: Synteny analysis between *P. falciparum* and *P. vivax* reveals 118 *P. falciparum*-specific and 173 *P. vivax*-specific genes in non-subtelomeric regions, 25 (21%) and 100 (58%) of which without annotated function. The 118 *P. falciparum*-specific genes are highly enriched for known virulence factors, suggesting possible virulence-associated roles for the identified genes of currently of unknown function. Between the more closely related parasites *P. vivax* and *P. knowlesi*, we identified 142 non-subtelomeric *P. vivax*-specific genes, three of which constitute prime human virulence candidate genes because a syntenic ortholog is present in *P. falciparum* but not in any of the four investigated non-human parasites. BLAST-based comparison followed by syntenic examination of candidate genes confirmed 16 genes to be present in all three primate parasites but absent from rodent pathogens, suggesting a possible role of these genes in the pathogenesis of human disease.

CONCLUSION: Species-specific genes are not confined to dynamic subtelomeric regions of *Plasmodium* genomes but are also abundant in chromosome 'core' regions. Our bioinformatics comparison of human and non-human parasites proposes new candidate human virulence genes in non-subtelomeric regions, which can now be followed-up by targeted computational and experimental analyses to further elucidate their functions.