New complete genome of *Giardia intestinalis* reveals interesting structural variation

Feifei Xu[1], Staffan G. Svärd[1]

1. Department of Cell and Molecular Biology, BMC, Box 596, Uppsala University, SE-751 24 Uppsala, Sweden

Giardia intestinalis, an unicellular heterotrophic eukaryote, is an intestinal parasite. It infects humans and other mammals, causing diarrhoea. The first draft genome of *G. intestinalis* was published in 2007 in Science. The genome was sequenced in the era of Sanger sequencing, and was assembled into 306 contigs on 92 scaffolds, leaving 137 gaps of 1.64 Mbp in size. We resequenced the genome with PacBio technology, and the assembly was then scaffolded with the optical maps. The new draft genome consists of 35 scaffolds, with the 5 biggest scaffolds corresponding to the 5 chromosomes. There are only 4 gaps of 0.87 Mbp in size in the new near complete genome, all falling into the repetitive regions. We start to have a more complete picture of for example the chromosomal gene distribution and structure variation with such a complete genome. The *Giardia* genome is basically composed of three components, conserved regions with the core genes, geneless regions with few genes, and arrays of repetitive elements. The conserved regions have well-defined transcription boundaries, while the geneless regions are often highly expressed with no corresponding open reading frames. The only genes found inside the geneless regions are variant-specific surface proteins (VSPs) and NEK kinases. Arrays of repetitive elements include pseudogenized VSPs, ribosomal RNAs and transposons. Those arrays can be found both at the end of as well as within the chromosomes. Both the repetitive and the geneless regions are GC rich. With the long read third generation sequencing technologies becoming maturer, we are getting more complete genomes over the long repetitive regions, which makes it feasible to study the chromosomal structure variation and evolution. We hope this project could not only provide a new complete reference genome for the *Giardia* community, but also serve as an inspiration for re-sequencing of the old-fragmented protist genomes.