Karolinska Institutet

Institutionen för Cell och Molekylärbiologi

Genome and transcriptome studies of the protozoan parasites *Trypanosoma cruzi* and *Giardia intestinalis*

AKADEMISK AVHANDLING
som för avläggande av medicine doktorsexamen vid Karolinska Institutet offentligen försvaras i Föreläsningssal CMB, Berzelius väg 21.

Fredagen den 23 november, 2012, kl 13.00

av
Oscar Franzén
Herr

Huvudhandledare:
Professor Björn Andersson
Karolinska Institutet
Institutionen för Cell och Molekylärbiologi

Bihandledare:
Professor Staffan Svärd
Uppsala Universitet
Institutionen för Cell och Molekylärbiologi

Fakultetsopponent:
Dr. Matthew Berriman
Wellcome Trust Sanger Institute
Parasite genomics

Betygsnämnd:
Professor Siv Andersson
Uppsala Universitet
Institutionen för Cell och Molekylärbiologi

Docent Lars Feuk
Uppsala Universitet
Institutionen för Immunologi, Genetik och Patologi

Docent Antonio Barragan
Karolinska Institutet
Institutionen för Medicin

Stockholm 2012

Abstract

Trypanosoma cruzi and *Giardia intestinalis* are two human pathogens and protozoan parasites responsible for the diseases Chagas disease and giardiasis, respectively. Both diseases cause suffering and illness in several million individuals. The former disease occurs primarily in South America and Central America, and the latter disease occurs worldwide. Current therapeutics are toxic and lack efficacy, and potential vaccines are far from the market. Increased knowledge about the biology of these parasites is essential for drug and vaccine development, and new diagnostic tests. In this thesis, high-throughput sequencing was applied together with extensive bioinformatic analyses to yield insights into the biology and evolution of *Trypanosoma cruzi* and *Giardia intestinalis*. Bioinformatics analysis of DNA and RNA sequences was performed to identify features that may be of importance for parasite biology and functional characterization. This thesis is based on five papers (*i*-*v*). Paper *i* and *ii* describe comparative genome studies of three distinct genotypes of *Giardia intestinalis* (A, B and E). The genome-wide divergence between A and B was 23% and 13% between A and E. 4557 groups of three-way orthologs were defined across the three genomes. 5 to 38 genotype-specific genes were identified, along with genomic rearrangements. Genes encoding surface antigens, *vsps*, had undergone extensive diversification in the three genotypes. Several bacterial gene transfers were identified, one of which encoded an acetyltransferase protein in the E genotype. Paper *iii* describes a genome comparison of the human infecting *Trypanosoma cruzi* with the bat-restricted subspecies *Trypanosoma cruzi marinkellei*. The human infecting parasite had an 11% larger genome, and was found to have expanded repertoires of sequences related to surface antigens. The two parasites had a shared 'core' gene complement. One recent horizontal gene transfer was identified in *T. c. marinkellei*, representing a eukaryote-to-eukaryote transfer from a photosynthesizing organism. Paper *iv* describes the repertoire of small non-coding RNAs in *Trypanosoma cruzi* epimastigotes. Sequenced small RNAs were in the size range 16 to 61 nucleotides, and the majority were derived from transfer RNAs and other non-coding RNAs. 92 novel transcribed loci were identified in the genome, 79 of which were without similarity to known RNA classes. One population of small RNAs were found to be derived from protein-coding genes. Paper *v* describes transcriptome analysis using paired-end RNA-Seq of three distinct genotypes of *Giardia intestinalis* (A, B and E). Gene expression profiles recapitulated the known phylogeny of the examined genotypes, and 61 to 176 genes were differentially expressed. 49,027 distinct polyadenylation sites were mapped and compared, and the median 3′UTR length was 80 nucleotides (A). One 36-nt novel intron was identified and the previously reported introns (5) were confirmed.

ISBN 978-91-7457-860-7