PROTEOME-WIDE EVOLUTIONARY ANALYSIS REVEALS LINEAGE-SPECIFIC ADAPTATIONS AND IMPROVES FUNTIONAL ANNOTATION OF Schistosoma mansoni PROTEINS

Larissa Silva, Marina Marcet-Houben, Laila Nahum, Adhemar Zerlotini, Toni Gabaldón, Guilherme Oliveira

Centro de Pesquisas René Rachou - CPqRR, Centro de Excelência em Bioinformática - CEBio; Fundação Oswaldo Cruz - FIOCRUZ, Belo Horizonte, MG., Bioinformatics and Genomics Programme, Center for Genomic Regulation (CRG) and UPF. Barcelona, Spain, Faculdade Infórium de Tecnologia; Centro de Pesquisas René Rachou - CPqRR; Fundação Oswaldo Cruz - FIOCRUZ, Belo Horizonte, MG, Laboratório Multiusuário de Bioinformática, Embrapa Informática Agropecuária, Campinas, São Paulo, Brazil, Bioinformatics and Genomics Programme, Center for Genomic Regulation (CRG) and UPF. Barcelona, Spain, Centro de Pesquisas René Rachou - CPqRR, Centro de Excelência em Bioinformática - CEBio; Fundação Oswaldo Cruz - FIOCRUZ, Belo Horizonte, MG.

Schistosomiasis, as many other helminthiases, is a neglected tropical disease and a major public health problem affecting about 230 million people in 77 countries. Measures to control schistosomiasis rely mainly on Praziquantel®, the only drug available for mass treatment. There is as yet no available vaccine. Aiming at moving towards schistosomiasis control and elimination, the World Health Organization (WHO) supported a genome sequencing initiative for Schistosoma species as an alternative to identify new targets for drug and vaccine development, as well as to understand the molecular basis of parasite biology. Although a significant effort has been made toward identifying schistosome gene and protein functions, a complete picture of the parasite genome/proteome and its biology and evolution remains a challenging goal. In this context, comparative proteomic analysis can shed new light on the evolutionary processes that shaped hostparasite interaction over evolutionary time. Taking advantage of the benefits provided by a largescale phylogenetic analysis, in the present work, we reconstructed the evolutionary history of each protein encoded in the S. mansoni genome in comparison with other 12 taxa to gain insights into lineage-specific evolutionary events, potentially related to the parasitic lifestyle, as well as to improve functional annotation. Results The collection of trees reconstructed in this work includes 7,964 phylogenies, which comprises the evolutionary histories of all parasite proteins and their homologs across 12 other organisms. This analysis allowed a deeper understanding of genomic complexity and evolutionary adaptations to a parasitic lifestyle. In particular, the identification of lineage-specific gene duplications pointed to the diversification of several protein families that are relevant for host-parasite interactions, including proteases, tetraspanins, fucosyltransferases, venom allergen-like proteins and tegumentalallergen-like proteins. In addition to the evolutionary knowledge, the phylome data enabled us to automatically improve the functional annotation of 3,451 protein-coding genes through a phylogenetic-based approach rather than just using sequence similarity based searches. To allow further exploitation of this valuable data, all information has been made available at PhylomeDB (http://www.phylomedb.org). Conclusions By using an evolutionary approach, we analyzed the S. mansoni proteome from a system-wide perspective to assess parasite biology, improve functional annotation of the parasite proteome, and provide insights into host-parasite interactions. Taking advantage of a large-scale analysis rather than focusing on individual proteins, we identified that this parasite has experienced specific gene duplication events over evolutionary time, particularly affecting genes/proteins that are potentially related to the parasitic lifestyle. These innovations may be related to the mechanisms that protect S. mansoni against host immune responses being important adaptations for the parasite survival in a potentially hostile environment. Continuing this work, a comparative analysis involving genomic, transcriptomic, and proteomic data from other helminth parasites and vectors will supply more information regarding parasite biology as well as host-parasite interactions.

Keywords: phylogenomics, maximum likelihood analysis, homology prediction, functional annotation, paralogous families, parasite genomics, schistosomiasis

Concentration area: Genomics Evolution