

Genomic Scale Analysis of Lateral Gene Transfer in Apicomplexan Parasites: insights into early eukaryotic evolution, host-pathogen interaction and drug target development

Lucia Peixoto and David S. Roos

Department of Biology and Penn Genomics Institute, University of Pennsylvania, Philadelphia, PA, USA.

The protozoan phylum Apicomplexa is comprised of >5000 species, including *Toxoplasma gondii* (a prominent opportunistic infection in AIDS), and the *Plasmodium* parasites responsible for malaria. Complete genome sequences are now available for several apicomplexan species, including both *T. gondii* and *Plasmodium falciparum*, making it possible to exploit comparative phylogenomic tools to identify targets for therapeutic development, and providing insight into adaptations associated with intracellular parasitism. Remarkably, these parasites harbor a novel organelle -- the apicoplast -- which was acquired by endosymbiotic capture of a eukaryotic alga, providing a dramatic example of lateral gene transfer (LGT). In order to determine the overall impact of LGT, we have employed phylogeny reconstruction to detect transfers on a genomic scale. Trees have been constructed for all genes in the *Toxoplasma* and *Plasmodium* genomes, and compared with a whole genome species tree derived from 55 taxa present in the orthologous protein database OrthoMCL-DB. Non-apicomplexan specific ortholog groups containing 4 to 100 sequences were examined, pruning outlying species based on their degree of connectivity to the group as a whole, and robustness of the resulting tree.

Our previous study of the *T. gondii* genes coding for metabolic enzymes estimates that about 15% of them were acquired by LGT. Turning to the entire parasite genome(s), Horizstory software identified 1552 groups where *T. gondii* gene placement contrasts with the species tree. Similarly, 1520 instances of potential LGT were identified within *P. falciparum*. To identify the most probable LGT events we developed the LGTsmart strategy, ranking probability of LGT based on the taxonomic consistency of sibling species placement within the tree, and bootstrap support for the corresponding internal node(s). The net result of this strategy is a ranked list of probable lateral gene transfer events. A closer look to the highest ranked genes reveals a subset of 228 high confidence LGT events that occurred at different times during apicomplexan evolution (or were differentially lost within apicomplexan subclades, Fig1). Not surprising a lot of the transfers involve plant-like genes, consistent with the endosymbiosis of an algae, including many promising new targets for drug development; some recent bacterial transfers were also observed. Interestingly we also detected LGT of some metazoan-like genes, possibly acquired during the adaptation to the parasitic lifestyle and the interaction with the host.

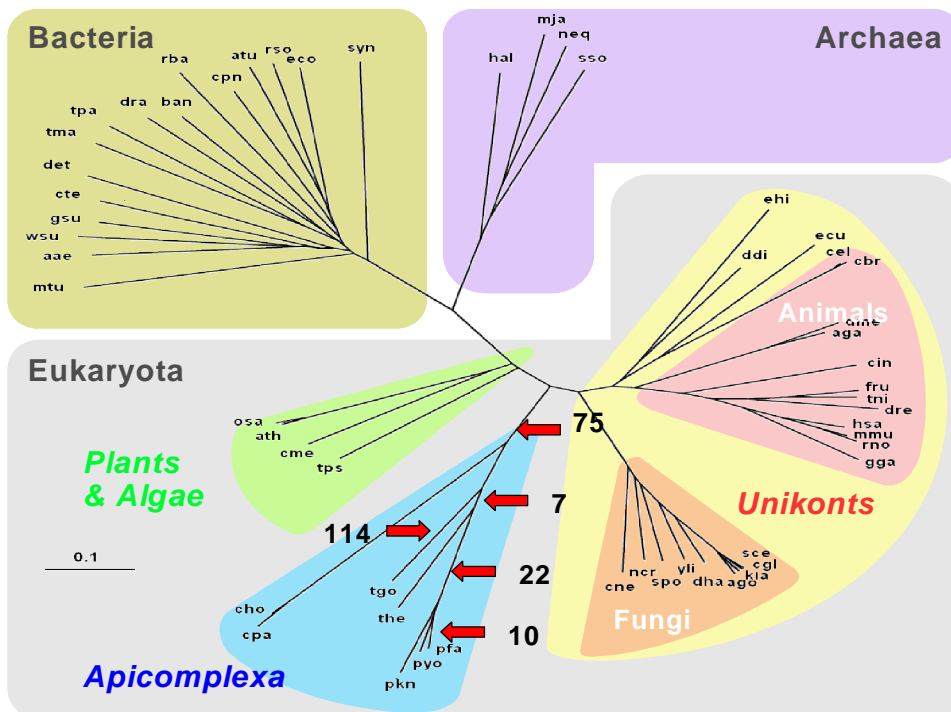


Fig1- Whole genome phylogeny based on gene content of the 55 species present in OrthoMCL-DB (see <http://orthomcl.cbil.upenn.edu> for species abbreviations)

High confidence LGT events occurring at different times during apicomplexan evolution are highlighted. These include 91 plant-like genes, 14 bacterial genes, and 95 metazoan-like genes