RAPID CONCURRENT BURSTS OF GENE DUPLICATIONS IN MAJOR SURFACE PROTEIN FAMILIES OF *TRYPANOSOMA CRUZI*: POTENTIAL LINK WITH EVOLUTION OF PARASITIC LIFESTYLE

Carlos A. Flores-López[*a], Carlo German González Vera [a], Carlos A. Machado [b]

[a] *Facultad de Ciencias, Universidad Autónoma de Baja California, Km. 103 Carretera Tijuana – Ensenada, Pedregal Playitas, 22860 Ensenada, Baja California, Mexico. b Department of Biology, University of Maryland, College Park, MD 20742.*

The advent of the genomic era has provided fundamental data to study how a parasitic lifestyle can evolve. In particular, it is now possible to address the genomic transitions that take place when: 1) a free-living organism makes the transition to a pathogenic lifestyle, 2) a pathogen evolves the ability to survive inside a cell (intracellular parasitism), 3) a pathogen evolves the capacity to invade different host cells, and/or 4) a pathogen acquires the ability to invade multiple host species. Most of the genomic studies about pathogen evolution that have been recently completed have focused on addressing the first question, where genome size reduction, gene length reduction, loss of metabolic pathways, and species-specific protein expansions have been the most common traits found amongst a diverse array of pathogens (Glaser et al, 2001; Heinz et al, 2012; Nerima et al, 2010; Razin, 1997; Toft & Andersson, 2010; Tsai et al, 2013; Wernegreen, 2005). In this study we address the last three questions using genomic data from *Trypanosoma cruzi* (the agent of Chagas disease) and closely related pathogens.

To explore the genomic changes associated with the evolution of several aspects of the evolution of a parasitic lifestyle we focused on *Trypanosoma cruzi* due to several characteristics of this parasite: 1) availability of an annotated genome sequence (El-Sayed et al, 2005)., 2) the parasite is an obligate intracellular pathogen during the mammal host stage (an adaptation that is not shared by the closely related *T. brucei*, the causative agent of sleeping sickness in the African continent), 3) the evolution of intracellular parasitism occurred more than once in the kinetoplastids (e.g. species of the genus *Leishmania* also adapted to an intracellular niche and became obligate intracellular parasites that infect mostly macrophages (Sibley, 2011)), 4) *T. cruzi* has the remarkable ability to invade most nucleated cells within the mammal host, 5) *T. cruzi* can infect a diverse range of mammal hosts (the parasite has been isolated from more than 70 genera (Zingales et al, 2012)), and 6) a recent evolutionary comparative genomic study of ours (Flores-Lopez & Machado, 2015) revealed that *T. cruzi spp*. had a significant larger amount of proteins under positive selection compared to *Leishmania spp*., which suggest recent evolutionary adaptations.

The *T. cruzi* genome-sequencing project found that approximately 50% of the protein-coding genes of this parasite were members of very large gene families (El-Sayed et al,

2005). The majority of those gene families were surface proteins such as trans-sialidases (TSs), mucins, mucin associated surface proteins (MASPs), dispersed gene family 1 (DGF-1), and gp63 peptidases (De Pablos & Osuna, 2012; El-Sayed et al, 2005). Higher sequence conservation observed within paralogs from the same strain haplotype, rather than between haplotypes, suggests these protein families are of recent origin (El-Sayed et al, 2005).

Using gene duplication age distributions we show that the expansions of those cell surface protein families in *T. cruzi* occurred recently and in rapid, concurrent, bursts. Convergent evolution of intracellular parasitism in another kinetoplastid (*Leishmania spp*.) suggests that major protein family duplication events were not required for the evolution of this trait. However, the time estimates, functions and phylogenetic distribution of these gene family expansions, suggest that these massive gene family expansions were linked to: 1) the parasite evolution from having a monoxenous development (development restricted to 1 host species) to a heteroxenous development (development requires 2 or more host species), and 2) the parasite's capacity for invading multiple cell tissues. Time estimates for the massive gene family expansions overlap with the predicted evolution of hematophagy in the insect vectors. Given that hematophagy increased the vector's host range, we hypothesize that this event may have indirectly selected for those rapid gene family expansions in the ancestor of *T. cruzi*.

The description of the sudden and great expansion of a few surface protein families occurring in rapid concurrent bursts represents the largest gene family expansions in all kinetoplastid pathogens sequenced thus far**.** The functions of these proteins and the estimated time of their expansions appear to have been driven by the evolution of the parasite's adaptation to invade mammal species, since their ancestors were only adapted to insect hosts. Our results suggest that the evolution of this parasitic lifestyle trait was the selective force driving the maintenance of these massive gene family expansions.

- De Pablos LM, Osuna A (2012) Multigene families in Trypanosoma cruzi and their role in infectivity. *Infect Immun* **80:** 2258-2264
- Glaser P, et al. (2001) Comparative genomics of Listeria species. *Science* **294:** 849-852
- Heinz E, et al. (2012) The genome of the obligate intracellular parasite Trachipleistophora hominis: new insights into microsporidian genome dynamics and reductive evolution. *PLoS Pathog* **8:** e1002979
- El-Sayed NM, et al (2005) The genome sequence of Trypanosoma cruzi, etiologic agent of Chagas disease. *Science* **309:** 409-415
- Sibley LD (2011) Invasion and intracellular survival by protozoan parasites. *Immunological reviews* **240:** 72-91
- Flores-Lopez CA, Machado CA (2015) Differences in inferred genome-wide signals of positive selection during the evolution of Trypanosoma cruzi and Leishmania spp. lineages: A result of disparities in host and tissue infection ranges? *Infect Genet Evol* **33:** 37-46
- Nerima B, Nilsson D, Maser P (2010) Comparative genomics of metabolic networks of free-living and parasitic eukaryotes. *BMC Genomics* **11:** 217
- Razin S (1997) Comparative genomics of mycoplasmas. *Wien Klin Wochenschr* **109:** 551-556
- Toft C, Andersson SG (2010) Evolutionary microbial genomics: insights into bacterial host adaptation. *Nat Rev Genet* **11:** 465-475

- Tsai IJ, et al (2013) The genomes of four tapeworm species reveal adaptations to parasitism. *Nature* **496:** 57-63
- Wernegreen JJ (2005) For better or worse: genomic consequences of intracellular mutualism and parasitism. *Curr Opin Genet Dev* **15:** 572-583
- Zingales B, et al (2012) The revised Trypanosoma cruzi subspecific nomenclature: rationale, epidemiological relevance and research applications. *Infect Genet Evol* **12:** 240-253

*cflores2@uabc.edu.mx