

Comparative genomics of *Cryptosporidium*

Simone M. Cacciò

European Union Reference Laboratory for Parasites, Istituto Superiore di Sanità, Rome, Italy

e-mail: simone.caccio@iss.it

Cryptosporidium is a protozoan parasite that infect humans and animals worldwide, causing a diarrheal disease that is particularly severe, and unfortunately poorly treatable, among the very young and the immunocompromised. Two species, *Cryptosporidium hominis* and *C. parvum*, are responsible for the vast majority of human cases of cryptosporidiosis. Not surprisingly, therefore, the first genome sequencing efforts focused on these two species and culminated about 15 years ago in the description of a small, compact genome (9.1 Mb) comprised of eight chromosomes, with a strong conservation in synteny between the *C. parvum* (IOWA) and *C. hominis* (TU502) isolates.

The introduction of high throughput sequencing techniques also referred to as Next Generation Sequencing (NGS), has had a big impact on the field of genomics allowing faster and cheaper data generation. However, intrinsic challenges in the *in vitro* and in vivo propagation of parasite isolates has limited the application of NGS to *Cryptosporidium*. Several groups have addressed this issue and improved methods for sample processing (including all steps from oocyst purification, robust DNA extraction and whole genome amplification) have been published. In turn, this has made possible to sequence *Cryptosporidium* genomes directly from fresh or archived fecal samples. Single cell whole genome sequencing techniques were also applied. In parallel to improved wet-lab procedures, efforts were also made to develop and test ad-hoc bio-informatics pipelines, as this should solve one of the perceived barrier in the use of NGS, that is to say, data analysis.

Research questions that have promoted genomic studies include understanding the extent of genetic variability of human and animal isolates, the role of recombination in the evolution of virulent strains and in the process of host adaptation, and the identification of species-specific genes.

These studies clearly demonstrated that large genetic variability is a common feature of *Cryptosporidium* isolates, and that the distribution of single nucleotide polymorphisms (SNPs) and indels is non-random, with more variability observed at telomeric and sub-telomeric regions of some chromosomes. These regions contain genes encoding for secreted proteins that are important in host-parasite relationships. Parasite isolates cluster differently based on genome-wide SNPs compared to what observed at the gp60 locus, indicating that better tools for molecular epidemiologic studies can be developed from genome data. The existence of *C. parvum* clusters related to different hosts (human, different ruminants) and/or different geographical origin is still unclear.

Other studies have shown that the essential role played by genetic recombination in the evolution of *Cryptosporidium*, and found that recombinant regions are enriched for positively

selected genes and potential virulence factors. Furthermore, levels of genetic variability and population structures vary among *C. hominis* isolates from different geographical regions, likely reflecting the relative transmission rates and the likelihood of mixed infections.