

Draft genome sequences of *Blastocystis* subtypes 1 and 8, and comparative analysis.

Ivan Wawrzyniak¹, Amandine Cian², Céline Nourrisson¹, Aldert Bart³, Magali Chabé², Philippe Poirier¹, Tom van Gool³, Eric Peyretailade¹, Eric Viscogliosi², Frédéric Delbac¹.

¹ Laboratoire Microorganismes Génome et Environnement, Université Clermont Auvergne, CNRS, UMR 6023, 1 impasse Amélie Murat, F 63000, Clermont-Ferrand, France

² Center of Infection and Immunity of Lille, Institut Pasteur de Lille, Inserm U1019, CNRS, UMR 9017, University of Lille, CHU of Lille, F 59019, Lille, France

³ Department of Medical Microbiology, Amsterdam Medical Center, 1105 AZ, Amsterdam, The Netherlands

Corresponding Author: ivan.wawrzyniak@uca.fr

Blastocystis is a highly prevalent anaerobic eukaryotic parasite found in the intestinal tract of human and various animals. Although the role of *Blastocystis* as human pathogen remains unclear, it can cause acute or chronic digestive disorders and some studies have suggested an association with irritable bowel syndrome. Seventeen subtypes (ST1-ST17), among which ten (ST1 to ST9 and ST12) are found in human with varying prevalence, have been identified based on the small-subunit ribosomal RNA. The genomes of three isolates belonging to ST1 [1], ST4 [2] and ST7 [3] have been previously sequenced and annotated. Using illumina HiSeq 2000 system technology, we conducted genome sequencing and annotation of a new *Blastocystis* ST1 isolate and one ST8 isolate. Assembly of reads generated genomic sequence of 15.01 Mbp possessing 6604 putative genes and 13.9 Mbp for 5579 putative genes in ST1 and ST8, respectively. The Mitochondrion-Like Organelle (MLO) genome full sequences were obtained for both ST1 (28.3 kbp) and ST8 (27.9 kbp). These circular mitochondrial genomes encompass 10 NADH subunits, 13 ribosomal proteins and 4 ORFs with unknown functions and present a highly conserved gene synteny. Comparative analysis of whole proteomes identified the core and the specific proteins between the five sequenced ST. Analysis of the core proteins revealed an over representation of gene families coding for proteins that may be involved in virulence including hydrolases. Some of these proteins are predicted to be secreted or targeted to the plasma membrane and may play important roles in processes such as cytoadherence, host cell invasion, molecule degradation, or host immune response evasion. A whole genome phylogeny of *Blastocystis* confirmed the *Blastocystis* lineage based on the complete SSU rDNA [4]. Our genomic comparative analysis also revealed a highly conserved gene synteny between the 5 ST. Finally, we demonstrated for the first time the presence of transposable elements in *Blastocystis* using TransposonPSI software and a tblastX approach.

References

- [1] Eleni Gentekaki, Bruce A Curtis, Courtney W stairs, Vladimir Klimes, Marek Elias, Dayana E Salas-Leiva, Emily K Herman, Laura Eme, Marias C Arias, Bernard Henrissat et al. Extreme genome diversity in the hyper-prevalent parasitic eukaryote *Blastocystis*. *Plos Biology*, (46/47):217–252, 2017.
- [2] Ivan Wawrzyniak, Damien Courtine, Marwan Osman, Christine Hubans-Pierlot, Amandine Cian, Céline Nourrisson, Magali Chabé, Philippe Poirier, Aldert Bart, Valérie Polonais et al. Draft genome sequence of the intestinal parasite *Blastocystis* subtype 4-isolate WR1. *Genomic Data*. Feb 2; 4:22-3, 2015.
- [3] France Denoeud, Michael Roussel, Benjamin Noël, Corinne Da Silva, Marie Diogon, Eric Viscogliosi, Céline Brochier-Armanet, Arnaud Couloux, Julie Poulain, Béatrice Segurens et al. Genome sequence of the stramenopile *Blastocystis*, a human anaerobic parasite. *Genome Biology*, 12(3): R29, 2011.
- [4] Hisao Yoshikawa, Yukiko Koyama, Erika Tsuchiya, Kazutoshi Takami. *Blastocystis* phylogeny among various isolates from humans to insects. *Parasitology International*, Dec;65:750–759, 2016.