

## A *Plasmodium falciparum* FcBI-schizont-EST collection providing clues to schizont specific gene structure and polymorphism

Isabelle Florent\*<sup>1</sup>, Betina M Porcel<sup>2</sup>, Elodie Guillaume<sup>1</sup>, Corinne Da Silva<sup>2</sup>, François Artiguenave<sup>2</sup>, Eric Maréchal<sup>3</sup>, Laurent Bréhélin<sup>4</sup>, Olivier Gascuel<sup>4</sup>, Sébastien Charneau<sup>1</sup>, Patrick Wincker<sup>2</sup> and Philippe Grellier<sup>1</sup>

Address: <sup>1</sup>FRE3206 CNRS/MNHN, UISM504, Biologie Fonctionnelle des Protozoaires, RDDM, Muséum National d'Histoire Naturelle, Paris, France, <sup>2</sup>UMR-CNRS 8030, Genoscope, Evry, France, <sup>3</sup>UMR-CNRS 5168, CEA, INRA, Université Joseph Fourier, Grenoble, France and <sup>4</sup>UMR-CNRS 5506, Laboratoire d'Informatique, de Robotique et de Micro-électronique de Montpellier, Université de Montpellier II, Montpellier, France

Email: Isabelle Florent\* - florent@mnhn.fr; Betina M Porcel - betina@genoscope.cns.fr; Elodie Guillaume - guillaume@mnhn.fr; Corinne Da Silva - dasilva@genoscope.cns.fr; François Artiguenave - artiguenave@genoscope.cns.fr; Eric Maréchal - eric.marechal@cea.fr; Laurent Bréhélin - brehelin@lirmm.fr; Olivier Gascuel - gascuel@lirmm.fr; Sébastien Charneau - charneau@unb.br; Patrick Wincker - pwincker@genoscope.cns.fr; Philippe Grellier - grellier@mnhn.fr

\* Corresponding author

Published: 19 May 2009

Received: 11 September 2008

BMC Genomics 2009, 10:235 doi:10.1186/1471-2164-10-235

Accepted: 19 May 2009

This article is available from: <http://www.biomedcentral.com/1471-2164/10/235>

© 2009 Florent et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

### Abstract

**Background:** The *Plasmodium falciparum* genome (3D7 strain) published in 2002, revealed ~5,400 genes, mostly based on *in silico* predictions. Experimental data is therefore required for structural and functional assessments of *P. falciparum* genes and expression, and polymorphic data are further necessary to exploit genomic information to further qualify therapeutic target candidates. Here, we undertook a large scale analysis of a *P. falciparum* FcBI-schizont-EST library previously constructed by suppression subtractive hybridization (SSH) to study genes expressed during merozoite morphogenesis, with the aim of: 1) obtaining an exhaustive collection of schizont specific ESTs, 2) experimentally validating or correcting *P. falciparum* gene models and 3) pinpointing genes displaying protein polymorphism between the FcBI and 3D7 strains.

**Results:** A total of 22,125 clones randomly picked from the SSH library were sequenced, yielding 21,805 usable ESTs that were then clustered on the *P. falciparum* genome. This allowed identification of 243 protein coding genes, including 121 previously annotated as hypothetical. Statistical analysis of GO terms, when available, indicated significant enrichment in genes involved in "entry into host-cells" and "actin cytoskeleton". Although most ESTs do not span full-length gene reading frames, detailed sequence comparison of FcBI-ESTs versus 3D7 genomic sequences allowed the confirmation of exon/intron boundaries in 29 genes, the detection of new boundaries in 14 genes and identification of protein polymorphism for 21 genes. In addition, a large number of non-protein coding ESTs were identified, mainly matching with the two A-type rRNA units (on chromosomes 5 and 7) and to a lower extent, two atypical rRNA loci (on chromosomes 1 and 8), TARE subtelomeric regions (several chromosomes) and the recently described telomerase RNA gene (chromosome 9).

**Conclusion:** This FcBI-schizont-EST analysis confirmed the actual expression of 243 protein coding genes, allowing the correction of structural annotations for a quarter of these sequences. In addition, this analysis demonstrated the actual transcription of several remarkable non-protein coding loci: 2 atypical rRNA, TARE region and telomerase RNA gene. Together with other collections of *P. falciparum* ESTs, usually generated from mixed parasite stages, this collection of FcBI-schizont-ESTs provides valuable data to gain further insight into the *P. falciparum* gene structure, polymorphism and expression.