SEQTRYPLANT

Obtaining the full sequence of the genome of two plants trypanosomatids.

ANR-08-GENM- 020-001

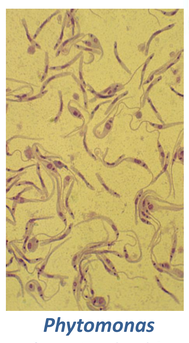
Coordinateur: Michel Dollet (CIRAD, Montpellier)
Partenaires: Patrick Wincker, Betina Porcel, Benjamin Noël, France Denoeud, Kamel Jabbari, Jean-Marc Aury, Corinne Da Silva, Arnaud Couloux, Julie Poulain (GENOSCOPE-CEA, Evry).

Aims

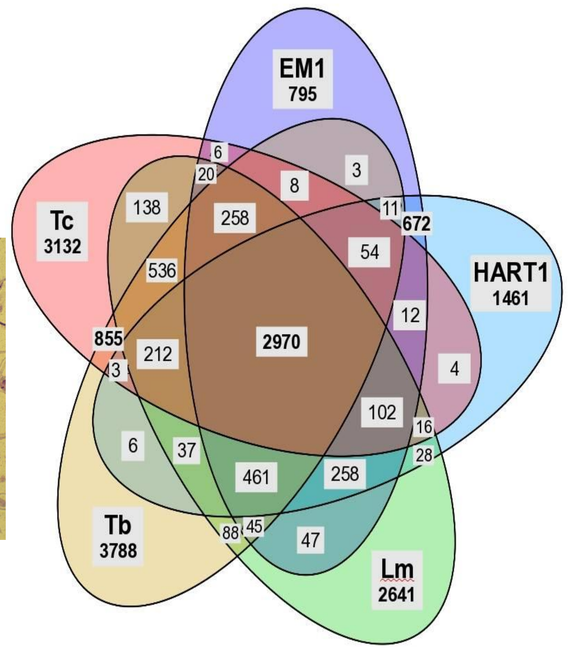
Some trypanosomatids multiplying in the phloem sap are responsible for wilts of tropical crops (coconut, oil palm) and have a major economic impact in Latin America and the Caribbean. Other trypanosomatids from laticiferous plants appear as “symbionts-like”- non pathogenic- and others multiply in fruits. Only one arbitrary genus name has been proposed as yet for all these trypanosomatids: “*Phytomonas”.* For this project we proposed the comparative sequencing of two plant trypanosomatids, one pathogenic isolate from a diseased coconut from Guiana and one latex isolate from *Euphorbia*.

Hartrot of coconut

Hartrot *Phytomonas*



Phytomonas from *Euphorbia*



Number of genes shared between EM1, HART1, *Leishmania major*, *Trypanosoma brucei* and *T.cruzi*

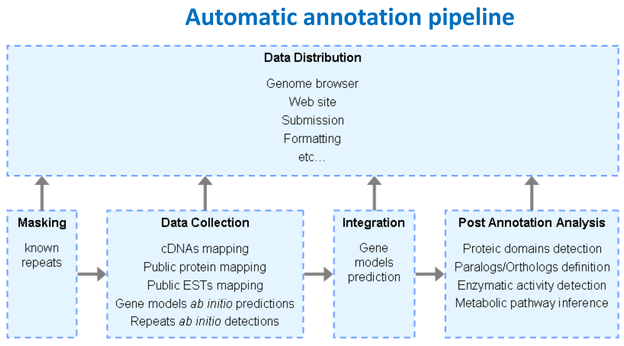
Methods and Results

The sequencing of the two plant trypanosomatids, **Hart1** (pathogenic phloem isolate from coconut – **Group H**-) and **EM1**(symbiont-like latex isolate from *Euphorbia*- **Group D**-*)* were obtained, using data from three different technologies (454 Titanium, Illumina GAIIx and Sanger). Both assemblies were annotated using a combination of gene evidence, for prediction of a reference set of gene models.

General features of *Phytomonas* EM1 and HART1 genomes

	EM1	HART1
Genome size	17.8 Mb	18.1 Mb
% masked	0.9%	1.2%
% ab initio repeats	0.2%	1.5%
Number of genes	6381	6451
Number of operons	298	334
Gene size (avg ;med)	1,614; 1,245	1,507; 1,185
Gene coverage	57.9%	53.7%
Intergenic lenght (avg ;med)	1,140 : 518	1,280 : 584

Automatic annotation pipeline



The total number of genes in both isolates is abridged, when compared to other trypanosomatids. The compaction in these parasite genomes was reflected by the short length of the intergenic regions and the relatively low repeat coverage as well. Moreover, no significant difference in gene size was observed between both isolates.

The genome of the two sequenced isolates consists essentially of single copy genes for the bulk of its metabolic enzymes, where *Leishmania* and *Trypanosoma* possess numerous duplicated genes or large gene families.
OrthoMCL was used to build clusters of orthologous genes from BLAST hits between trypanosoma proteoms. The number of genes shared between EM1, HART1, *Leishmania major*, *Trypanosoma brucei* and *T.cruzi* are displayed on a Venn diagram.

Genes not shared by EM1 and HART1:

Genes from each strain with no orthologous gene found in the other strain were curated in order to remove genes missing from the assembly/annotation.

13 HART1 genes absent from EM1:

gene id	Description	homologues in other tryps
GSHART1T00000090001	-	no
GSHART1T00000095001	-	no
GSHART1T00000658001	-	no
GSHART1T00000732001	-	no
GSHART1T00001296001	-	yes (doubtful)
GSHART1T00003148001	-	no
GSHART1T00003542001	-	no
GSHART1T00003697001	Signal peptide and two transmembrane domains	no
GSHART1T00007428001	IPR domain: Zinc finger, GATA-type	no
GSHART1T00007485001	-	no
GSHART1T00007721001	IPR domain: Thioredoxin-like	no
GSHART1T00007887001	-	no

19 EM1 genes absent from HART1:

gene id	Description	homologues in other tryps
GSEM1T00000365001	-	yes
GSEM1T00000681001	HAD like domain IPRO23214	yes
GSEM1T00000818001	transmembrane domain; IPR domains: calmodulin-binding region	yes
GSEM1T00000902001	-	yes
GSEM1T00001405001	-	no
GSEM1T00002173001	secreted cysteine proteinase with signal peptide and transmembrane region	yes
GSEM1T00002561001	DUS3L_XENLA Q7ZWS1 tRNA-dihydrouridine synthase 3-like (1.-.-.-)	yes
GSEM1T00002618001	IPR domain: Meckelin, transmembrane	yes
GSEM1T00002868001	-	no
GSEM1T00002885001	LBR_PONAB Q5R7H4 Lamin-B receptor (Integral nuclear envelope	yes
GSEM1T00004696001	-	no
GSEM1T00005374001	-	no
GSEM1T00005834001	-	no
GSEM1T00006049001	-	no
GSEM1T00006247001	Inositol-polyphosphate phosphatase with PTS1 (SHL)	yes
GSEM1T00006477001	YCF45_PORPU P51281 Uncharacterized protein ycf45 (ORF565)	yes
GSEM1T00006759001	-	no
GSEM1T00006777001	-	no

Conclusions and Perspectives

Analysis of these plant parasites genome sequences provides a global view of the metabolic potential of plant trypanosomatids. Indeed, the comparison of these plant parasites to the other sequenced Trypanosomatids revealed a streamlined genome, encoding for a minimal system, which conserves the major complexes and pathways, indicating retention of all major organelles, but with an apparent lack of complexity. Adaptation has possibly been via minimization rather than novelty, considering a likely less complex life cycle of these parasites in their host, the plants.

These results will open several doors for future research programmes and applications as far as phytopathology protistology, parasitology, biochemistry, and evolution, are concerned.
An important consequence of these results concerns the application to taxonomical changes in the plant/insect trypanosomatids. It would be very interesting to sequence other plant trypanosomatids belonging to other groups as well “insect trypanosomatids” able to multiply in insects, plants and mammals.

The knowledge of the genes involved in the different aspects of metabolism will help for improvement of the *in vitro* culture protocols of these fastidious parasites, and may provide new tools for the control of the diseases they cause.

CONTACT :
michel.dollet@cirad.fr
pwincker@genoscope.cns.fr

P0	Thematic: «GENOMIQUE VEGETALE» **ANR n°: «ANR-08-GENM- 020-001»** **Acronym: «SEQTRYPLANT»** Total Cost: **«1236528»** euros Total Grant: **«277295»** euros Date of beginning: **«01.01.2009»** – End: **«31.12.2011»**	**Edition 2008**

«Obtaining the full sequence of the genome of two plants trypanosomatids»

Autors: Dollet Michel[(1)], Fabre sandrine[(1)],
Wincker Patrick[(2)], Porcel Betina[(2)], Noël Benjamin[(2)], Denoeud France[(2)], Jabbari Kamel[(2)], Aury Jean-Marc[(2)], Da Silva Corinne[(2)], Couloux Arnaud[(2)], Poulain Julie[(2)],
Bringaud Frédéric[(3)], Campbell David[(4)], Cintron Roxana[(5)], Dickens Nicholas [(9)], Docampo Roberto[(5)], Field Mark[(6)], Hammarton Tansy[(9)], Koumandou Lila[(7)], Matthews Keith [(11)], Mottram Jeremy[(9)], Michaeli Shula[(8)], Opperdoes Fred[(10)], Szoor Balazs[(11)], Sturm Nancy[(4)], Zilberstein Dan[(12)]
[(1)] CIRAD, Montpellier, [(2)] Institut de Génomique, Evry,
[(3)] Bordeaux, [(4)] Los Angeles (USA), [(5)] Athens, (USA), [(6)] Cambridge (UK) , [(7)] Athens (Greece), [(8)] Ramat-Gan (Israel), [(9)] Glasgow (UK) , [(10)] Louvain la Neuve (Belgium), [(11)] Edinburgh (UK), [(12)] Haifa (Israel).

Aims

Some trypanosomatids multiplying in the sap are responsible for wilts of tropical crops like the hartrot disease of coconut palm. They have a major economic impact in Latin America and the Caribbean. The goal of the research on these parasites is to attain an integrated control of these diseases. In the latex vessels of plants, other trypanosomatids appear as “symbionts-like” without any negative effect on their host and others multiply in fruits and seeds. The latter usually are like trypanosomatids formerly known as “lower trypanosomatids” (*Crithidia*, *Herpetomonas* and *Leptomonas*). Only one arbitrary genus name has been proposed as yet for all these trypanosomes living in plants, in different tissues, with different consequences: “*Phytomonas*”. Parasitologists have become increasingly interested in the relationship between “*Phytomonas*” and the “lower trypanosomatids” due to the recent discovery of one of them in immuno-suppressed patients affectd by AIDS.

For this project we proposed the comparative sequencing of two plant trypanosomatids, one pathogenic isolate from coconut and one latex isolate from Euphorbia.

The benefits from the sequencing effort will be:

- The insights gained into the molecular evolutionary relationships within the Trypanosomatidae (*Trypanosoma* and *Leishmania* cause severe human diseases like sleeping sickness, Chagas disease and Kala-azar.)
- Identification of subsets of genes linked to different pathogenesis mechanisms.
- A better understanding of the biology of “*Phytomonas*,” which will eventually result in better and safer control methods.
- New keys for taxonomical changes in the former “lower trypanosomatids”. Because plant trypanosomatids have not been *in vitro* cultured before 1986, they are poorly known compared to human trypanosomatids.

Results

The sequencing of two plant trypanosomatids, one phloem restricted -Hart 1, Group H - responsible for a disease of coconut (hartrot) in Latin America, and one non-pathogenic, intra- laticiferous tubes - EM1, Group D - from *Euphorbia pinea* from South of France have been obtained. Sequencing of the isolates was performed using data from three different technologies (454 Titanium, Illumina GAIIx and Sänger). Assemblies obtained by Newbler (Roche) showed high continuity. We also obtained cDNA sequences using 454 Titanium to help annotation. The assemblies were of sufficient continuity to start an automatic annotation phase, using procedures that involve cDNA, matches to protein data, and *de novo* gene finding. We annotated 6288 (for EM1) and 6360 (for Hart1) gene models. These genes are being compared between the two plant trypanosomatids isolates and between these isolates and other sequenced Trypanosomatid genomes.

In October 2010 we created a consortium with 15 scientists from 10 laboratories for specific annotation of the two genomes: UMR 5536 CNRS- University Bordeaux 2, Department of Microbiology- University California Los Angeles (USA), Department of Cellular Biology- University of Georgia, Athens (USA), Department of Pathology -

Cambridge University (UK), Biomedical Research Foundation – Academy of Athens (greece), Mina and Everad Goodman Faculty of Sciences –Ramat-Gan (Israel), Welcome Trust Center for Molecular Parasitology – University of Glasgow (UK), Research Unit for Tropical Diseases-University of Louvain La Neuve (Belgium), Institute of Immunology and Infection Research – University of Edinburgh (UK), Faculty of Biology – Technion- Israel Institute of Technology, Haifa (Israel).

The total number of genes in both *Phytomonas* strains is abridged, when compared to other trypanosomatids. The compaction in these parasite genomes was reflected by the short length of the intergenic regions and the relatively low repeat coverage as well. Moreover, no significant difference in gene size was observed between both isolates.

Even though these plant parasites correspond to different groups (a pathogen versus a non-pathogenic isolate), most of the two isolates genes are shared, similar in gene size and intergenic length. As an example, both species of *Phytomonas* shared essentially identical membrane transport systems, with the simplest yet described amongst trypanosomatids. Analysis of these plant parasites genome sequences provides a global view of the metabolic potential of plant trypanosomatids. The genome of the two sequenced isolates consists essentially of single copy genes for the bulk of its metabolic enzymes, where *Leishmania* and *Trypanosoma* possess numerous duplicated genes or large gene families. Indeed, the comparison of these plant parasites to the other sequenced Trypanosomatids revealed a streamlined genome, encoding for a minimal system, which conserves the major complexes and pathways, indicating retention of all major organelles, but with an apparent lack of complexity. Adaptation has possibly been via minimization rather than novelty, considering a likely less complex life cycle of these parasites in their host, the plants.

Perspectives

These results will open several doors for future research programmes and applications as far as protistology, parasitology, molecular biology, biochemistry, evolution, taxonomy and phytopathology are concerned.

An important aspect of the results of this sequencing concerns the application to taxonomical changes in the former “lower trypanosomatids” (insect/plant trypanosomatids). It is known there is a need of revision of the classification of these organisms (particularly *Leptomonas*, *Crithidia*, *Herpetomonas*, and the so called “*Phytomonas*”). The sequencing of one or two more isolates from the other groups of plant trypanosomatids (10 identified so far) and two or three *Herpetomonas* isolates, known to multiply in insects, plants and mammals, included immuno-suppressed patients, would likely bring paramount understanding in the limits, if there are, between what were called, “monoxenous insect trypanosomes” like the *Herpetomonas* and “plant trypanosomes”.

On the agronomic/phytopathology aspects, efforts must be focused on further comparative analyses of their metabolic chains and hypothetic genes involved in their pathogenicity which will eventually result in better and safer control methods because so far there are no satisfactory control methods for trypanosomatid diseases of palms. New targets must be tried inside the sequences at different levels, according the results obtained in order to better fight palm diseases with non polluting and more direct methods.

The knowledge of the genes involved in the different aspects of metabolism will help for improvement of the *in vitro* culture protocols – mainly primocultures- of these parasites, because at the present time they are very fastidious to culture. Most of the research laboratories in parasitology have abandoned the idea to work on pathogenic *Phytomonas* because they are too difficult to culture – and time consuming! New adapted media will be designed in order to facilitate their *in vitro* cultures. These new media could also be new opportunities for the *in vitro* culture of other fastidious or non culturable intraphloemic organisms like proteobacteria or phytoplasmas associated with hundreds of severe diseases of plants.

Publications / Congress

Dollet M., Couloux A., Noël B., Aury J.M, Jaillon O., Fabre S., Da Silva C., Artiguenave F., Poulain J., Wincker P. Current status of the sequencing project on two plant trypanosomatids (*Phytomonas spp.*). XIIth International Congress of Parasitology (ICOPA). Melbourne, Australia. 15-20 August 2010.

Dollet M., Couloux A., Noel B., Aury J.M., Jaillon O.,Fabre S., Da Silva C., Artiguenave F., Poulain J., Wincker P. 2010.Genome sequencing project has begun on two plant trypanosomatids (“*Phytomonas*”). British Society for Parasitology, Spring Meeting and Trypanosomiasis & Leishmaniasis Seminar. Cardiff University, Wales, UK. March 29th- April 1rst.

Total permanent scientist

CIRAD: 1 scientist, one technician
Genoscope: 4 scientists

Temporary contracts

CIRAD : 1 technician, Ophélie Dagail from February 2009 to October 2009.
Genoscope : Maria Bernard from August 2010 to February 2012.