First complete chromosomal organization of a protozoan plant parasite (Phytomonas spp.)

Clotilde Marin a,b,1, Blandine Albergeb, Michel Dolletta, Michel Pagèsb, Patrick Bastien b,*

a CIRAD, Department Biological Systems, Research Unit 29 Etiology of Wilts, TA A-29/F, 34398 Montpellier Cedex, Montpellier, France
b CNRS/Université Montpellier I FRE 3013 "Biologie Moléculaire, Biologie Cellulaire et Biodiversité des Protozoaires Parasites,” Laboratoire de Parasitologie-Mycologie, UFR Médecine, 163 Rue Auguste Broussonet, F-34090 Montpellier, France

Received 24 May 2007; accepted 26 July 2007
Available online 26 November 2007

Abstract

Phytomonas spp. are members of the family Trypanosomatidae that parasitize plants and may cause lethal diseases in crops such as Coffee Phloem necrosis, Hartrot in coconut, and Marchitez sorpresiva in oil palm. In this study, the molecular karyotype of 6 isolates from latex plants has been entirely elucidated by pulsed-field gel electrophoresis and DNA hybridization. Twenty-one chromosomal linkage groups constituting heterologous chromosomes and sizing between 0.3 and 3 Mb could be physically defined by the use of 75 DNA markers (sequence-tagged sites and genes). From these data, the genome size can be estimated at 25.5 (±2) Mb. The physical linkage groups were consistently conserved in all strains examined. Moreover, the finding of several pairs of different-sized homologous chromosomes strongly suggest diploidy for this organism. The definition of the complete molecular karyotype of Phytomonas represents an essential primary step toward sequencing the genome of this parasite of economical importance.

© 2007 Elsevier Inc. All rights reserved.

Keywords: Phytomonas; Trypanosomatids; Genome; Chromosomes; Molecular karyotype; Genes; STS

Introduction

The genus Phytomonas was born in 1909 to classify trypanosomatids isolated from plants when they were observed in the latex tubes of some Euphorbiaceae [1]. Subsequently, the term Phytomonas spp. has been used for digenetic trypanosomatids isolated from different plant tissues in a wide variety of plants, including fruits and seeds, with a wide geographical distribution [2,3]. Phytomonas are grouped in the rank Trypanosomatida [4] and appear evolutionarily more closely related to Leishmania spp. and monoxenous insect trypanosomatids than to Trypanosoma spp. [5].

In plants, trypanosomatids multiply in different kinds of tissues and organs and are associated with lethal syndromes such as Coffee Phloem necrosis [6], coconut Hartrot [7], and oil palm Marchitez sorpresiva [2]. These vector-borne trypanosomatid infections can inflict considerable economic losses and have important ecological implications because of intensive crop treatment with insecticides. The classification of Phytomonas remains obscure, but has been recently clarified using molecular data. Comparison of sequences of SL RNA [8–10] and 5S rRNA genes [11] as well as, more recently, of minicircle conserved regions [12] and of the ITS of the ribosomal operon (Dollet, Sturm, and Campbell, unpublished data) allowed the identification of at least 8 (more probably10) taxonomic groups.

Although the genomes of three trypanosomatid parasites causing human diseases (Leishmania major, Trypanosoma brucei, and T. cruzi) have been entirely sequenced [13], the genetics and molecular biology of Phytomonas are almost completely ignored. Only a few gene sequences are available in sequence databases and basic data, such as genome size and organization and ploidy, are not known.
Determining the “molecular karyotype” of a protozoon is a fruitful “pregenomics” step toward systematic sequencing of the genome. It allows mapping of DNA markers and contig anchoring on defined chromosomes, thus generating invaluable information during the process of genome sequence assembly. The term molecular karyotype (MK) was forged when chromosomes from single-celled organisms, that do not condense during mitosis, could be separated and visualized using pulsed-field gel electrophoresis (PFGE) [14]. The MK cannot be reduced to a mere PFGE profile but, rather, should include the identification of heterologous chromosomes as physical linkage (or synten) groups, using DNA markers as tags. Although they reflect the large-scale organization of the genome [15], MKs usually show a high degree of polymorphism between different strains, including of the same species [15–18], this being mostly due to size variations among homologous chromosomes [19]. Yet, a high level of conservation of the synten groups among isolates of the same species is usually observed [15–18,20]. Such a conservation, combined with interstrain chromosome size polymorphisms, has actually been used as a tool to determine the genomic organization of a number of pathogenic protozoa [17,21–23].

Here, to increase our knowledge about plant trypanosomatids genetics and to pave the way toward systematic sequencing of their genome, we have used a similar strategy to define the complete molecular karyotype of 6 Phytomonas isolates from latex of Euphorbia. This includes the identification of all homologous and heterologous chromosomes with the help of 75 DNA markers.

Results

Ethidum bromide-stained molecular karyotype of Phytomonas sp.

The complete resolution of the Phytomonas karyotype was obtained using three different PFGE running conditions. Genome sizes ranged from 0.3 to 3 Mb (Figs. 1A and B). Sixteen to 21 nonstoichiometrically staining bands were clearly individualized, 2 to 4 bands of which showed a higher (“double”) staining intensity. The uneven distribution of the ethidium bromide fluorescence intensity among all bands suggested the presence of more than one chromosome comigrating in these bands. If each single and each double band represented 1 and 2 chromosomes, respectively, the number of chromosomes would be 20 in E.M.8 and E.hi Ind, 21 in E.M.1, 22 in E.M.2 and E.C.2, and 25 in E.C.3. This made clearly necessary the identification of heterologous chromosomes in this genome by chromosome-specific markers.

Identification of heterologous and homologous chromosomes

The strategy used for identifying chromosomes was based on the identification of synten groups with the use of chromosome-specific DNA markers [17,20,21]. A total of 216 genomic DNA clones from our reference strain E.M.1 were screened by hybridization onto Southern blots of MKs of different Phytomonas isolates. Out of these, 72 were found specific for one (or sometimes 2 or 3) chromosomal bands in each of the karyotypes (Fig. 1C). On top of these, three known genes (namely iso-
propanol dehydrogenase (iPDH), SL RNA, and 5S rRNA genes) were PCR-amplified and the PCR products used as probes.

Two or more markers constantly hybridizing on the same chromosomal band in the six karyotypes defined a synteny group [17,20,21]. Specific size variations were observed for each synteny group among the six isolates (Fig. 1C) and pairs of different-sized homologues in the same karyotype were seen as “doublets.” This allowed us to infer that each synteny group corresponds to one heterologous chromosome, as in previous protozoan genome projects. In total, the 75 DNA markers enabled us to identify 21 synteny groups, therefore heterologous chromosomes, in the *Phytomonas* genome (Table S1, Fig. 2). The chromosomes were arbitrarily numbered by order of increasing size in our reference strain E.M.1. A few synteny groups were difficult to individualize because, in most MKs, they were comigrating with one or more others (e.g., chromosomes 6–7–8, Fig. 2); each of these was considered as a specific chromosome from the fact that their entire linkage group clearly hybridized onto a distinct chromosomal band in at least one MK. It is noteworthy that four chromosomes could be identified in this way from the sole polymorphic MK of a geographically distant isolate, E. hi. Ind: these are chromosome 7 and 8, on the one hand, and 17 and 18, on the other hand (Fig. 2).

![Fig. 2. Schematic representation of the complete karyotypes of six representative *Phytomonas* isolates from latex: from left to right, E.M.1, E.M.2, E.M.8, E.C.2, E.C.3, and E.hi.Ind. Each chromosome is indicated by a different color code and each box represents a chromosomal band (that may contain more than one chromosome). Molecular size markers (LmjF chromosomes) are indicated on the left in kb [from http://www.ebi.ac.uk/parasites/LGN/chromsum.html]. The scale is not directly proportional, but tries to represent the actual resolution seen on different gels for different-size classes.](image-url)
Considering the chromosomes of the *L. major* genome project reference strain (LmjF) as molecular size markers (Fig. 2), we could estimate the minimum size of the haploid genome of our isolates as 25.5 Mb for *E.M.1*, 26.2 Mb for *E.M.2*, 24.9 Mb for *E.M.8*, 24.5 Mb for *E.C.2*, 29.7 Mb for *E.C.3*, and 21.4 Mb for *E hi.*.

Among the 72 anonymous chromosome-specific markers, 49 were sequenced. Careful manual sequence annotation showed that these comprise a number of genes, e.g., malate dehydrogenase (MDH), beta-tubulin, rDNA (18S or SSU, and 28S or LSU-alpha), fructose-1,6-bisphosphatase, transaldolase, and malic enzyme (Table S2). On top of those, a number of DNA markers were sequenced at first but did not hybridize onto the karyotypes: 11 of them whose annotation yielded significant TBLASTX hits are shown in Table S2. Interestingly, the 18S rDNA and SSU and 28S (LSU-alpha) rRNA genes are dispersed on 15 out of 21 chromosomes (Table S1), whereas the 5S rDNA and SL are present as single loci on chromosomes 4 and 19, respectively. Of note also is the presence of microsatellites in some of the markers screened and sequenced, in particular a (CA)\textsubscript{15} (not shown), this being the first report of these simple repeats in this organism.

**Discussion**

For the first time, we have established the complete molecular karyotype of *Phytomonas* parasitizing latex of *Euphorbia* by identifying all homologous chromosomes as syntenic groups. The genome comprises 21 homologous chromosomes, size-distributed evenly between 0.3 and 3 Mb. This genomic organization is relatively close to that of *Leishmania* spp. or *T. cruzi* that, however, exhibit a higher number of chromosomes, i.e., 34 to 36 and >55, respectively [22–25]. It differs widely from that seen in *T. brucei* that comprises only 11 Mb chromosomes, 2 intermediary chromosomes, and ≈100 “minichromosomes” [26]. From our data, the total haploid genome of *Phytomonas* can be estimated as 25.5 Mb (±2 Mb depending on strain), and the mean G/C rate as 50%. The relatively large size differences observed between isolates are due to the presence of homologous chromosomes with different sizes (see below): in the Mb range, these size differences may reach several hundreds of kb each. Such genome size differences have also been reported for other protozoa, in particular for *Leishmania* spp. [22] and *Trypanosoma* spp. [25]. As a comparison, the genome size of *T. brucei* is about 25 Mb [26], that of *Plasmodium falciparum* 25–30 Mb, that of *P. vivax/ovale/malariae* 35–40 Mb [27], that of *Leishmania* spp. 33–37 Mb [22], and that of *Eimeria* spp. about 60 Mb [28] (see also http://www.sanger.ac.uk/Projects/Protozoa/).

Our data were established using 75 DNA markers, 52 of which constitute sequence-tagged sites (STSs). These STSs will undoubtedly be useful in contig mapping in a future genome project for *Phytomonas*. Among the DNA clones sequenced, 29 new *Phytomonas* genes could be identified with relative confidence. The rest of the nucleotide sequences did not yield any significant BLAST hits, suggesting, therefore, that they are highly divergent and that they may be either noncoding sequences or highly specific genes.

A number of DNA markers mapped to repeated loci present on two (9 markers) or three (3 markers) homologous chromosomes (indicated by an asterisk in Table S1). Most of these duplications were not associated on the same linkage groups, suggesting that they are not large and that the *Phytomonas* genome is probably not very redundant.

Our data also revealed the genomic organization of rRNA genes in *Phytomonas*. The dispersion of the 18S (SSU) and 28S (LSU) rRNA genes (likely representing the typical eukaryotic rDNA locus: highly repeated and organized in head-to-tail tandem arrays) on 15 out of 21 chromosomes is reminiscent of the organization seen in *T. brucei* [17], *T. cruzi* [13], and *Giardia lamblia* [29] but highly different from the single locus localized in the chromosome 27 in *L. major* [30]. On the other hand, the 5S rDNA is present as a single locus, as in *T. brucei* and *T. cruzi*, but different than the gene dispersion observed in *L. major* [13]. Similarly, the beta-tubulin gene was found on a single chromosome here, as in *T. brucei* [17] whereas it is present on three chromosomes in *Leishmania* spp. [13] and more dispersed in *T. cruzi* [31]. *Phytomonas* sp. have a single locus encoding the SL RNA, as in the other known trypanosomatid genomes [13,17,31].

Finally, we report here for the first time the presence of microsatellites in *Phytomonas*. Several other microsatellites, of two or three bases [(CTT)n, (CT)n, (TAG)n, (GC)n, and (GT)n], have recently been found in the internal spacers (ITS 1 and ITS 2) between the SSU and the LSU genes of several trypanosomatids isolated from plants (Dollet, Sturm, and Campbell, submitted for publication).

As seen in all protozoan genomes studied to date [16,18,22,28,32–36], chromosomal size variations were observed among the different isolates. These were not as extensive as those seen in *Trypanosoma* spp. [25,26] but rather within the range of those seen in *Leishmania* spp. [22]. However, this should be relativized by the fact that the isolates analyzed here were geographically closely related. The only isolate originating from a widely different geographical area (India) exhibited larger size variations (e.g., chromosome 8, Fig. 2). In other protozoa, such chromosomal size polymorphisms are mostly attributable to amplification/deletion events in subtelomeric repetitive regions, or more seldom, in internally located, tandemly repeated genes. However, the immature state of knowledge on *Phytomonas* chromosome or gene organization does not permit us to draw any conclusion about this matter at this stage.

Finally, the presence of several “pairs” of different-sized homologous chromosomes in these karyotypes suggests that *Phytomonas* sp. is “mainly” diploid, like *Leishmania* spp. and other trypanosomatids [37].

The whole body of the data presented here (genome size, G/C rate, ploidy, number and size of chromosomes, and chromosome-mapped STSs) constitutes essential preliminary knowledge for a future systematic genome sequencing project on *Phytomonas* sp. Such a project would be extremely useful.
for a better knowledge and future control of *Phytomonas*-related plant diseases.

**Materials and methods**

**Organisms and in vitro culture**

The *Phytomonas* isolates from latex used in this study (Table 1) all belong to group D [11]. Five originated from the area of Montpellier (France) while one was isolated in India. They were routinely cultured in Grace’s Insect medium (Gibco) supplemented with 10% heat inactivated fetal bovine serum at 28 °C. All strains were checked by sequencing of the Spliced Leader RNA (SL) and 5S ribosomal RNA genes against the sequences published previously [8,11]. *Leishmania major* “Friedlin” (MHOM/IL/81/FRIEDLIN) was grown in RPMI 1640 supplemented with 10% heat inactivated fetal bovine serum at 28 °C.

**DNA preparation and PFGE**

For preparation of chromosomal DNA, all strains were cultivated up to a final concentration of \( 7 \times 10^8 \) cell ml\(^{-1}\). DNA agarose blocks were prepared and processed and PFGE was carried out on home-made devices as described in [23]. PFGE was performed at 15 °C in 1.5% Seakem GTG (FMC) agarose gels with 0.5X TBE running buffer (1X TBE=89 mM Tris-HCl/89 mM boric acid/2 mM EDTA, pH 7.4). The voltages and pulse conditions necessary to obtain a fine resolution for every chromosomal size class are described in the legend to Fig. 1. As chromosome size markers, we used the chromosomes of the reference strains (Table 1) available at http://www.ebi.ac.uk/parasites/LGN/chromosome.html. For very large size classes, we also used *Schizosaccharomyces pombe* genomic DNA (Bio-Rad). After migration, the gels were stained with 0.25 µg/ml ethidium bromide, photographed under UV transillumination, and transferred onto nylon filters (Hybond N\(^{+}\), Amersham) by alkaline transfer, according to the manufacturer’s instructions.

**DNA probes and hybridization**

Two sources of marker DNA probes were used to establish the linkage groups. First we constructed a genomic library consisting of genomic DNA from strain E.M.1 digested with HpaII, and the resulting fragments were ligated into plasmid vector pBluescript (Stratagene) digested with Clal. The recombinant plasmids were transformed into *Escherichia coli* JM109 by standard techniques [38]. Four hundred colonies harboring recombinant plasmids were chosen at random and a “minipreparation” of DNA was performed by standard techniques [38]. Single fragments with a molecular weight ranging from 1 to 2 kb were selected after electrophoresis in 1% agarose gel and the DNA was purified according to standard techniques [38]. The clones from this library were termed Phi-EM followed by a number. A second source of marker DNA probes was obtained by PCR amplification from E.M.1 genomic DNA using primers specific for known gene sequences: isopropyl alcohol dehydrogenase (iPDH) [39], SL RNA [10], and 5S rRNA [11]. Markers probes were radioactively labeled by random primed synthesis [40]. Following overnight hybridization at 65 °C, Southern blots were washed at 65 °C to a final stringency of 0.1X SSPE (1X SSPE: 180 mM NaCl/10 mM Na\(_2\)HPO\(_4\), pH 7.5/1 mM EDTA) containing 0.1% SDS. For repeated probing of the same filter, blots were stripped at 50 °C using 1 M NaOH, followed by neutralization with 0.2 M Tris-HCl, pH 7.5/0.1X SSPE/0.5% SDS [20].

**DNA sequencing and sequence annotation**

A total of 60 Phi-EM markers were double sequenced by MWG-Biotech (France) and Genome Express (France) using primers T3 and T7. DNA sequences were examined for homology with known genes in the NCBI (National Center for Biotechnology Information) database [http://www.ncbi.nlm.nih.gov/BLAST/Blast.cgi?%3FCMD%3DWeb%26PAGE_TYPE%3D BlastHome], using the TBLASTX program.

**Acknowledgments**

We thank Sandrine Fabre for expert technical help in cultivating *Phytomonas* isolates and Lucien Crobu for help in genomic DNA libraries and sequence annotation. We also acknowledge the help of Christine Blaineau and Laurence Lachaud for mass cultivation of *Leishmania major*, and Valérie Gil and Yves Balard for their kind assistance in the preparation of the figures. Clotilde Marín was the recipient of a grant from the Postdoctoral Program of the University of Granada (Spain).

**Appendix A. Supplementary data**

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ygeno.2007.07.015.

**References**

[1] C. Donovan, Kala-azar in Madras, especially with regard to its connexion with the dog and the bug (Conorrhinus), Lancet 177 (1909) 1495–1496.

**Table 1**

<table>
<thead>
<tr>
<th>Isolate code name</th>
<th>Plant/vector</th>
<th>Country/year</th>
</tr>
</thead>
<tbody>
<tr>
<td>E.M.1</td>
<td><em>Euphorbia</em> <em>pinea</em></td>
<td>Montpellier, France/1980</td>
</tr>
<tr>
<td>E.M.2</td>
<td><em>E. pinea</em></td>
<td>Montpellier, France/1982</td>
</tr>
<tr>
<td>E.M.8</td>
<td><em>E. pinea</em></td>
<td>Montpellier, France/1998</td>
</tr>
<tr>
<td>E.C.2</td>
<td><em>E. characias</em></td>
<td>Montpellier, France/1982</td>
</tr>
<tr>
<td>E.C.3</td>
<td><em>E. characias</em></td>
<td>Montpellier, France/1987</td>
</tr>
<tr>
<td>Ehli.Ind.</td>
<td><em>E. hirta</em></td>
<td>Andhra Pradesh, India/1988</td>
</tr>
</tbody>
</table>


