Infection, Genetics and Evolution 12 (2012) 539–548

Contents lists available at SciVerse ScienceDirect

Infection, Genetics and Evolution

journal homepage: www.elsevier.com/locate/meegid

Horizontal gene transfer confers fermentative metabolism in the respiratory-deficient plant trypanosomatid *Phytomonas serpens*

Susan Ienne [a], Georgios Pappas Jr. [b,c], Karim Benabdellah [d,1], Antonio González [d], Bianca Zingales [a,*]

[a] Departamento de Bioquímica, Instituto de Química, Universidade de São Paulo, Av. Prof. Lineu Prestes, 748, São Paulo, SP 05508-000, Brazil
[b] Programa de Ciências Genômicas e Biotecnologia, Universidade Católica de Brasília, Av. W5 Norte – SGAN Quadra 916 Asa Norte, Brasília, DF 70790-160, Brazil
[c] Embrapa Recursos Genéticos e Biotecnologia, PqEB, Av. W5 Norte, Brasília, DF 70770-917, Brazil
[d] Instituto de Parasitología y Biomedicina. IPBLN-CSIC. Parque Tecnológico de Ciencias de la Salud, Av. del Conocimiento, s/n, 18100 Armilla (Granada), Spain

ARTICLE INFO

Article history:
Received 8 August 2011
Received in revised form 11 January 2012
Accepted 14 January 2012
Available online 25 January 2012

Keywords:
Trypanosomatids
Genealogy
Pyruvate decarboxylase
Horizontal gene transfer
γ-Proteobacteria
Ethanol
NADH oxidation

ABSTRACT

Among trypanosomatids, the genus *Phytomonas* is the only one specifically adapted to infect plants. These hosts provide a particular habitat with a plentiful supply of carbohydrates. *Phytomonas* sp. lacks a cytochrome-mediated respiratory chain and Krebs cycle, and ATP production relies predominantly on glycolysis. We have characterised the complete gene encoding a putative pyruvate/indolepyruvate decarboxylase (PDC/IPDC) (548 amino acids) of *P. serpens*, that displays high amino acid sequence similarity with phytobacteria and *Leishmania* enzymes. No orthologous *PDC/IPDC* genes were found in *Trypanosoma cruzi* or *T. brucei*. Conservation of the *PDC/IPDC* gene sequence was verified in 14 *Phytomonas* isolates. A phylogenetic analysis shows that *Phytomonas* protein is robustly monophyletic with *Leishmania* spp. and *C. fasciculata* enzymes. In the trees this clade appears as a sister group of indolepyruvate decarboxylases of γ-proteobacteria. This supports the proposition that a horizontal gene transfer event from a donor phytobacteria to a recipient ancestral trypanosome has occurred prior to the separation between *Phytomonas*, *Leishmania* and *Crithidia*. We have measured the PDC activity in *P. serpens* cell extracts. The enzyme has a Km value for pyruvate of 1.4 mM. The acquisition of a PDC, a key enzyme in alcoholic fermentation, explains earlier observations that ethanol is one of the major end-products of glucose catabolism under aerobic and anaerobic conditions. This represents an alternative and necessary route to reoxidise part of the NADH produced in the highly demanding glycolytic pathway and highlights the importance of this type of event in metabolic adaptation.

1. Introduction

Kinetoplastids are a remarkable group of protists that includes pathogens of invertebrates, vertebrates and even plants. The genus *Phytomonas* designates trypanosomatids found in latex, phloem, fruits and seeds of different plant species with a wide geographical distribution (reviewed by Dollet, 1984; Camargo, 1999). *Phytomonas* are thought to be transmitted by phytophagous insects and, accordingly, are classified as dixenous trypanosomatids.

Besides *Phytomonas* other trypanosomatids of the *Leptomonas*, *Herpetomonas* and *Crithidia* genera have been isolated from plants (Conchon et al., 1989). These monogenetic organisms are insect-specific and their presence in fruits, seeds and flowers results from accidental inoculation into the insect feeding sites (Conchon et al., 1989; Jankevicius et al., 1989; Serrano et al., 1999a; Fiorini et al., 2001). Genetic markers allow the distinction of the *Phytomonas* genus from the monogenetic insect kinetoplastids (Teixeira et al., 1996; Muller et al., 1997; Serrano et al., 1999b; Sturm et al., 2007).

Multiple horizontal gene transfer (HGT) events have been implicated in the acquisition of structural and biochemical peculiarities of the trypanosomatids that allowed the adaptation of these organisms to a vast range of different hosts (Opperdoes and Michels, 2007). The authors speculate that these genes were probably originated from bacterial donors that could have been endosymbionts or organisms resident in the midgut of an insect host.

Accordingly, profound metabolic adaptations are expected to permit *Phytomonas* to thrive in the phloem and laticiferous tubes of host plants. Due to the facility of *in vitro* cultivation, some characteristics of *Phytomonas* metabolism have been investigated in detail for the tomato isolate *Phytomonas serpens*. These studies concluded that several mitochondrial functions, such as cytochrome-mediated respiration, ATP production and Krebs cycle, are missing, and cell energetics are based predominantly on glycolysis (Maslov et al., 1999; Nawathean and Maslov, 2000). Another distinctive characteristic of *Phytomonas*, shared only by *Crithidia*, is ethanol production

* Corresponding author. Tel.: +55 11 30912686; fax: +55 11 38155579.
E-mail address: zingales@iq.usp.br (B. Zingales).
[1] Present address: GENYO (Pfizer-University of Granada-Andalusian Government Centre for Genomics and Oncological Research), Armilla, Granada, Spain.

1567-1348/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.meegid.2012.01.016

as one of the major end-products of glycolysis (Cazzulo et al., 1985; Chaumont et al., 1994).

Little genomic and proteomic information regarding members of the genus *Phytomonas* is available. As a first step toward the definition of the *Phytomonas* genome, the molecular karyotypes of isolates from latex and phloem have been reported (Marín et al., 2008, 2009). Aiming to expand basic knowledge on the biology and metabolism of *Phytomonas*, our group undertook a moderate gene survey by means of the generation, sequencing and analysis of *P. serpens* expressed sequence tags (ESTs) (Pappas et al., 2005). A number of enzymes of metabolic pathways were identified, as well as several ESTs that can shed some light on physiological aspects of the parasite lifestyle.

The 540-bp consensus sequence of one EST cluster (cluster 53) showed high sequence similarity at the amino acid level with pyruvate/indole-pyruvate decarboxylases of phytobacteria. Pyruvate decarboxylases (PDC; EC 4.1.1.1) are key enzymes in alcoholic fermentation, functioning as a "gateway" leading from glycolysis to fermentation, while indolepyruvate decarboxylases (IPDCs; EC 4.1.1.74) are key enzymes in the biosynthetic pathway of indole-3-acetic acid of many bacteria (reviewed by Koga, 1995). The two classes of decarboxylases (PDC and IPDC) display high structural similarities, but different affinities for their substrates (Schütz et al., 2003).

The presence of *P. serpens PDC/IPDC* gene displaying high similarity to genes of phytobacteria raised the stimulating hypothesis that a HGT event may have occurred.

In this study we have characterised the complete gene encoding a putative PDC/IPDC of *P. serpens*, shown that this sequence is conserved in *Phytomonas* members and obtained phylogenetic evidence of a probable HGT event providing the molecular basis for alcoholic fermentation in *Phytomonas*.

2. Material and methods

2.1. Trypanosomatid isolates, cultivation and DNA extraction

Phytomonas isolates used in this study were obtained from two cryopreserved collections of the Universidade de São Paulo (USP, São Paulo, Brazil) (Table 1). PCR assays targeted to the spliced-leader region were used to identify *Phytomonas isolates* (Catarino et al., 2001 and references therein). Information on other trypanosomatids used in this study is provided in Supplementary material (Table A1). *P. serpens* and *L. major* stocks were kindly provided, respectively, by Dr. Marta M. G. Teixeira (Instituto de Ciências Biológicas, USP) and Dr. Lucile Floeter-Winter (Instituto de Biociências, USP). Promastigotes of *P. serpens* were grown at 28 °C in Grace's insect medium (Sigma) supplemented with 10% foetal calf serum; promastigotes of *L. major* were grown at 25 °C in 199 medium (Vitrocell) supplemented with 10% foetal calf serum. *P. serpens* and *L. major* genomic DNA was prepared by the conventional phenol-chloroform extraction method. DNA preparations of other trypanosomatids were kindly provided by Dr. Marta M. G. Teixeira (Instituto de Ciências Biológicas, USP).

2.2. PCR amplification

Three pairs of primers were employed for the amplification of *P. serpens PDC/IPDC* gene (Table 2; Fig. 1) designed based on the nucleotide sequence of this gene here determined (GenBank JN400885). For the amplification of *P. serpens* housekeeping gene *glycosomal glyceraldehyde dehydrogenase (gGAPDH)* the pair of primers Ps.GAPDH.For/Ps.GAPDH.Rev (Table 2) was designed based on the nucleotide sequence (GenBank No EU084892). Amplification reactions were performed in 25 μL total volume, containing 50 ng DNA, 1 unit *Taq* DNA polymerase Biolase (Bioline), 1× enzyme buffer, 1.5 mM $MgCl_2$, 50 mM KCl (for Ps.IPDC.N.For/Ps.IPDC.N.Rev and Ps.IPDC.M.For/Ps.IPDC.C.Rev primers), 0.2 μm dNTPs, 0.4 μM of each primer, 5% DMSO (for Ps.IPDC.U.For/Ps.IPDC.U.Rev primers). PCR conditions included 1 min denaturation at 95 °C, followed by 30 cycles at three temperatures: 1 min denaturation at 95 °C; 1 min primer annealing at the temperature specified in Table 2; and 1 min elongation at 72 °C; followed by 9 min elongation at 72 °C. PCR products were separated in agarose gels and stained with ethidium bromide.

2.3. Cloning and sequencing of the PDC/IPDC gene

The 2400 bp insert of clone pBSSK$^+$ p228, belonging to the EST cluster 53 was fully sequenced by standard procedures using subclones obtained with a number of restriction enzymes. Sequencing was carried out with an ABI 377 automated DNA sequencer (Applied Biosystems Inc.).

2.4. Sequencing of the PDC/IPDC intergenic region

The intergenic region was amplified from *P. serpens* genomic DNA with the Ps.IPDC.U.For/Ps.IPDC.U.Rev primers (Table 2; Fig. 1) and *Pfu* DNA polymerase (Fermentas). The amplification

Table 1
Characteristics of *Phytomonas* isolates used in this study.

Isolate	Identification[a]	Host		
		Species	Source	Country
P. serpens	TCC 060	*Lycopersicon esculentum* (tomato)	Fruit	Brazil
P. françai	TCC 064	*Manihot esculenta* (cassava)	Latex	Brazil
P. mcgheei 163	TCC 300	Zea mays (maize)	Seed	Brazil
1G	Trycc 56	*Nezara viridula* (insect)	NA	Brazil
412 U	Trycc 55	*Bixa orellana* (annatto)	Seed	Brazil
492 FG	Trycc 72	*Cajanus flavus* (bean)	Seed	Brazil
Bni	TCC 086	*Blepharodon nitidus*	Latex	Surinam
EC2	TCC 052	*Euphorbia characias*	Latex	France
Epi 053 (=Em1)	TCC 053	*Euphorbia pínea*	Latex	France
Jma	TCC 066	*Jatropha macrantha*	Latex	Peru
Msc 084	TCC 084	*Mandevilla scabra*	Latex	Surinam
Sam 225	TCC 225	*Solanum americanum*	Fruit	Brazil
Ser 226 (Seri 226)	TCC 226	*Solanum erianthum*	Fruit	Brazil
Svi 223	TCC 223	*Solanum viarum*	NA	Brazil

insert NA = not aware
[a] TCC: Trypanosomatids Culture Collection of the Universidade de São Paulo; TryCC: Trypanosomatids Culture Collection of the Department of Parasitology – Universidade de São Paulo.

Table 2
Primers for PCR amplification of *P. serpens* genes.

Primer name	Sequence (5′–3′)	Localisation	Product size (bp)	Primer annealing temperature (°C)
PDC/IPDC gene				
Ps.IPDC.N.For	CCTCGGAAACCTACAACGTC	N-domain[a]	403	57
Ps.IPDC.N.Rev	CAGGAGATCTGCTCGCTCAT			
Ps.IPDC.M.For	CATGAAGCCACCGAGCTT	Middle and C-domain[a]	297	57
Ps.IPDC.C.Rev	CACTCCTGCACCGTCATCT			
Ps.IPDC.U.For	AGATGCCCGTCGTTTCTCTT	C- to N- domain[a,b]	769	57
Ps.IPDC.U.Rev	TCCGAGGGCATCGATATTAG			
GAPDH gene				
Ps.GAPDH.For	GTGCTTGTGGTGAAC	N-terminal domain [NAD(P) binding fold]	269	51
Ps.GAPDH.Rev	CACGACGCGTTCGAC			

[a] Domains for the ThDP coenzyme. See localisation in Fig. 1.
[b] Primers used for the amplification of the intergenic region.

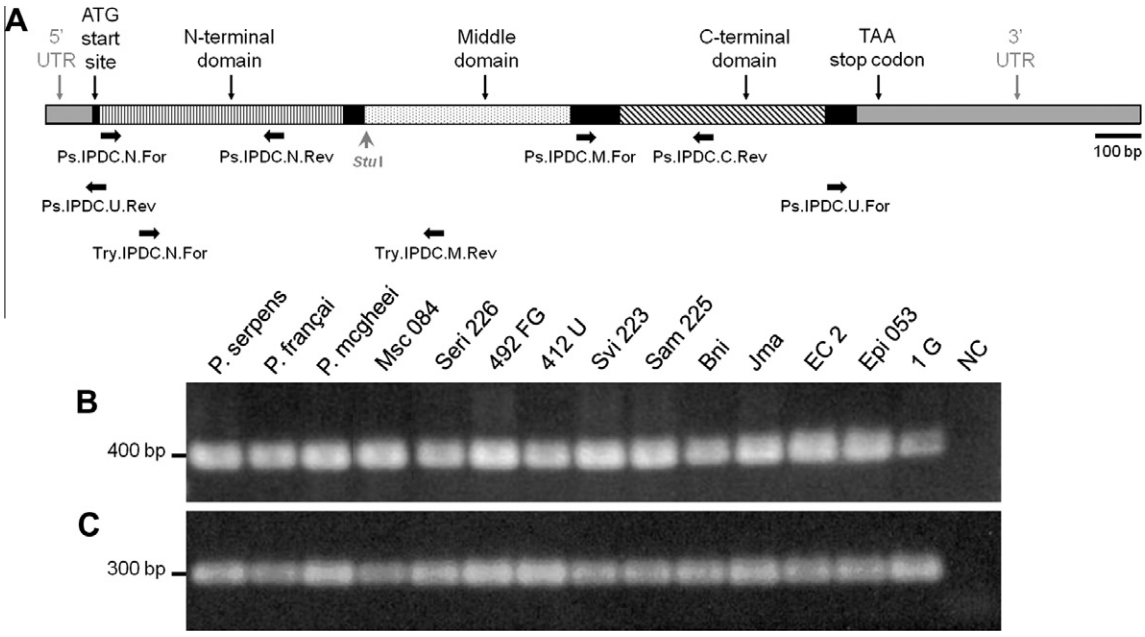

Fig. 1. *Phytomonas PDC/IPDC* gene. Schematic representation of *P. serpens PDC/IPDC* gene (Panel A). The localisation of the three ThDP binding domains and primers used for PCR amplification is indicated (see Table 2). The *PDC/IPDC* gene of *Phytomonas* isolates was PCR amplified with Ps.IPDC.N.For/Ps.IPDC.N.Rev primers (Panel B) and Ps.IPDC.M.For/Ps.IPDC.C.Rev primers (Panel C). Ethidium bromide stained agarose gels. Molecular mass markers are indicated. NC, negative control.

product was purified using a PCR purification kit (RBC – Real Biotech Corporation), incubated with dATP and *Taq* polymerase (Fermentas) and cloned into the TOPO TA plasmid (Invitrogen) according to the manufacturer's instructions. Recombinant plasmids were purified and sequenced as described above with TOPO TA primers (M13 Forward and M13 Reverse).

2.5. Phylogenetic analyses

The deduced protein sequence of *P. serpens* PDC/IPDC together with protein sequences of PDCs and lPDCs from *Leishmania* spp., bacteria, fungi and plants were subjected to phylogenetic analyses. The databases accessed for selection of the sequences were GenBank and SWISS-PROT. Preference was given to the inclusion of sequences that have been characterised experimentally, corroborating the enzymatic function. Two *PDC/IPDC* homologs were found in *C. fasciculata* genome following a search in the TriTryp database (http://tritrypdb.org/tritrypdb/) using as query *P. serpens* protein sequence: contig 653 (nucleotide position 1531–3168) and contig 1012 (nucleotide position 116893–118548). The corresponding protein sequences were included in the analysis. Sequences were aligned with CLUSTALW (Thompson et al., 1994) and visualised with BioEdit (Biological Sequence Alignment Editor for Windows) (Hall, 1999). PAUP* version 4.0b10 (Swofford, 2003) was used for the reconstruction of phylogenetic trees. The reconstructions were performed using the neighbor-joining distance method (Saitou and Nei, 1987), maximum parsimony method (Fitch, 1971), with heuristic search algorithm by stepwise addition, and with the maximum likelihood method (Felsenstein, 1981). To determine the best substitution model for the dataset the Likelihood Ratio Test (LRT) with ProtTest program (Drummond and Strimmer, 2001; Guindon and Gascuel, 2003; Abascal et al., 2005) was applied. Bootstrap values (1000 pseudo-replicates) were used for internal consistency of data in the maximum parsimony method, assessed with PAUP*. The generated trees were displayed with TreeView (version 1.6.6; Page, 1996).

2.6. Gene sequence characteristics and codon usage

The nucleotide sequences of the following genes were analysed: *P. serpens* and *L. major PDC/IPDCs* and GAPDHs; *E. cloacae* and *E. herbicola IPDCs* and *A. parasiticus* and *R. oryzae PDCs*. GenBank accession numbers are indicated in Section 3. The nucleotide sequences were subjected to codon usage analysis with the CodonW program (John F. Peden; available at http://codonw.sourceforge.net/).

2.7. Southern blot

P. serpens DNA (7 μg) was digested with *Stu* I restriction enzyme (New England Biolabs), run on 1% agarose gels and blotted to nylon membranes (Hybond-N, Amersham Biosciences) using standard protocols. Hybridisation was performed using as probe the 297-bp purified amplification product of the *P. serpens PDC/IPDC* gene obtained with the Ps.IPDC.M.For/Ps.IPDC.C.Rev primers (Table 2). The probe was labelled with [α^{32}P]dATP using a Random Primer DNA Labeling Kit (Invitrogen). Blots were hybridised at 60 °C in 0.05% Ficoll, 0.05% PVP, 0.5 mM EDTA, 3× SSC, 0.1% SDS, 100 μg/mL salmon sperm DNA. Blots were washed to a final stringency of 0.1×SSC/0.1% SDS at 60 °C. The radioactive images were recorded on Phosphor Screens (Kodak).

2.8. Relative abundance of PDC/IPDC transcripts

P. serpens and *L. major* total RNA was isolated using TRIzol® (Invitrogen) and treated with DNAse RQ1 RNAse-free (Promega) according to the manufacturer's instructions. Two micrograms of total RNA were used for cDNA synthesis with random hexamers as primers and Super Script II Reverse Transcriptase First-Strand Synthesis Kit (Invitrogen), as indicated by the manufacturer. For quality control, RNA absorbance was measured at 260 nm and individual samples were analysed by agarose gel electrophoresis. Sequences of the primer sets used for real-time PCR analysis were designed with the PrimerExpress program (Applied Biosystems) and are shown in Table 3. All real-time PCR runs were performed in triplicate and each reaction mixture (20 μL final volume) contained 1× SYBR Green PCR Master Mix (Applied Biosystems), 200 nM each primer and 96 ng of cDNA template. Amplification was performed in Real-Time PCR 7300 (Applied Biosystems) instrument with the following thermal cycling protocol: 10 min initial denaturation at 95 °C, 40 cycles of 15 s at 95 °C and 1 min at 60 °C. The fluorescence signal was measured at the end of each extension step at 60 °C. After amplification, a melting peak analysis with a temperature gradient from 60 to 95 °C was performed to confirm that only the specific products were amplified. The threshold cycle (C_t) was determined with the 7500 Software Version 2.0.1 (Applied Biosystems). In all experiments, the species-specific housekeeping gene glyceraldehyde-3-phosphate dehydrogenase was used as normaliser.

2.9. PDC activity in P. serpens cell extracts

P. serpens (10^9 cells) were disrupted by sonication. The protein concentration of the 16,000*g* supernatant was determined with the Bradford's method (Bradford, 1976). PDC activity was measured using a coupled photometric assay with alcohol dehydrogenase (ADH) as an auxiliary enzyme. The reaction mixture (1 mL) contained 10 mM potassium phosphate, pH 6.5, 5 mM $MgCl_2$, 0.35 mM NADH, 0.1 mM thiamin diphosphate (ThDP), 10 U *S. cerevisiae* ADH (Sigma), 100 μg protein *P. serpens* 16,000*g* supernatant and different pyruvate concentrations. NADH oxidation was followed at 340 nm with a Cary Win UV spectrophotometer (Varian Inc.). The reaction was followed for 10 min at 25 °C. The data were plotted and fitted according to Michaelis–Menten kinetics.

3. Results

3.1. Characteristics of P. serpens PDC/IPDC gene

Clone p228 from the *P. serpens* EST survey (Pappas et al., 2005) is one of the members of EST cluster 53, which shows homology to *PDC/IPDC* genes. This clone contains a full-size cDNA including spliced leader (SL) (also named mini-exon) and polyA sequences. The SL sequence is followed by a 102-bp 5′-untranslated region (UTR), a 1647-bp open reading frame (ORF) that encodes 548 amino acids, and a 615-bp 3′ UTR (Fig. 1A).

Protein domain searches against the Pfam database (Bateman et al., 2004) revealed three ThDP binding domains, schematically shown in Fig. 1A. The N-terminus domain (residues 6-181) binds the pyrimidine site of ThDP; the middle domain (residues 199-347); and the C-terminal domain (residues 383-530) that contains a conserved motif for binding the diphosphate region of ThDP + Mg^{2+} (Fig. 1A). The N- and C-terminal domains are present in all families of ThDP-dependent enzymes, while the middle domain is present only in the decarboxylase and transketolase families (reviewed by Duggleby, 2006).

BLASTX similarity searches against the non-redundant (nr) NCBI database indicated high similarity to pyruvate/indole-pyruvate carboxylases of *Leishmania* species and bacterial IPDCs (Table 4). Interestingly, no orthologous genes of *Leishmania* enzyme were found in the *Trypanosoma cruzi* or *T. brucei* genomes.

The similarity of *P. serpens* PDC/IPDC with *Leishmania* pyruvate carboxylases (PC; EC 6.4.1.1) was intriguing because the two classes of enzymes catalyse different reactions, both irreversible: PDCs catalyse the reaction pyruvate → acetaldehyde + CO_2, having ThDP as coenzyme, whereas PCs catalyse the reaction pyruvate + HCO_3^- + ATP → oxaloacetate + ADP + P_i , and have biotin as prosthetic group. To elucidate this aspect, the putative pyruvate/indole-pyruvate carboxylase of *L. major* was searched for conserved domains against Pfam database. Three ThDP binding domains were disclosed, indicating that the correct designation of *Leishmania* enzyme should be PDC/IPDC and not PC/IPC, as automatically annotated in the genome project (Ivens et al., 2005).

3.2. Presence of PDC/IPDC in Phytomonas isolates

To investigate the distribution of the *PDC/IPDC* gene in *Phytomonas* isolates two pairs of primers were designed: Ps.IPDC.N.For and Ps.IPDC.N.Rev primers, both located in the ThDP N-terminal domain; and Ps.IPDC.M.For primer (located in the middle domain) and Ps.IPDC.C.Rev primer (located in the C-terminal domain) (see Table 2; Fig. 1A). For all 14 *Phytomonas* isolates the two pairs of primers amplified products of the sizes observed for

Table 3
Primers for Real-Time PCR.

Primer	Sequence (5′–3′)	Localisation	Product size (bp)
Ps.IPDC.RTP.For	GGTGTTCCCGGTGACTATAACC	*P. serpens PDC/IPDC* gene; N-domain	76
Ps.IPDC.RTP.Rev	AGCAGCCAACCCAATTGAGT		
Ps.GAPD.RTP.For	AGACGATTGTGATGGGTGTGAA	*P. serpens gGAPDH* gene	70
Ps.GAPD.RTP.Rev	GACGCGTTCGACACGATGT		
Lm.IPDC.RTP.For	CACTTGATTCACCGCCATCA	*L. major PDC/IPDC* gene; N-domain	70
Lm.IPDC.RTP.Rev	CGTACGGGATGCGGATGT		
Lm.GAPD.RTP.For	GTGATCGAGTCTACTGGCCTGTT	*L. major gGAPDH* gene	71
Lm.GAPD.RTP.Rev	TTCTTCGCACCACCCTTGAT		

Table 4
List of ten best BLASTX hits against NCBI's nr database ranked by the alignment bit-score using as query *P. serpens PDC/IPDC* nucleotide sequence.

GenBank Accession	Description	E-value	Bit-score	Identity	Similarity
XP_001564563	Pyruvate/indole-pyruvate carboxylase [*Leishmania braziliensis* MHOM/BR/75/M2904][a]	6e-137	493	46	61
XP_001686429	Pyruvate/indole-pyruvate carboxylase [*Leishmania major* strain Friedlin][a]	1e-136	492	47	62
XP_001468661	Pyruvate/indole-pyruvate carboxylase [*Leishmania infantum* JPCM5][a]	4e-136	490	47	61
ZP_06639356	Indolepyruvate decarboxylase [*Serratia odorifera* DSM 4582]	2e-130	471	44	62
ZP_06191191	Thiamine pyrophosphate binding domain-containing protein [*Serratia odorifera* 4Rx13]	2e-130	471	45	64
CAX60783	Indolepyruvate decarboxylase [*Erwinia billingiae* Eb661]	2e-128	464	46	64
YP_001479638	Thiamine pyrophosphate binding domain-containing protein [*Serratia proteamaculans* 568]	9e-128	462	44	63
AAB06571	Indolepyruvate decarboxylase [*Pantoea agglomerans*]	2e-126	457	45	62
YP_051693	Indole-3-pyruvate decarboxylase [*Pectobacterium atrosepticum* SCRI1043]	2e-126	457	45	60
YP_001005545	Indole-3-pyruvate decarboxylase [*Yersinia enterocolitica* subsp. *enterocolitica* 8081]	8e-126	456	44	60

[a] As discussed in the text, the correct designation of the *Leishmania* enzyme should be pyruvate/indole-pyruvate decarboxylase.

P. serpens (Fig. 1B and C), suggesting high conservation of the *PDC/IPDC* gene sequence in this trypanosomatid genus. No amplification products were obtained with the same primers in 18 isolates of other trypanosomatid genera (*Blastocrithidia*, *Crithidia, Endotrypanum, Herpetomonas, Leishmania, Leptomonas* and *Trypanosoma*) (Table A1 – Supplementary material). The lack of amplification observed in four *Leishmania* species and *C. fasciculata* is due to the fact that the primers do not recognise the sequence of the *PDC/IPDC* genes in these organisms (data not shown), which serves as a negative control.

3.3. Phylogenetic analysis

Given the similarity of the *P. serpens* PDC/IPDC protein with bacterial IPDCs (Table 4), we performed a phylogenetic analysis to infer the origin of this sequence in *Phytomonas*. Twenty-six protein sequences of IPDCs, IPDC/PPDCs (PPDCs: phenylpyruvate decarboxylases, E.C. 4.1.1.43; Baudoin et al., 2010) and PDCs of bacteria, fungi and plants were selected for comparison. The selected sequences are summarised in Table 5; and fully described in Table A2 (Supplementary material). Preference was given to enzymes whose function has been confirmed experimentally because PDCs and IPDCs have a very similar amino acid sequence and structure (Koga et al., 1992; Schütz et al., 2003), which precludes unequivocal functional assignment based solely on sequence data. Sequences of PDCs and IPDC/PPDCs were included because both have binding domains to ThDP and Mg^{2+} similarly to IPDCs, showing significant differences only in the specificity and affinity for the substrate (Koga et al., 1992; Schütz et al., 2003; Baudoin et al., 2010). PDC/IPDC protein sequences of three *Leishmania* species and two PDC/IPDC homologs, identified in the genome of *C. fasciculata* using as query the protein sequence of *P. serpens*, were also incorporated in the analysis.

After multiple sequence alignment, phylogenetic trees were generated by maximum likelihood (ML; Fig. 2), neighbor-joining and maximum parsimony methods (NJ and MP; Fig. 3A and B, respectively). The resulting trees revealed distinct clusters of plant PDCs, fungi PDCs, bacterial IPDCs and bacterial IPDCs/PPDCs. The phylogenies suggested that *P. serpens, Leishmania* spp. and *C. fasciculata* PDC/IPDC form a monophyletic group with IPDCs of the γ-proteobacteria *P. agglomerans*, *E. cloacae* and *P. putida*, with bootstrap support of 93% in the MP tree (Fig. 3B). This raises the possibility that a HGT event from an ancestral phytobacteria was responsible for the acquisition of the *PDC/IPDC* gene in the trypanosomatids branch. Interestingly, the protein sequences of *L. major*, *L. braziliensis*, *L. infantum* and *C. fasciculata* grouped in a monophyletic clade with the *P. serpens* protein. This clade was supported by bootstrap analysis, appearing in 91% of pseudo-replicates in the MP tree (Fig. 3B), suggesting that a putative HGT event may have occurred before the separation of *Phytomonas*, *Leishmania* and *Crithidia*. Phylogenetic trees obtained with the inclusion of PDCs and IPDCs electronically annotated sequences were also constructed by NJ and ML methods (Fig. A1, Supplementary material). The trees show that the main clusters are retained.

The trees also evoke a common ancestor for the fungal PDCs and bacterial IPDCs. Considerable differences between the IPDCs/PPDCs of α-proteobacteria (*Azospirillum lipoferum* and *A. brasilense*) and the IPDCs of γ-proteobacteria (*P. agglomerans*, *P. putida* and *E. cloacae*) were in evidence, in agreement with Unweighted Pair Group Method with Arithmetic Mean (UPGMA) that showed two distinct clades, one composed by strict IPDCs and another by *A. brasilense* IPDCs/PPDCs and other related sequences (Spaepen et al., 2007a).

The trees also suggest a putative HGT event for the acquisition of *Neurospora crassa* PDC, because, as opposed to what is observed for the other fungal PDCs, *N. crassa* PDC showed a closer relationship to plant and bacteria PDCs.

Table 5
Summary of protein sequences selected for the phylogenetic analyses.

Group	IPDCs/ PPDCs[a]	IPDCs[a]	PDCs[a]
Bacteria – Proteobacteria – Alphaproteobacteria	2	0	0
Bacteria – Proteobacteria – Gammaproteobacteria	0	3	1
Eukaryota – Fungi – Ascomycota	0	0	11
Eukaryota – Fungi – Zygomycota	0	0	1
Eukaryota – Viridiplantae	0	0	8
TOTAL	2	3	21

[a] Enzymatic function confirmed experimentally.

3.4. Gene sequence characteristics and codon usage

To explore the hypothesis that lateral gene transfer has participated in the acquisition of the *P. serpens* and *Leishmania* spp. *PDC/IPDC* gene we analysed the GC content and codon usage of the gene, given that horizontally transferred DNA often has anomalous nucleotide compositions differing from the rest of the genome (Monier et al., 2007). For this purpose, two phytobacterial *IPDC*s and two fungal *PDC*s were selected (Table 6). *P. serpens* and *L. major* g*GAPDH* housekeeping genes were used as controls (Table 6). The phytobacterial and fungal species were selected based on their proximity to *P. serpens* and *Leishmania* spp. in the *PDC/IPDCs* phylogenies (Fig. 2 and 3) and to the fact that their enzymatic activity has been demonstrated experimentally. The analysis of the proportion of codons (Relative Synonymous Codon Usage – RSCU) in the ORFs of *IPDC*, *PDC* and *gGAPDH* genes was conducted with the CodonW program and is shown in Table A3 of Supplementary material. From this analysis the GC3s index was calculated, which describes the G + C content at the third position of synonymous codons (Table 6). The data indicated a similar codon usage for the *PDC/IPDC* and *gGAPDH* genes of *P. serpens* and *L. major*. The GC3s index of these genes is higher than the indexes of the bacterial and fungal genes examined.

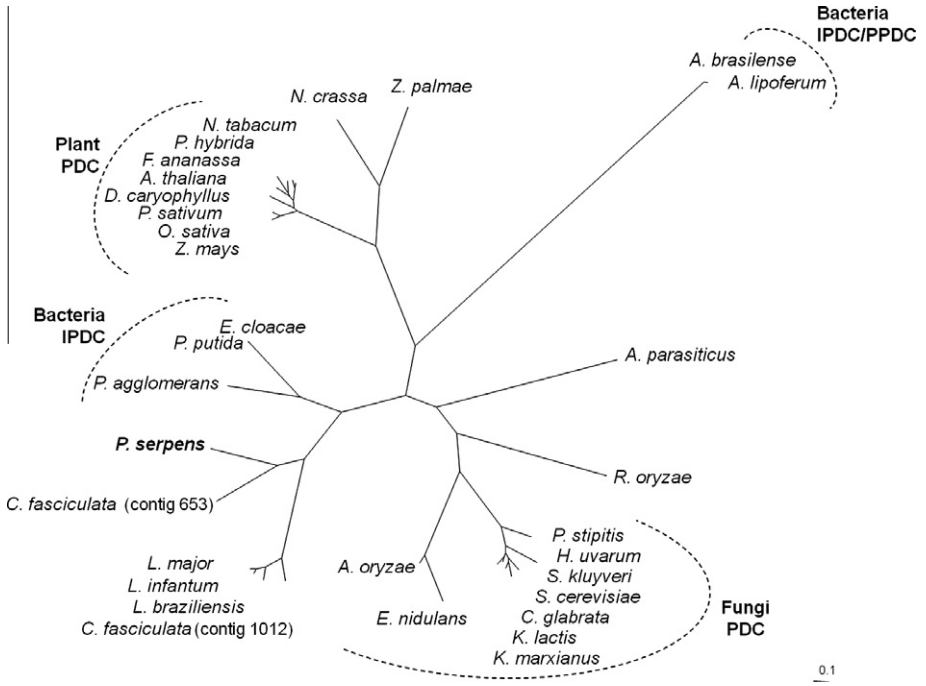

Fig. 2. Maximum Likelihood tree generated with *P. serpens* PDC/IPDC (*P. serpens*), the putative PDC/IPDC of *L. major*, *L. infantum* , *L. braziliensis*, and *C. fasciculata*, and protein sequences of IPDCs, PDCs and IPDCs/PPDCs of Bacteria, Fungi and Plants (see Table 5). The bar indicates the number of substitutions per site. (Evolutionary model: LG + I + G + F = −21627.59).

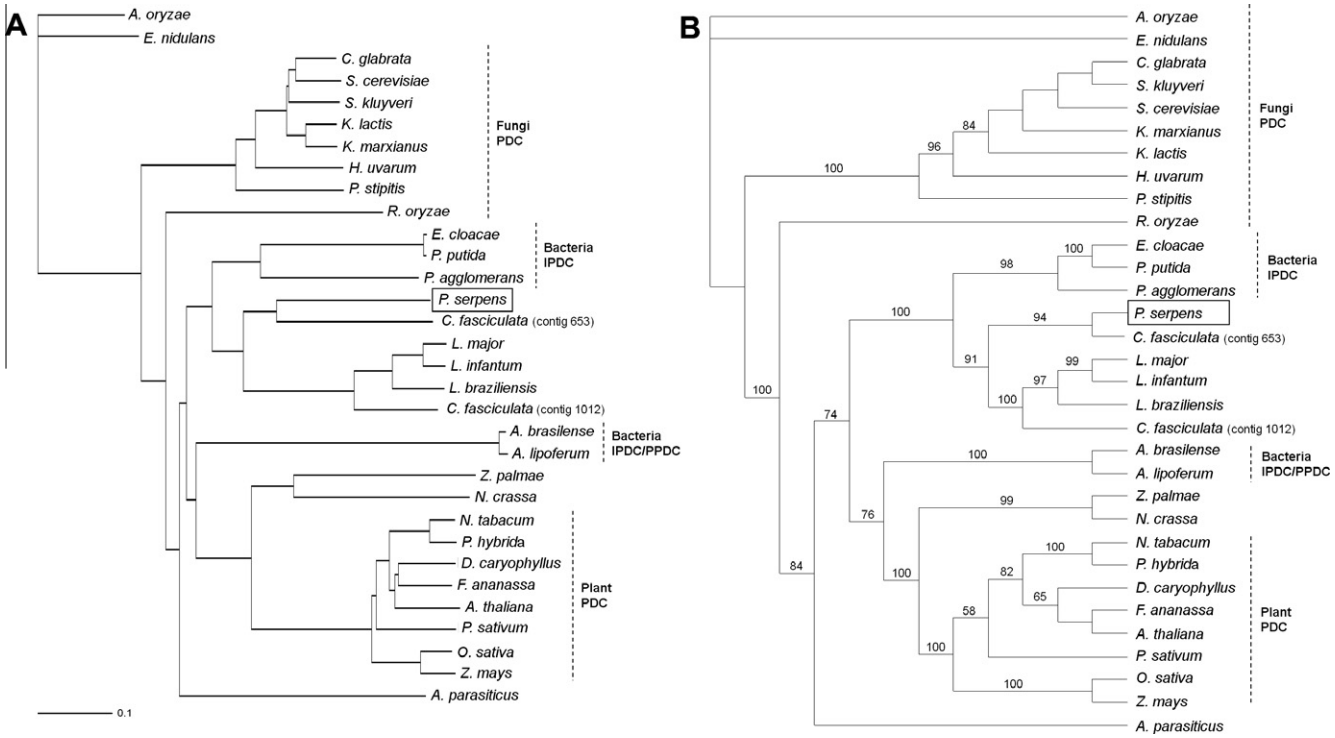

Fig. 3. Neighbor-Joining (A) and Maximum parsimony (B) trees generated with *P. serpens* PDC/IPDC, the putative PDC/IPDC of *L. major*, *L. infantum* and *L. braziliensis*, and protein sequences of IPDCs, PDCs and IPDCs/PPDCs of Bacteria, Fungi and Plants (Table 5). The bar indicates the number of substitutions per site.

3.5. P. serpens PDC/IPDC gene copy number and sequence of the intergenic region

The gene copy number of *PDC* and *IPDC* varies in different organisms. In *S. cerevisiae* and *S. kluyveri* three copies of *PDC* are reported (Hohmann, 1991; Møller et al., 2004), whereas *IPDC* is a single copy gene in *Leishmania* spp. (genome data at http://www.genedb.org) and in phytobacteria (Costacurta et al., 1994; Patten and Glick, 2002). We investigated the gene copy number in *P. serpens*. For this purpose *P. serpens* genomic DNA was digested

Table 6
GC content and codon usage indexes.

Organism	Gene	GenBank	GC[a]	GC3s[b]
P. serpens	PDC/IpdC	JN400885	0627	0849
P. serpens	gGAPDH	EU084892	0599	0758
L. major	PDC/IpdC	CAJ08046	0590	0801
L. major	GAPDH	XM_001684852	0617	0884
Enterobacter cloacae	IpdC	P23234	0567	0470
Erwinia herbicola	IpdC	ERWIPDC	0595	0617
Aspergillus parasiticus	PDC	P51844	0504	0523
Rhizopus oryzae	PDC	AF282846	0475	0446

[a] GC content of the gene.
[b] GC3s index, G + C content at the third position of synonymous codons (excluding Met, Trp and termination codons).

with *Stu* I (position 961 of the gene, indicated in Fig. 1A) for 30 min or 16 h and hybridised with the 297-bp *PDC/IPDC* probe (Table 2; Fig. 4A). This probe identified a series of bands of similar intensity (a 2.5-Kb band, followed by five visible bands at multiples of 2.5 Kb) in the 30-min digested DNA (Fig. 4A, lane a) and a prominent 2.5-Kb band in the 16-hour digestion (Fig. 4A, lane b). This result is indicative of at least six tandemly-repeated gene copies.

Next the size and sequence of the intergenic region was determined. *P. serpens* genomic DNA was amplified with Ps.IPDC.U.For and Ps.IPDC.U.Rev primers (Table 2), localised, respectively on the 3′ UTR of the gene and the 5′ UTR of the putative contiguous gene (Fig. 4B). A product of ~900 bp was obtained, cloned and sequenced (GenBank JN400886 to JN400894). Sequence analysis of nine independent clones and alignment with the 3′ and 5′ *PDC/IPDC* UTRs indicated that the intergenic region is 124 bp in length. Three 3' UTR variants were obtained, differing at only three positions (Fig. 4C).

3.6. Relative abundance of PDC/IPDC transcripts

The relative abundance of *PDC/IPDC* transcripts in *P. serpens* and *L. major* was assessed by real time RT-PCR. To test the primer efficiencies, eightfold serial dilution series of linearised plasmids containing *P. serpens* or *L. major* PDC/IPDC or *gGAPDH* inserts were used to construct standard curves. The C_t values were plotted against the logarithm of their initial template copy concentrations. The standard curves were generated for each pair of primers by a linear regression of the plotted points. From the slope of each curve, PCR amplification efficiency (E) was calculated according to Lee et al. (2008, and references therein). The efficiency of the target and reference gene amplifications were similar (data not shown) and could be used to calculate the delta C_t.cDNA extracted from *P. serpens* and *L. major* were used as template for the real time PCR assays. The C_t values obtained from the *PDC/IPDC* and *gGAPDH* were used to calculate the relative abundance of *PDC/IPDC* transcripts by the delta-C_t method. The results of three biological replicates indicate a 5.7-fold increase in the abundance of gene transcripts in *P. serpens* as compared with *L. major* (Table 7).

3.7. PDC activity in P. serpens

Although the production of ethanol has been reported in *Phytomonas* sp. cultured in glucose-rich medium (Chaumont et al., 1994), a PDC activity has not been demonstrated in these organisms. Accordingly, we measured PDC activity in the 16,000*g* supernatant of *P. serpens* extracts using a coupled photometric assay with ADH as an auxiliary enzyme. Plots of initial reaction rates versus pyruvate concentrations indicate classical Michaelis-Menten kinetics (Fig. 5). Based on the kinetic data, we determined that *P. serpens* PDC has a Km value for pyruvate of 1.4 mM.

4. Discussion

Here we describe a gene in *P. serpens* encoding a PDC/IPDC enzyme potentially related to fermentation metabolism. We estimate that *P. serpens* contains at least six *PDC/IPDC* gene copies in tandem, separated by a small 124-bp intergenic sequence. The G + C content of the 1647-bp ORF is 62.5%, which is identical to the average G + C content reported for *L. major* protein-coding genes (Ivens et al., 2005). *P. serpens* PDC/IPDC protein has high similarity with the enzyme of three *Leishmania* species, C. *fasciculata* and bacterial IPDCs. No orthologous genes were found in the *T. cruzi* or *T. brucei* genomes, an intriguing observation as many protein-coding genes are conserved in the three trypanosomatid species (TriTryp, El-Sayed et al., 2005). Accordingly, this suggests a putative common origin of *Leishmania, Crithidia* and *P. serpens* gene.

PCR assays with two pairs of primers targeting the ThDP domains indicate conservation of the *PDC/IPDC* gene sequence in

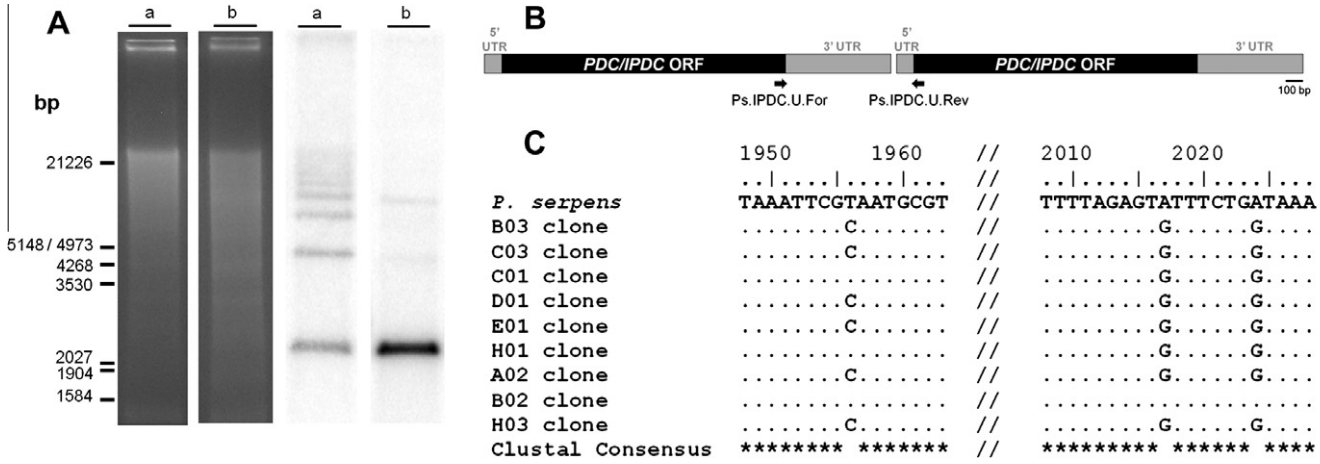

Fig. 4. Multiple copies of the *PDC/IPDC* gene. Panel A: Southern blot of *P. serpens* DNA digested with *Stu* I for 30 min (lane a) and 16 h (lane b). Ethidium bromide stained gel and and hybridisation with the 297-bp *PDC/IPDC* ^{32}P-labelled probe. Molecular mass markers are indicated in bp. Exposure time in Phosphor Screen: 24 h. Panel B: Schematic representation of two tandem *P. serpens PDC/IPDC* genes. The localisation of the primers used for the amplification of the intergenic region is indicated. Panel C: Nucleotide sequence variation in the 3′ UTR region of the *PDC/IPDC* gene copies. Nine clones derived from the amplification product obtained with Ps.IPDC.U.For and Ps.IPDC.U.Rev primers (Table 2), localised as indicated in Fig. 2B, were sequenced. The nt positions indicated in the upper line correspond to the sequence of the *P. serpens PDC/IPDC* gene (GenBank JN400885).

Table 7
Relative abundance of *PDC/IPDC* transcripts of *P. serpens* and *L. major*.

Biological replicate	cDNA	Target gene	C_t mean and SD [a]	ΔC_t and SD	$2^{-\Delta C_t}$	Fold difference (*P. serpens*/*L. major*)	Mean fold difference and SD
1	*P. serpens*	PDC/IPDC	11.19 ± 0.058	−1.08 ± 0.18	2.13	**5.3**	**5.73 ± 1.69**
		GAPDH	12.27 ± 0.178				
	L. major	PDC/IPDC	15.59 ± 0.123	1.33 ± 0.14	0.4		
		GAPDH	14.26 ± 0.071				
2	*P. serpens*	PDC/IPDC	10.59 ± 0.07	−1.80 ± 0.09	3.5	**7.6**	
		GAPDH	12.39 ± 0.06				
	L. major	PDC/IPDC	15.76 ± 0.06	1.12 ± 0.08	0.46		
		GAPDH	14.65 ± 0.06				
3	*P. serpens*	PDC/IPDC	10.36 ± 0.07	−1,03 ± 0.08	2.05	**4.3**	
		GAPDH	11.39 ± 0.03				
	L. major	PDC/IPDC	15.67 ± 0.03	1.06 ± 0.10	0.48		
		GAPDH	14.61 ± 0.10				

[a] SD: standard deviation.

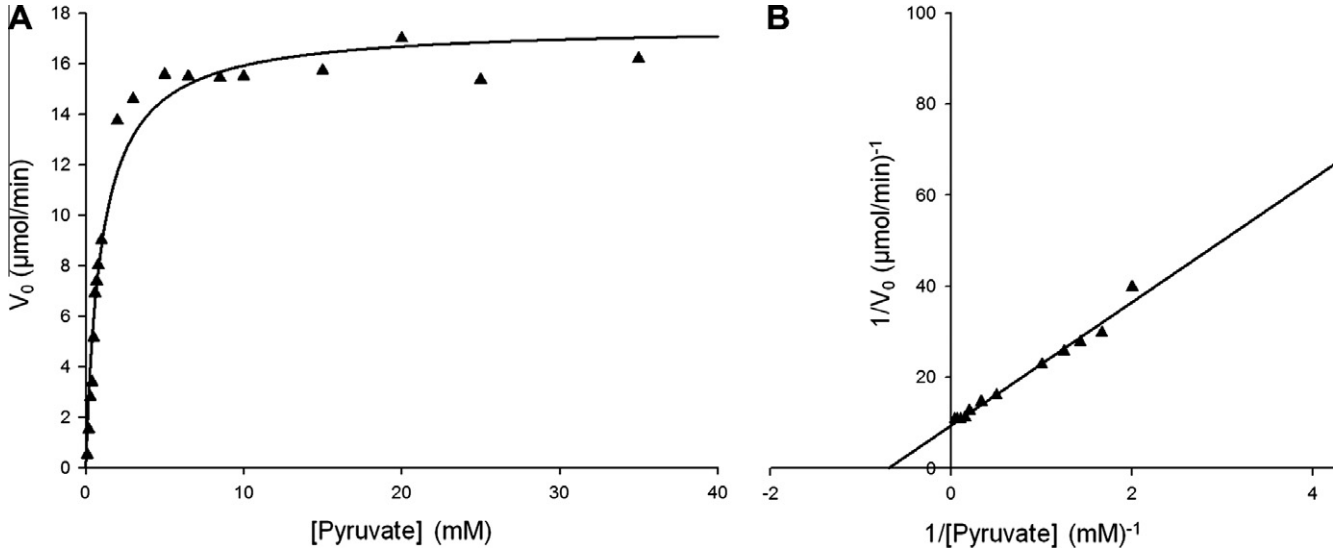

Fig. 5. Kinetics for PDC of *P. serpens*. The data represent mean results from triplicate determinations of PDC activity by the ADH coupled assay using 100 µg of the 16,000*g* supernatant of *P. serpens* sonicated cells. Panel A: Plot of initial reaction rate vs. substrate concentration. Panel B: Lineweaver-Burk plot of kinetic data.

14 analysed *Phytomonas* isolates of different origins. No amplification products in trypanosomatids of the genera *Blastocrithidia*, *Crithidia, Endotrypanum, Herpetomonas, Leishmania, Leptomonas* and *Trypanosoma* were obtained.

The possible origin of the *PDC/IPDC* gene in *Phytomonas* was inferred from the phylogenetic analysis of protein sequences of IPDCs, IPDC/PPDCs and PDCs of bacteria, fungi and plants. In the process of sequence selection we verified that no IPDC sequences were annotated for fungi or plants, nor PDCs for Archaea. The lack of Archaea PDCs likely reflects the limited amount of data from these microorganisms in sequence databases.

The phylogenetic analyses revealed distinct clusters of plant, fungal and bacterial enzymes. *Phytomonas* PDC/IPDC was robustly monophyletic with the *Leishmania* and *C. fasciculata* enzymes. In the trees obtained by NJ, MP and ML this clade appeared as a sister group of IPDCs of the γ-proteobacteria *P. agglomerans*, *E. cloacae* and *P. putida*, with 98% bootstrap support in the MP tree. The proximity of the trypanosomatid enzymes with IPDCs of phytobacteria suggests a HGT event from a donor phytobacteria to a recipient ancestral trypanosome. We propose that this event occurred prior to the separation between *Phytomonas*, *Leishmania* and *Crithidia.*

Interestingly, two recent studies based on extensive phylogenetic analysis concluded that the genomes of trypanosomatids other than members of the genus *Trypanosoma* seem to have incorporated at least three γ-proteobacterial genes that encode enzymes required for haeme biosynthesis (Kořený et al., 2010; Alves et al., 2011). The localisation of these genes in different *Leishmania* chromosomes is suggestive that more than one HGT event from γ-proteobacteria may have occurred before the diversification of non-*Trypanosoma* trypanosomatids.

In *Phytomonas* sp., as in the long-slender stages of African trypanosomes, both living in glucose-rich medium, ATP is produced via substrate level phosphorylation instead of the classical cytochrome-based respiratory chain (reviewed by Bringaud et al., 2006). The acquisition of a PDC activity by *Phytomonas* offers a parallel route to reoxidise part of the NADH instead of exclusive reliance on the plant-like alternative oxidase as previously proposed (Tielens and Van Hellemond, 1998). This leads to the possibility that higher glycolytic rates could be sustained in *Phytomonas* by alcoholic fermentation, where PDC converts the excess of pyruvate to acetaldehyde, which is subsequently reduced by an ADH to ethanol, with the concomitant regeneration of oxidative equivalents (NAD^+). Here we report for the first time a PDC activity in *P. serpens* cell extracts. The enzyme displays a classical Michaelis–Menten kinetics with a Km value for pyruvate of 1.4 mM. The substrate affinity of *P. serpens* enzyme is very similar to the affinity for pyruvic acid of *S. cerevisiae* PDC (Km = 1.3 mM) and *E. cloacae* IPDC (Km = 2.5 mM) (Koga et al., 1992). *E. cloacae* enzyme also catalises decarboxylation of indolepyruvate with a Km of 15 µM (Koga et al., 1992). We have tried to obtain the recombinant *P. serpens* enzyme

in order to perform more broad kinetic studies. The gene was successfully cloned and expressed in three different vector/bacterial systems. However in all cases the protein was produced as inclusion bodies most probably due to the presence of 16 cysteine residues. In the future other cloning and expression systems will be explored.

One attractive aspect to be considered is that *Phytomonas* putative ADH could be represented by an isopropyl alcohol dehydrogenase (iPDH) (Uttaro and Opperdoes, 1997) supposedly acquired by a HGT event from a strictly aerobic bacterium (Molinas et al., 2003). Interestingly iPDH activity was detected in 20 isolates of *Phytomonas* and was absent in the genera *Trypanosoma*, *Leishmania*, *Leptomonas*, *Crithidia*, *Blastocrithidia*, and *Herpetomonas* (Uttaro et al., 1997).

At this point the functional relevance of the PDC/IPDC in *Leishmania* arises. The promastigote insect stage of these parasites relies on amino acid catabolism to obtain energy, whereas amastigotes proliferate in the lysosomal compartment of macrophages and prefer fatty acids as energy sources (Bringaud et al., 2006). Since, to our knowledge, no reports describe ethanol production in *Leishmania*, the PDC function in this genus is a matter that deserves further investigation. The 6-fold increase in the abundance of the *PDC*/*IPDC* gene copy number and corresponding transcripts in *P. serpens* as compared with *Leishmania* further supports that the acquisition of the *PDC*/*IPDC* gene represents an adaptive advantage for *Phytomonas* sp.

In addition to phylogenetic analyses, clues about HGT can be obtained following the examination of genome (A + T) content, GC content at first and third codon positions, dinucleotide frequencies, and synonymous codon usage (Lawrence and Ochman, 1997). The comparison of GC3s index and codon usage of *P. serpens* and *L. major PDC*/*IPDCs* genes with homologous genes of phytobacteria and fungi does not support a recent HGT event. On the other hand, this seems to be the case of the acquisition of *Phytomonas* sp. *iPDH* gene (Molinas et al., 2003). We verified that the GC3s index of this gene is 0.602, which is lower than the GC3s index of *P. serpens PDC*/*IPDC* and *GAPDH* genes (0.849 and 0.758, respectively) and closer to the indexes of bacteria *IPDC*s (0.617 for *E. herbicola*). Criteria based on codon usage bias and differential base compositions have undergone several criticisms (Kuo and Ochman, 2009). Over time introgressed genes are subject to the same mutational pressures affecting all genes in the recipient genome (Lawrence and Ochman, 1998). In this way, these criteria may allow the identification of recent gene transfer events but are inefficient in detecting events that occurred a long time ago (Boto, 2010).

Among trypanosomatids, the genus *Phytomonas* is the only one found to be specifically adapted to infect plants. These hosts provide a particular habitat with a steady and plentiful supply of carbohydrates. Several molecular adaptations reflect *Phytomonas* mobilisation to this environment, such as high expression levels of at least two glucose transporter isoforms (Pappas et al., 2005) necessary for replenishing the intracellular carbohydrate pool for glycolysis. Notwithstanding, several HGT events seem to have mediated a systematic introduction of functional novelties along *Phytomonas* evolutionary record. The potential HGT process for the acquisition of *PDC*/*IPDC* and *iPDH* in *Phytomonas* increases the list of foreign genes that may have entered a trypanosomatid ancestor through lateral gene transfer, conferring adaptive advantages (Opperdoes and Michels, 2007). Alcoholic fermentation ensued from these events might have introduced an extensive metabolic adjustment, sustaining high glycolytic rates in anaerobic conditions. This might have significantly contributed to the adaptation to a unique environment for kinetoplastids and transition to obligate parasitism.

Acknowledgements

We are greatly indebted to Dr. Marta M. G. Teixeira for providing *P. serpens* culture and DNA preparations of several trypanosomatids and to Dr. Lucile Floeter-Winter for providing *L. major* culture. We are grateful to Dr. Nancy Sturm for insightful suggestions. This work was supported by Grants of Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), Brazil (BZ), Grant BIO2002-02228 of Ministerio de Ciencia y Tecnologia, Spain (AG). SI received a fellowship from FAPESP.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.meegid.2012.01.016.

References

Abascal, F., Zardoya, R., Posada, D., 2005. ProtTest: selection of best-fit models of protein evolution. Bioinformatics 21, 2104–2105.

Alves, J.M., Voegtly, L., Matveyev, A.V., Lara, A.M., da Silva, F.M., Serrano, M.G., Buck, G.A., Teixeira, M.M., Camargo, E.P., 2011. Identification and phylogenetic analysis of heme synthesis genes in trypanosomatids and their bacterial endosymbionts. PLoS One 6, e23518.

Bateman, A., Coin, L., Durbin, R., Finn, R.D., Hollich, V., Griffiths-Jones, S., Khanna, A., Marshall, M., Moxon, S., Sonnhammer, E.L., Studholme, D.J., Yeats, C., Eddy, S.R., 2004. The Pfam protein families database. Nucleic Acids Res. 32, D138–141.

Baudoin, E., Lerner, A., Mirza, M.S., El Zemrany, H., Prigent-Combaret, C., Jurkevich, E., Spaepen, S., Vanderleyden, J., Nazaret, S., Okon, Y., Moënne-Loccoz, Y., 2010. Effects of *Azospirillum brasilense* with genetically modified auxin biosynthesis gene *ipdC* upon the diversity of the indigenous microbiota of the wheat rhizosphere. Res. Microbiol. 161, 219–226.

Boto, L., 2010. Horizontal gene transfer in evolution: facts and challenges. Proc. R Soc. B 277, 819–827.

Bradford, M., 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal. Biochem. 72, 248–254.

Bringaud, F., Rivière, L., Coustou, V., 2006. Energy metabolism of trypanosomatids: adaptation to available carbon sources. Mol. Biochem. Parasitol. 149, 1–9.

Camargo, E.P., 1999. *Phytomonas* and other trypanosomatid parasites of plants and fruit. Adv. Parasitol. 42, 29–112.

Catarino, L.M., Serrano, M.G., Cavazzana, M.Jr., Almeida, M.L., Kaneshina, E.K., Campaner, M., Jankevicius, J.V., Teixeira, M.M., Itow-Jankevicius, S., 2001. Classification of trypanosomatids from fruits and seeds using morphological, biochemical and molecular markers revealed several genera among fruit isolates. FEMS Microbiol. Lett. 201, 65–72.

Cazzulo, J.J., Franke de Cazzulo, B.M., Engel, J.C., Cannata, J.J., 1985. End products and enzyme levels of aerobic glucose fermentation in trypanosomatids. Mol. Biochem. Parasitol. 16, 329–343.

Chaumont, F., Schanck, A.N., Blum, J.J., Opperdoes, F.R., 1994. Aerobic and anaerobic glucose metabolism of Phytomonas sp. isolated from *Euphorbia characias*. Mol. Biochem. Parasitol. 67, 321–331.

Conchon, I., Campaner, M., Sbravat, C., Camargo, E.P., 1989. Trypanosomatids, other than *Phytomonas* spp. isolated and cultured from fruit. J. Eukaryot. Microbiol. 36, 412–414.

Costacurta, A., Keijers, V., Vanderleyden, J., 1994. Molecular cloning and sequence analysis of an *Azospirillum brasilense* indole-3-pyruvate decarboxylase gene. Mol. Gen. Genet. 243, 463–472.

Dollet, M., 1984. Plant diseases caused by flagellate protozoa. Ann. Rev. Phytopathol. 22, 115–132.

Drummond, A., Strimmer, K., 2001. PAL: an object-oriented programming library for molecular evolution and phylogenetics. Bioinformatics 17, 662–663.

Duggleby, R.G., 2006. Domain relationships in thiamine diphosphate-dependent enzymes. Acc. Chem. Res. 39, 550–557.

El-Sayed, N.M., Myler, P.J., Blandin, G., Berriman, M., Crabtree, J., Aggarwal, G., Caler, E., Renauld, H., Worthey, E.A., Hertz-Fowler, C., Ghedin, E., Peacock, C., Bartholomeu, D.C., Haas, B.J., Tran, A.N., Wortman, J.R., Alsmark, U.C., Angiuoli, S., Anupama, A., Badger, J., Bringaud, F., Cadag, E., Carlton, J.M., Cerqueira, G.C., Creasy, T., Delcher, A.L., Djikeng, A., Embley, T.M., Hauser, C., Ivens, A.C., Kummerfeld, S.K., Pereira-Leal, J.B., Nilsson, D., Peterson, J., Salzberg, S.L., Shallom, J., Silva, J.C., Sundaram, J., Westenberger, S., White, O., Melville, S.E., Donelson, J.E., Andersson, B., Stuart, K.D., Hall, N., 2005. Comparative genomics of trypanosomatid parasitic protozoa. Science 309, 404–409.

Felsenstein, J., 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. J. Mol. Evol. 17, 368–376.

Fiorini, J.E., Takata, C.S., Teofilo, V.M., Nascimento, L.C., Faria-e-Silva, P.M., Soares, M.J., Teixeira, M.M., De Souza, W., 2001. Morphological biochemical and molecular characterization of *Herpetomonas samuelpessoai camargoi* n. subsp.,

a trypanosomatid isolated from the flower of the squash *Cucurbita moschata*. J. Eukaryot. Microbiol. 48, 62–69.
Fitch, W.M., 1971. Toward defining the course of evolution: minimum change for a specific tree topology. Syst. Zool. 20, 406–416.
Guindon, S., Gascuel, O., 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst. Biol. 52, 696–704.
Hall, T.A., 1999. BIOEDIT: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp. Ser. 41, 95–98.
Hohmann, S., 1991. Characterization of *PDC6*, a third structural gene for pyruvate decarboxylase in *Saccharomyces cerevisiae*. J. Bacteriol. 173, 7963–7969.
Ivens, A.C., Peacock, C.S., Worthey, E.A., Murphy, L., Aggarwal, G., Berriman, M., Sisk, E., Rajandream, M.A., Adlem, E., Aert, R., Anupama, A., Apostolou, Z., Attipoe, P., Bason, N., Bauser, C., Beck, A., Beverley, S.M., Bianchettin, G., Borzym, K., Bothe, G., Bruschi, C.V., Collins, M., Cadag, E., Ciarloni, L., Clayton, C., Coulson, R.M., Cronin, A., Cruz, A.K., Davies, R.M., De Gaudenzi, J., Dobson, D.E., Duesterhoeft, A., Fazelina, G., Fosker, N., Frasch, A.C., Fraser, A., Fuchs, M., Gabel, C., Goble, A., Goffeau, A., Harris, D., Hertz-Fowler, C., Hilbert, H., Horn, D., Huang, Y., Klages, S., Knights, A., Kube, M., Larke, N., Litvin, L., Lord, A., Louie, T., Marra, M., Masuy, D., Matthews, K., Michaeli, S., Mottram, J.C., Müller-Auer, S., Munden, H., Nelson, S., Norbertczak, H., Oliver, K., O'neil, S., Pentony, M., Pohl, T.M., Price, C., Purnelle, B., Quail, M.A., Rabbinowitsch, E., Reinhardt, R., Rieger, M., Rinta, J., Robben, J., Robertson, L., Ruiz, J.C., Rutter, S., saunders, D., Schäfer, M., Schein, J., Schwartz, D.C., Seeger, K., Seyler, A., Sharp, S., Shin, H., Sivam, D., Squares, R., Squares, S., Tosato, V., Vogt, C., Volckaert, G., Wambutt, R., Warren, T., Wedler, H., Woodward, J., Zhou, S., Zimmermann, W., Smith, D.F., Blackwell, J.M., Stuart, K.D., Barrell, B., Myler, P.J., 2005. The genome of the kinetoplastid parasite *Leishmania major*. Science 309, 436–442.
Jankevicius, J.V., Jankevicius, S.I., Campaner, M., Conchon, I., Maeda, L.A., Teixeira, M.M.G., Freymuller, E., Camargo, E.P., 1989. Life cycle and culturing of *Phytomonas serpens* (Gibbs) a trypanosomatid parasite of tomatoes. J. Protozool. 36, 265–271.
Koga, J., Adachi, T., Hidaka, H., 1992. Purification and characterization of indolepyruvate decarboxylase A novel enzyme for indole-3-acetic acid biosynthesis in *Enterobacter cloacae*. J. Biol. Chem. 267, 15823–15828.
Koga, J., 1995. Structure and function of indolepyruvate decarboxylase, a key enzyme in indole-3-acetic acid biosynthesis. Biochim. Biophys. Acta 1249, 1–13.
Kořený, L., Lukes, J., Oborník, M., 2010. Evolution of the haem synthetic pathway in kinetoplastid flagellates: an essential pathway that is not essential after all? Int. J. Parasitol. 40, 149–156.
Kuo, C., Ochman, H., 2009. Inferring clocks when lacking rocks: the variable rates of molecular evolution in bacteria. Biol. Direct 2009 4, 35.
Lawrence, J.G., Ochman, H., 1997. Amelioration of bacterial genomes: rates of change and exchange. J. Mol. Evol. 44, 383–397.
Lawrence, J.G., Ochman, H., 1998. Molecular archaeology of the *Escherichia coli* genome. Proc. Natl. Acad. Sci. USA 95, 9413–9417.
Lee, C., Lee, S., Shin, S.G., Hwang, S., 2008. Real-time PCR determination of rRNA gene copy number: absolute and relative quantification assays with *Escherichia coli*. Appl. Microbiol. Biotechnol. 78, 371–376.
Marín, C., Alberge, B., Dollet, M., Pagès, M., Bastien, P., 2008. First complete chromosomal organization of a protozoan plant parasite (*Phytomonas* spp.). Genome 91, 88–93.
Marín, C., Dollet, M., Pagès, M., Bastien, P., 2009. Large differences in the genome organization of different plant Trypanosomatid parasites (*Phytomonas* spp) reveal wide evolutionary divergences between taxa. Infect. Genet. Evol. 9, 235–240.
Maslov, D.A., Nawathean, P., Scheel, J., 1999. Partial kinetoplast-mitochondrial gene organization and expression in the respiratory deficient plant trypanosomatid *Phytomonas serpens*. Mol. Biochem. Parasitol. 99, 207–221.
Molinas, S.M., Altabe, S.G., Opperdoes, F.R., Rider, M.H., Michels, P.A., Uttaro, A.D., 2003. The multifunctional isopropyl alcohol dehydrogenase of *Phytomonas* sp could be the result of a horizontal gene transfer from a bacterium to the trypanosomatid lineage. J. Biol. Chem. 278, 36169–36175.
Møller, K., Langkjaer, R.B., Nielsen, J., Piskur, J., Olsson, L., 2004. Pyruvate decarboxylases from the petite-negative yeast *Saccharomyces kluyveri*. Mol. Genet. Genomics 270, 558–568.
Monier, A., Claverie, J.M., Ogata, H., 2007. Horizontal gene transfer and nucleotide compositional anomaly in large DNA viruses. BMC Genomics 8, 456.
Muller, E., Gargani, D., Banuls, A.L., Tibayrenc, M., Dollet, M., 1997. Classification of plant trypanosomatids (*Phytomonas* spp.): parity between random-primer DNA typing and multilocus enzyme electrophoresis. Parasitol. 115, 403–409.
Nawathean, P., Maslov, D.A., 2000. The absence of genes for cytochrome *c* oxidase and reductase subunits in maxicircle kinetoplast DNA of the respiration-deficient plant trypanosomatid *Phytomonas serpens*. Curr. Genet. 38, 95–103.
Opperdoes, F.R., Michels, P.A., 2007. Horizontal gene transfer in trypanosomatids. Trends Parasitol. 23, 470–476.
Page, R.D., 1996. TREEVIEW: an application to display phylogenetic trees on personal computers. Comput. Appl. Biosci. 12, 357–358.
Pappas Jr., G.J., Benabdellah, K., Zingales, B., González, A., 2005. Expressed sequence tags from the plant trypanosomatid *Phytomonas serpens*. Mol. Biochem. Parasitol. 142, 149–157.
Patten, C.L., Glick, B.R., 2002. Role of *Pseudomonas putida* indoleacetic acid in development of the host plant root system. Appl. Environ. Microbiol. 68, 3795–3801.
Saitou, N., Nei, M., 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4, 406–425.
Schütz, A., Sandalova, T., Ricagno, S., Hübner, G., König, S., Schneider, G., 2003. Crystal structure of thiamin diphosphate-dependent indolepyruvate decarboxylase from *Enterobacter cloacae* an enzyme involved in the biosynthesis of the plant hormone indole-3-acetic acid. Eur. J. Biochem. 270, 2312–2321.
Serrano, M.G., Nunes, L.R., Campaner, M., Buck, G.A., Camargo, E.P., Teixeira, M.M., 1999a. Trypanosomatidae: *Phytomonas* detection in plants and phytophagous insects by PCR amplification of a genus-specific sequence of the spliced leader gene. Exp. Parasitol. 91, 268–279.
Serrano, M.G., Campaner, M., Buck, G.A., Teixeira, M.M., Camargo, E.P., 1999b. PCR amplification of the spliced leader gene for the diagnosis of trypanosomatid parasites of plants and insects in methanol-fixed smears. FEMS Microbiol. Lett. 176, 241–246.
Spaepen, S., Versées, W., Gocke, D., Pohl, M., Steyaert, J., Vanderleyden, J., 2007a. Characterization of phenylpyruvate decarboxylase, involved in auxin production of *Azospirillum brasilense*. J. Bacteriol. 189, 7626–7633.
Sturm, N.R., Dollet, M., Lukes, J., Campbell, D.A., 2007. Rational sub-division of plant trypanosomes (*Phytomonas* spp) based on minicircle conserved region analysis. Infect. Genet Evol 7, 570–576.
Swofford, D. L., 2003. PAUP* Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4.0b10. Sinauer Associates, Sunderland, Massachusetts.
Teixeira, M.M., Serrano, M.G., Nunes, L.R., Campaner, M., Buck, G.A., Camargo, E.P., 1996. Trypanosomatidae: a spliced-leader-derived probe specific for the genus *Phytomonas*. Exp. Parasitol. 84, 311–319.
Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions – specific gap penalties and weight matrix choice. Nucleic Acids Res. 22, 4673–4680.
Tielens, A.G., Van Hellemond, J.J., 1998. Differences in energy metabolism between trypanosomatidae. Parasitol. Today 14, 265–272.
Uttaro, A.D., Opperdoes, F.R., 1997. Purification and characterisation of a novel *iso*-propanol dehydrogenase from *Phytomonas* sp. Mol. Biochem. Parasitol. 85, 213–219.
Uttaro, A.D., Sánchez-Moreno, M., Opperdoes, F.R., 1997. Genus-specific biochemical markers for *Phytomonas* spp. Mol. Biochem. Parasitol. 90, 337–342.