Infection, Genetics and Evolution 20 (2013) 171–176

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

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

A protocol to improve genotyping of problematic microsatellite loci of *Trypanosoma brucei gambiense* from body fluids

Jacques Kaboré [a,e,*], Thierry De Meeûs [a,b], Annette MacLeod [c], Hamidou Ilboudo [a], Paul Capewell [c], Mamadou Camara [d], Adrien Marie Gaston Belem [e], Bruno Bucheton [a,b], Vincent Jamonneau [a,b]

[a] Centre International de Recherche-Développement sur l'Élevage en zones Subhumides (CIRDES), Unité de recherches sur les bases biologiques de la lutte intégrée, 01 BP 454 Bobo-Dioulasso 01, Burkina Faso
[b] Institut de Recherche pour le Développement, Unité Mixte de Recherche IRD-CIRAD 177, TA A-17/G, Campus International de Baillarguet, F-34398 Montpellier, France
[c] Wellcome Centre for Molecular Parasitology, University of Glasgow, Biomedical Research Centre, 120 University Place, Glasgow G12 8TA, United Kingdom
[d] Programme National de Lutte contre la Trypanosomose Humaine Africaine, BP 851 Conakry, Guinea
[e] Université Polytechnique de Bobo-Dioulasso, 01 BP 1091 Bobo-Dioulasso 01, Burkina Faso

ARTICLE INFO

Article history:
Received 20 May 2013
Received in revised form 5 August 2013
Accepted 7 August 2013
Available online 15 August 2013

Keywords:
Trypanosoma brucei gambiense
Microsatellite loci
Body fluids
Genotyping
Null alleles
Dropouts

ABSTRACT

Microsatellite genotyping of *Trypanosoma brucei gambiense*, the causative agent of human African trypanosomiasis or sleeping sickness, and population genetics tools, are useful for inferring population parameters such as population size and dispersal. Amplifying parasite DNA directly from body fluids (i.e., blood, lymph or cerebrospinal fluid) allows avoiding costly and tedious isolation phases. It is however associated to increased frequencies of amplification failures (allelic dropouts and/or null alleles) at some loci. In this paper, we present a study focused on three *T. brucei gambiense* microsatellite loci suspected to present amplification problems when amplified from body fluids sampled in Guinean sleeping sickness foci. We checked for the real nature of blank and apparent homozygous genotypes of parasite DNA directly amplified from body fluids and tested the effect of three different DNA quantities of trypanosomes. Our results show that some initially blank and homozygous genotypes happen to be actual heterozygous genotypes. In Guinea, lymph from the cervical nymph nodes, known to contain the highest concentrations of parasites, appeared to provide the best amplification results. Simply repeating the PCR may be enough to retrieve the correct genotype, but we also show that increasing initial DNA content provides better results while undertaking first amplification. We finally propose an optimal protocol for amplifying trypanosome's DNA directly from body fluids that should be adapted to local characteristics and/or constraints.

1. Introduction

Human African trypanosomiasis (HAT), or sleeping sickness, is a parasitic disease whose pathogenic agent, *Trypanosoma brucei* (*T. brucei*) can be found in different biological fluids (blood, lymph and cerebrospinal fluid). Among *T. brucei*, *Trypanosoma brucei gambiense* group 1 is a monophyletic genetic group (Gibson, 2007; Koffi et al., 2007, 2009) responsible of HAT in West and Central Africa. Genotyping this pathogen remains an important goal, for a number of applications, for example, to study the role of *T. brucei gambiense* group 1 genetic polymorphism in the diversity of responses to HAT infections observed in the field (Garcia et al., 2006; Bucheton et al., 2011) and infer basic population parameters such as population size and dispersal (Koffi et al., 2009; Simo et al., 2010). Microsatellite markers were shown to be polymorphic enough to study the genetic diversity among *T. brucei gambiense* group 1 (Biteau et al., 2000; Koffi et al., 2007, 2009; Balmer et al., 2011). An important advantage of microsatellite markers is that they allow working from small amounts of targeted DNA and should thus allow studying *T. brucei* directly from biological fluids (Koffi et al., 2007; Morrison et al., 2007) thus avoiding time costly amplification steps, either *in vivo* or *in vitro*.

However, it has been observed in recent studies that genotyping *T. brucei* directly from body fluids is confronted with unusually frequent missing data (Morrison et al., 2007; Koffi et al., 2009; Kaboré et al., 2011) and also to frequent homozygous profiles (Kaboré et al., 2011) at some loci that are unexpected given the theoretically mostly clonal nature of *T. brucei gambiense* reproductive strategy (Koffi et al., 2009; Simo et al., 2010). In the most recent of these studies, Kaboré et al. (2011) genotyped 213 parasite samples using six microsatellite loci from biological fluids of Guinean patients

* Corresponding author at: Centre International de Recherche-Développement sur l'Élevage en zones Subhumides (CIRDES), Unité de recherches sur les bases biologiques de la lutte intégrée, 01 BP 454 Bobo-Dioulasso 01, Burkina Faso. Tel.: +226 20 97 62 15; fax: +226 20 97 23 20.
E-mail address: jacqueskabore@yahoo.fr (J. Kaboré).

1567-1348/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.meegid.2013.08.006

(Kaboré et al., 2011). Among these samples, they noted great frequencies of missing data (20.03%) and unusual frequency of homozygous genotypes (5.44%) as compared to a previous study on the same site (Koffi et al., 2009). This unusually frequent homozygosity was assumed to be due to null alleles and/or allelic dropouts (Kaboré et al., 2011). Null alleles (or dropouts) produce an underestimation of heterozygosity, together with an increase across loci variance, which could be interpreted as resulting from rare sexual recombination events (Balloux et al., 2003; De Meeûs et al., 2006). Avoiding such problems thus represents a crucial goal. In the present paper we wanted to improve genotyping at three loci that displayed unusually high homozygosity and missing data after amplified from body fluids. We present here the results obtained with four kinds of fluids (whole blood, buffy coat, cervical node lymph and cerebrospinal fluid) with varying DNA quantities. We demonstrate that most initially homozygous profiles found for these loci were in fact heterozygous and that increasing DNA concentration or simply repeating DNA amplifications can help considerably revealing the actual microsatellite genotype of *T. brucei gambiense* from body fluids at such problematic loci. We also show that, for Guinean foci, cervical node lymph is the body fluid that provides the best results. We finally propose the best strategy to be used for field collected body fluids in order to be able studying the reproductive mode and population structure of parasites such as trypanosomes at as many loci as possible.

2. Materials and methods

2.1. Microsatellite markers and samples

The three microsatellite markers (Misatg4, Micbg6 and Micbg5) that previously generated most homozygous profiles and missing data in Kaboré et al. (2011) were selected for this study. Among these three loci, Micbg5 was responsible for most missing data. Thus, we selected more samples for Micbg5 than for the two other loci.

DNA samples were selected on the basis of the results given in the Supplementary Table S1 (Kaboré et al., 2011). DNA extraction had already been performed during a previous work (Kaboré et al., 2011) using the DNeasy® Tissue kit (Qiagen). Genotyping of these DNA samples has shown different profiles such as two alleles (heterozygous), one allele (real homozygous or resulting from an allele dropout) and blank genotypes (missing data) (see Supplementary Table S1). DNA samples were extracted from whole blood, buffy-coat, cervical node lymph and cerebrospinal fluid. These samples were taken from HAT active foci (Dubreka, Boffa and Forecariah) in the coastal mangroves area in Guinea during medical surveys conducted by the National Control Programs (NCP) according to the national HAT diagnostic procedures. All participants were informed about the objective of the study in their own language and signed an informed consent. Children less than 12 years old were excluded from the study. For participants between 12 and 18 years old, informed consent was obtained from their parents. This study is part of a larger project aiming at improving HAT diagnosis for which ethical clearance was obtained from WHO and IRD ethical committees.

2.2. Amplification conditions

For each body fluid type and each locus, we designed three different protocols. For the first protocol (P1), 1 μl of extracted DNA was used as template for PCR, in a volume of 10 μl. For other protocols (P2; P3), 2.5 and 5 μl of extracted DNA were used respectively as template for PCR. Such protocol represented the optimal strategy for preserving as much as possible body fluid samples, which are always available at limited quantities for obvious ethical reasons, and optimising genotyping.

Fully nested PCR was used with two distinct primer pairs (Koffi et al., 2007; Kaboré et al., 2011) (see Koffi et al., 2007 for more details). Nested-PCR amplifications were carried out using the following conditions, PCR buffer (45 mM Tris–HCl pH 8.8, 11 mM $(NH_4)_2SO_4$, 4.5 mM $MgCl_2$, 6.7 mM 2-mercaptoethanol, 4.4 μM EDTA, 113 μg per ml BSA, 1 mM of each of four deoxyribonucleotide triphosphates), 1 mM of each primer and 0.1 unit of Taq polymerase (ABgene) per 10 μL reaction. The reactions were covered with mineral oil to prevent evaporation and amplification performed in a thermocycler Robocycler gradient 96 (Stratagene, La Jolla, CA, UK). All PCR reactions were amplified under the following conditions: 28 cycles of 95 °C for 50 s, 55 °C for 50 s and 65 °C for 1 min. One micro liter of a 1/280 dilution of the first round product was used as template in the second round PCR. Alterations were not considered in how dilute the template for the second reaction was because the dilution used is standard. One primer out of every second round pair for the nested PCR amplification of microsatellite marker included a 5′-FAM or HEX modification, allowing size separation of products on a capillary-based sequencer (ABI 3100 Genetic Analyser; Applied Biosystems). A set of ROX-labelled size standards (GS 500 HD markers; Applied Biosystems) was included in each sample, allowing determination of PCR amplicon size to the level of 1 bp using Peak Scanner v1.0 Software (Applied Biosystems). The products of the second round of PCR were also analysed by separation by electrophoresis on 3% NuSieve GTG agarose gel in 0.5× TBE buffer containing 50 ng per ml ethidium bromide and visualised with UV. The results of electrophoresis are photographed for analysis.

2.3. Genescan

Depending of the intensity of the electrophoresis profile, an appropriate dilution (1/10–1/500) was performed for each product of the second step of the nested PCR. Then, 20 μl of each dilution was distributed in the different wells of the genescan plate. The wells were then covered with mineral oil to prevent evaporation. Plates were then ready for genescan. Genotyping was performed using an automatic sequencer with 8 capillaries (ABI 3730 Capillary DNA Sequencer, Genetic Analyser, Applied Biosystems; Dundee Sequencing Service http://www.dnaseq.co.uk), allele size determination was obtained as in Kaboré et al. (2011) using Peak Scanner v1.0 Software (Applied Biosystems) and definitive genotype recorded after being read by two independent researchers (JK and BB).

2.4. Data analyses

In order to test which factor explained best the success or failure of our amplifications, we undertook a generalised modelling approach. The success of amplification was analysed through generalised linear models (GLiM) with the software R 2.12.0 (R-Development-core-team, 2010). The models were always of the form Response–Protocol + Fluid + Locus + Protocol: Fluid + Protocol: Locus + Fluid: Locus, where ":" stands for the interaction between corresponding variables. Response was either the number of amplified alleles observed instantaneously (with a Poisson's distribution and a log normal error), the presence of amplification failure at both alleles (blanks) or the presence of success of amplification of two different alleles (revealed heterozygous individuals) in which case logistic regressions were used (with a logit link). In any case, we selected the best model following a stepwise procedure and parameter significance was tested with Chi-square tests. We also compared the amplification success resulting from DNA quantity increase (P2 or P3) or simply from an additional