Peer Community Journal

Section: Infections

RESEARCH ARTICLE

Published
2022-11-16

Cite as
Adeline Ségard, Audrey Romero, Sophie Ravel, Philippe Truc, Gauthier Dobigny, Philippe Gauthier, Jonas Etougbetche, Henri-Joel Dossou, Sylvestre Badou, Gualbert Houéménou, Serge Morand, Kittipong Chaisiri, Camille Noûs and Thierry de Meeûs (2022) *Development of nine microsatellite loci for* Trypanosoma lewisi, *a potential human pathogen in Western Africa and South-East Asia, and preliminary population genetics analyses*, Peer Community Journal, 2: e69.

Correspondence
adeline.segard@ird.fr

Peer-review
Peer reviewed and recommended by PCI Infections, https://doi.org/10.24072/pci.infections.100004

Development of nine microsatellite loci for *Trypanosoma lewisi*, a potential human pathogen in Western Africa and South-East Asia, and preliminary population genetics analyses

Adeline Ségard[1], Audrey Romero[1], Sophie Ravel[1], Philippe Truc[1], Gauthier Dobigny[2], Philippe Gauthier[2], Jonas Etougbetche[3], Henri-Joel Dossou[3], Sylvestre Badou[3], Gualbert Houéménou[3], Serge Morand[4,5], Kittipong Chaisiri[6], Camille Noûs[7], and Thierry de Meeûs[1]

Volume **2** (2022), article e69

https://doi.org/10.24072/pcjournal.188

Abstract

Trypanosoma lewisi belongs to the so-called atypical trypanosomes that occasionally affect humans. It shares the same hosts and flea vector of other medically relevant pathogenic agents as Yersinia pestis, the agent of plague. Increasing knowledge on the population structure (reproductive mode, population size, dispersal) of this parasite thus represents a challenging but important issue. The use of polymorphic genetic markers, together with suitable population genetics tools, is a convenient way to achieve such objectives. To date, the population biology of *T. lewisi* is poorly known and, to our knowledge, no population genetics studies have ever been conducted. Here, we present the development of nine microsatellite markers of this species. We investigated their polymorphism in different countries from Africa and South-East Asia from DNAs extracted from the spleen of their rodent reservoirs (essentially rat species). Several amplification problems arose, especially with South-East Asian individuals. This led to retain only those individuals with complete genotypes (most of them originating from West Africa, notably Cotonou, Benin) to ensure an optimal estimate of heterozygosity. Our results pointed towards a mainly (at least 95-99%) clonal mode of propagation, a strong subdivision at the smallest scale available (i.e., urban neighborhoods, i.e. 0.250 km²), and a generation time most probably shorter than 4 months. In future studies, more extensive sampling at smaller geographic scales (i.e., households), within a one- or two-months window and with improved amplification conditions, should lead to a more precise picture of the fine population structure of this parasite.

[1]Univ Montpellier, Cirad, IRD, Intertryp, Montpellier, France, [2]CBGP, IRD, Montpellier, France, [3]URIB, LARBA, EPAC, Univ. Abomey-Calavi, Cotonou, Benin, [4]MIVEGEC, CNRS, IRD, Montpellier Université, Montpellier, France, [5]Faculty of Veterinary, Kasetsart University, Bangkok, Thailand, [6]Faculty of Tropical Medicine, Mahidol University, Bangkok, Thailand, [7]Cogitamus laboratory, France

Introduction

The two classic forms of human trypanosomiasis are sleeping sickness or human African trypanosomiasis (HAT, also known as gHAT, due to Trypanosoma brucei gambiense, and rHAT, due to T. b. rhodesiense) and Chagas disease, or American trypanosomiasis, due to Trypanosoma cruzi. Other trypanosomoses normally exclusively infect animals in Africa where they are called "African animal trypanosomoses" (or AAT; Nagana, surra and dourine) as well as in Asia and America (surra and dourine). Consequently, the latter parasites are considered as atypical when they infect humans (Truc, Büscher et al., 2013). Atypical human trypanosomoses (a-HT) that have been documented to induce pathologies are: Trypanosoma evansi, *Trypanosoma lewisi*, Trypanosoma congolense, Trypanosoma brucei brucei and Trypanosoma vivax. These trypanosomes usually infect cattle, equines, camelids, suids and rodents. In humans, these parasites are considered to be non- or poorly pathogenic. They have infected individuals with either spontaneous cures, pathologic profiles requiring treatment, and sometimes with fatal issues (Doke & Kar, 2011). Following advances in molecular biology technologies, more recent human cases have been described in India, Gambia, Egypt, Thailand and Vietnam (Wabale, Nalage et al., 2015; Chau, Chau et al., 2016). In Africa, confusion of these atypical forms with sleeping sickness is suspected. Indeed, a patient with a mixed infection with *T. brucei* and *T. congolense* was identified in Côte d'Ivoire, and successfully treated (Truc, Jamonneau et al., 1998). More recently, during an investigation, *T. congolense*'s DNA was detected in the blood of 11 out of 480 subjects tested in the Maro sleeping sickness focus in Chad (Ibrahim, Weber et al., 2021).

Trypanosoma lewisi is a worldwide blood parasite of rodents transmitted by fleas (Hoare, 1972). This parasite is resistant to normal human serum (Lun, Wen et al., 2015). Several surveys found that small mammals (rodents and shrews) were infected with *T. lewisi* in West African villages and cities (e.g. (Tatard, Garba et al., 2017; Rossi, Kadaouré et al., 2018; Dobigny, Gauthier et al., 2019)) as well as in Southeast Asian villages (Pumhom, Morand et al., 2015). The high prevalence found in small mammals, especially rats, within the domestic and peri-domestic environment, suggested that many people, especially infants, may be at high risk of trypanosomes spill-over from rodents.

Heterozygosity levels and population structure of *T. lewisi* remain unknown, and so is its reproductive mode. Yet, such knowledge is crucial to understand the epidemiology of this potential zoonotic disease and get clues on its transmission dynamics. From there, we decided to develop *T. lewisi*-specific microsatellite markers and to investigate further the genetic diversity and population structure of this parasite. Such markers would indeed open the gate to investigations of population genetic co-structure of rodent reservoir and flea vector, together with other zoonotic pathosystem that are responsible for major zoonotic diseases such as plague and murine typhus.

Methods

Ethical statement

In Benin, researches were conducted within the framework of the research agreement between the Republic of Benin and the French National Institute for Sustainable Development (IRD) that was reapproved on the 6th April 2017, as well as the partnership agreement between IRD and the University of Abomey-Calavi (signed on the 30th September 2010 and renewed on the 3rd July 2019).

In Senegal, researches were carried out under the framework agreement established between IRD, the Republic of Senegal and the Senegalese Head Office of Waters and Forests (available upon request). At the time of sampling, no ethic agreement was required to investigate pest rodents in these two countries.

In Lao Republic and Thailand, ethic agreements were obtained from the National Ethics Committee of Health Research (Ministry of Health Council of Medical Sciences, 51/NECHR) and the Ethical Committee of Mahidol University, Bangkok (0517.1116/661), respectively. Samples from Cambodia were used under the courtesy of the Pasteur Institute of Cambodia (CeroPath project, coord. P. Buchy).

In all countries, explicit oral agreements were systematically obtained from local traditional (e.g. family and household heads, shop, firm and garden owners) as well as administrative (City Hall services, urban district chiefs) authorities before rodent trapping.

None of the rodent species captured for this study has protected status according to IUCN/CITES. Rodents were captured and brought alive to the lab where they were treated in a respectful manner in accordance with the guidelines of the American Society of Mammalogists (Sikes & Gannon, 2011), sedated and then sacrificed by cervical dislocation as recommended by Mills et al. (Mills, Yates et al., 1995). Handling procedures were performed under our laboratory agreement for experiments on wild animals (no. 34-169-1).

Access to and benefit-sharing of genetic resources in Benin produced during the course of the present study was authorized by the Benin national authorities following the Nagoya international protocol (permit 608/DGEFC/DCPRN/PF-APA/SA). The other samples were collected before Nagoya protocol-associated procedures implementation. Moreover, there is no possibility of commercial use of any of the genetic diversity evidenced during this work, and the co-authorship with our partners from the countries involved testifies of the access and benefit sharing on the utilization of the genetic diversity studied in this paper.

Biological material transfers to France have been systematically approved by the Regional Head of Veterinary Service Hérault, France.

Samples and associated data were deposited in the Small Mammal Collection at the IRD/CBGP (https://doi.org/10.15454/WWNUPO) as well as at URIB/LARBA/EPAC and Kasetsart University (Thailand). They are available upon request.

Sampling

Isolates (or ramets) of *T. lewisi* came from two continental landmasses with heterogeneous subsamples sizes and cohort compositions (Table 1) (225 isolates). Most of DNAs were extracted directly from qPCR-positive rodent spleens: in Thailand, Cambodia and Lao RP (Pumhom et al., 2015); in Niger and Nigeria (Tatard et al., 2017); in Senegal: (Cassan, Diagne et al., 2018); and in Benin (Dobigny et al., 2019). However, 12 ramets from Thailand and DRC were extracted from isolated strains cultivated in vivo in rats. Incubation in the mammal host lasts five to six days, followed by a multiplication phase of 7 to 10 days (Hoare, 1972; Zhang, Li et al., 2019), after which non-multiplying adult forms (trypomastigotes) appear and stay in the blood for weeks if not months ((Hoare, 1972), page 221). Inside the flea, the entire cycle lasts five days, but the parasite then remains infective up to a year ((Hoare, 1972), page 229). Assuming large variances around these median values, we assumed that the entire cycle is completed within two months. Considering a two months generation time, sampling corresponded to 39 different cohorts (Table 1). For African isolates (Benin and Niger), rats could be sampled in different neighborhoods (quarters) of two cities (Cotonou and Niamey) (Table 1).

Table 1: Date of capture (Month, Year) of infected rodents, corresponding cohort with a 2-months generation time, location (Country, City or Province, Quarter) and number of *Trypanosoma lewisi* isolates (N). *: Province name; NA: not available

Month	Year	Cohort	Country	Town/Province (*)	Quarter	N
12	2007	1	Niger	Boumba	NA	1
7	2008	5	Lao-PDR	Luang-Prabang	NA	1
7	2008	5	Lao-PDR	Luang-Prabang	NA	1
7	2008	5	Lao-PDR	Luang-Prabang	NA	1
7	2008	5	Lao-PDR	Luang-Prabang	NA	1
7	2008	5	Lao-PDR	Luang-Prabang	NA	1
7	2008	5	Lao-PDR	Luang-Prabang	NA	1
12	2008	7	Cambodia	Sihanouk*	NA	1
11	2008	7	Cambodia	Sihanou*k	NA	1
11	2008	7	Cambodia	Sihanouk*	NA	1
11	2008	7	Cambodia	Sihanouk*	NA	1
11	2008	7	Cambodia	Sihanouk*	NA	1
11	2008	7	Cambodia	Sihanouk*	NA	1
11	2008	7	Cambodia	Sihanouk*	NA	1
11	2008	7	Cambodia	Sihanouk*	NA	1
11	2008	7	Cambodia	Sihanouk*	NA	1
11	2008	7	Cambodia	Sihanouk*	NA	1
3	2009	9	Cambodia	Mondolkiri*	NA	1
4	2009	9	Lao-PDR	Champasak*	NA	1

Month	Year	Cohort	Country	Town/Province (*)	Quarter	N
4	2009	9	Lao-PDR	Champasak*	NA	1
4	2009	9	Lao-PDR	Champasak*	NA	1
4	2009	9	Lao-PDR	Champasak*	NA	1
4	2009	9	Lao-PDR	Champasak*	NA	1
10	2009	11	Niger	Niamey	Corniche-Yantala	1
7	2009	11	Cambodia	Sihanouk*	NA	1
7	2009	11	Cambodia	Sihanouk*	NA	1
7	2009	11	Cambodia	Sihanouk*	NA	1
7	2009	11	Cambodia	Sihanouk*	NA	1
1	2010	12	Niger	Niamey	Gamkalley	19
4	2010	13	Niger	Niamey	Corniche-Gamkalley	4
4	2010	13	Niger	Niamey	Gamkalley	7
11	2009	13	Lao-PDR	Champasak*	NA	1
11	2009	13	Lao-PDR	Champasak*	NA	1
11	2009	13	Lao-PDR	Champasak*	NA	1
11	2009	13	Lao-PDR	Champasak*	NA	1
2	2010	14	Lao-PDR	Luang-Prabang	NA	1
2	2011	19	Thailand	Bangkok	Bangkaen	3
7	2011	21	Niger	Gaya	NA	1
7	2011	21	Niger	Gaya	NA	1
7	2011	21	Niger	Gaya	NA	1
3	2013	31	Nigeria	Kano	NA	1
3	2013	31	Nigeria	Sokoto	NA	1
3	2013	31	Nigeria	Zaria	NA	1
11	2013	35	Senegal	Badi-Nieriko	NA	1
11	2013	35	Senegal	Badi-Nieriko	NA	1
11	2013	35	Senegal	Kedougou	NA	1
11	2013	35	Senegal	Kedougou	NA	1
1	2014	36	Senegal	Diakene-Wolof-	NA	1
2	2014	36	Senegal	Diattacounda	NA	1
1	2014	36	Senegal	Marsassoum	NA	1
2	2014	36	Senegal	Mereto	NA	1
1	2014	36	Senegal	Tobor	NA	1
1	2014	36	Senegal	Tobor	NA	1
3	2017	37	Benin	Cotonou	Agla	26
3	2017	37	Benin	Cotonou	Ladji	23
3	2017	37	Benin	Cotonou	Saint-Jean	23
6	2017	39	Benin	Cotonou	Agla	12
6	2017	39	Benin	Cotonou	Ladji	20
6	2017	39	Benin	Cotonou	Saint-Jean	21
NA	2008	2-7	Thailand	Bangkok	Bangkaen	1
NA	2013	30-35	Thailand	Kanchanaburi*	NA	1
NA	1968	NA	DRC	Kinshasa	Ref-Wery	1
NA	NA	NA	Thailand	Buriram*	NA	1
NA	NA	NA	Thailand	Buriram*	NA	1
NA	NA	NA	Thailand	Buriram*	NA	1
NA	NA	NA	Thailand	Buriram*	NA	1
NA	NA	NA	Thailand	Buriram*	NA	1
NA	NA	NA	Thailand	Buriram*	NA	1
NA	NA	NA	Thailand	NA	NA	1
NA	NA	NA	Thailand	NA	NA	1
NA	NA	NA	Thailand	NA	NA	1
NA	NA	NA	Thailand	NA	NA	1
NA	NA	NA	Thailand	NA	NA	1
NA	NA	NA	Thailand	NA	NA	1
NA	NA	NA	Thailand	NA	NA	1
NA	NA	NA	Thailand	Nan*	NA	1
NA	NA	NA	Thailand	Nan*	NA	1

Table 1 (continued)

Microsatellite loci

DNA extraction and detection of Trypanosoma lewisi*-carrying samples*

Total DNA was extracted from ethanol-preserved spleen tissue and pellets of in vivo cultured trypanosomes using the DNeasy 96 Blood and Tissue Kit (Qiagen) according to manufacturer instructions. Whole DNA was eluted with 200 µL of elution buffer.

Screening for the presence of *Trypanosoma* in rodent samples were carried out with a 131 bp-long fragment of the 18S rRNA gene qPCR-based assay with two primers (TRYPA1: AGGAATGAAGGAGGGTAGTTCG, TRYPA2: CACACTTTGGTTCTTGATTGAGG) and a pair of hybridization probes (TRYPA3: LCRed640 – AGAATTTCACCTCTGACGCCCAGT – Ph, TRYPA4: GCTGTAGTTCGTCTTGGTGCGGTCT – FITC), using a LightCycler® 480 (Roche Diagnostics). Each reaction was duplicated and set in 10 µL final volume using the LC480 Probes Master Kit (Roche Diagnostics, Meylan, France) with 0.5 µM of each primer, 0.25 µM of each probe and 0.25 µL of uracil-DNA-glycosylase (UDG) (Biolabs, Courtaboeuf, France). After an initial incubation step at 50 °C for 1 min and a denaturation step at 95 °C for 10 min, cycling conditions were performed for 50 cycles with a denaturation step at 95 °C for 10 s, annealing step at 56 °C for 10 s and extension step at 72 °C for 15 s. All qPCR-positive samples were sequenced for 400 bp-long fragments of the SSU rDNA gene to determine *Trypanosoma* species, using the primer pair TRYPASEQ1 (ACACTGCAAACGATGACACC) and TRYPASEQ2 (TCAACCAAACAAATCACTCCA). Reaction was carried out in a 50 µL final volume containing 2.5 mM MgCl2, 50 mM KCl, 10 mM Tris-HCl (pH 8.3), 2 mM of each dNTP, 0.2 µM of each primer and 1.25 U of Fast Start Taq DNA polymerase (Roch Diagnostics, Meylan, France). An initial denaturation step was performed for 10 min at 95 °C. Then, the amplification went for 45 cycles with a denaturation at 95 °C for 30 s, annealing at 60 °C for 30 s and extension at 72 °C for 1 min (Dobigny, Poirier et al., 2011; Pumhom et al., 2015; Tatard et al., 2017; Cassan et al., 2018; Dobigny et al., 2019).

Development of the microsatellite loci

The DNA library was prepared from the strain Wery L307 24/9/68 (isolated in Kinshasa in 1968 by Pr Wéry of Institute of Tropical Medicine in Antwerp Belgium, kindly provided by Etienne Pays, Université Libre de Bruxelles Belgium) and using the Nextera DNA sample kit (Illumina, San Diego, CA, USA). Sequencing was performed on a MiSeq Sequencer (Illumina). We used the QDD identification tool (Meglécz, Pech et al., 2014) to detect fragments containing microsatellite markers and to design primers. These primers sequences are available in the supplementary file S1.

A total of 58 primers pairs were screened on 17 *T. lewisi* DNA from Benin and Thaïland. Amplification products were first visualized using 2% agarose gel electrophoresis. A total of 29 primer pairs with amplification products of the expected size were retained. These primer pairs were then evaluated for their polymorphism on all extracts using an ABI3500xL sequencer (Applied Biosystems, Waltham, Massachussets, USA) and using GENEMAPPER 4.1 software. This allowed retaining nine primer pairs (Table 2).

Table 2: Primers and characteristics of the nine microsatellite markers retained

Locus	Forward Primer 5'-3'	Reverse Primer 5'-3'	Microsatellite Motif	Tm °C	Amplification Size bp
LEW2	CCAAGCATCCACAATCGTAA	GTTGGCCGCTATTTGTAAGC	AC	58	120-128
LEW12	AATATGGATAACTCCGCCACA	CCCTACGCAGTACATTATCCAA	AC	58	140-160
LEW16	TTTCAAGACGAGCCTGGATTA	CTTTCAGGCAAATACCTGGTC	AC	58	152-214
LEW32	GACAGGACTTAGCCTCCATGA	GCTAACTCTTTGACCATCTGGC	AC	58	241-291
LEW35	GAAACACTCCTCCAAGCACC	GGTGTCAAGAATTCCATCGC	AGG	58	200-356
LEW42	GTAGATTGGTGCGGGAGG	AATGAGTGACTGTGAGCAACAC	AC	58	296-316
LEW44	CATTGGGACTGTTGACAAGC	CGCCCAACCTGTCTTTGTA	AAT	58	309-339
LEW53	ACATCGGAACTGCCGTTTAC	GCAAACCAACAAATGGCAC	AAT	59	336-360
LEW55	GAAAGCCAGTGTTGTCTAATTTCA	TTGGCTGTATCACCAGAAGAAA	AC	58	335-373

The specificity of these primer pairs was tested using purified DNA from 15 Trypanosomatidae (Table 3).

Table 3: DNA identity and Trypanosomatidae species used to evaluate the specificity of the 9 microsatellite primer pairs

DNA identity	Species
40/1	*Trypanosoma brucei gambiense*
Farakoba	*Trypanosoma brucei brucei*
IL3000	*Trypanosoma congolense savannah*
DIN80	*Trypanosoma congolense forest*
GUB323	*Trypanosoma evansi*
IL3905	*Trypanosoma vivax*
L.t	*Leishmania tropica*
L.d	*Leishmania donovani*
Ops21cl11	*Trypanosoma cruzi I*
IVVcl4	*Trypanosoma cruzi II*
M6241cl6	*Trypanosoma cruzi III*
DogTheis	*Trypanosoma cruzi IV*
SC43cl1	*Trypanosoma cruzi V*
Tulacl2	*Trypanosoma cruzi VI*
TcmM1909	*Trypanosoma cruzi marinkellei*

PCR conditions

Touch-down PCR reactions were as follow: 3’ at 95 °C for the first denaturation, followed by 10 cycles at: 96 °c for 30”, annealing temperature + 5 °C for 30” and 72 °C for 1’, then followed by 30 cycles at: 96 °C for 30”, annealing temperature for 30” and 72 °C for 1’, and finally 5’ at 72 °c for final elongation. PCRs were carried out in a thermocycler (Eppendorf® Mastercycler® nexus) in 10 µL final volume, containing 0.5 U MP Biomedicals Taq DNA polymerase, 1 X reaction buffer, 200 µmol/L dNTPs, 20 pmol of each primer and 1 µL DNA. PCR tests for specificity were carried out in 25 µL final volume containing the same mix.

Data analyses

Quality tests for loci and samples

There was a substantial number of blanks (Bs) (no amplification) and unreadable genotypes (Us) (Supplementary file S1, available at https://zenodo.org/record/7234790). We thus used the same approach as the one developed by Kaboré et al. (Kaboré, MacLeod et al., 2011). If missing data (Bs and Us) translates into poor quality of corresponding isolates (poor conservation conditions of the extract, poor PCR conditions and/or outlier individuals displaying poor match of primers with targeted flanking sequences), this should then also correlate with an increase of dropouts of one allele in heterozygous individuals (fake homozygous profiles). We thus expected a negative correlation between the number of blanks (NBs) and the number of heterozygous loci (NHz), in the same individual, and between the number of unreadable loci (NUs) and NHz. Additionally, NBs and NUs should be positively correlated. We measured and tested these correlations with one-sided Spearman's rank correlation tests under the package R-commander (rcmdr) (Fox, 2005; Fox, 2007) for R version 4.0.5 (R-Core-Team, 2020).

Finding the relevant structure levels

Population genetics analyses were performed on dataset previously converted into the appropriate format by Create v2.37 (Coombs, Letcher et al., 2008).

We first tried to study the effect of cohorts. To do this, we tested subdivision between each possible cohort of the same quarter. This was done with the G-based test (Goudet, Raymond et al., 1996) over all loci with 10,000 randomizations of individuals between subsample pairs. This procedure was identified as the most powerful way to combine tests over loci (De Meeûs, Guégan et al., 2009). This procedure could only be undertaken in Cotonou subsamples, more precisely in Agla, Ladji and Saint-Jean neighborhoods, and between cohorts 37 and 39 (see Table 1 and Supplementary File S1, available at https://zenodo.org/record/7234790).

We also computed Wright's F_{ST} (Wright, 1965), which measures the relative effect of subdivision on inbreeding. It was estimated by Weir and Cockerham's unbiased estimator (Weir & Cockerham, 1984). These computations were undertaken with Fstat2.9.4 (Goudet, 2003), updated from Fstat1.2 (Goudet, 1995). We also estimated the normalized $F_{ST}'=F_{ST}/(1-H_S)$, where H_S is Nei's (Nei & Chesser, 1983) unbiased estimator of local genetic diversity (Hedrick, 2005). This was done after having verified the Wang's criterion (Wang, 2015). According to Wang (2015), F_{ST}' represents the most appropriate measure, if the correlation between Nei's G_{ST} (Nei & Chesser, 1983; Meirmans & Hedrick, 2011) and Nei's H_S is negative. This correlation was tested with a one-sided Spearman's rank correlation test with rcmdr. The values obtained between the three cohort pairs were averaged, and the *p*-values combined with the generalized binomial procedure (Teriokhin, De Meeûs et al., 2007), computed with MultiTest v1.2 (De Meeûs et al., 2009). Since there were only three tests, the whole series was kept for the test (k'=3), as previously recommended (De Meeûs, 2014).

We also used subsamples from Benin and Niger to test for the existence of a Wahlund effect when cohorts were ignored. To do so, we compared the number of significant linkage disequilibrium (LD) tests between locus pairs, and Wright's F_{IS} (Wright, 1965) (measure of the relative contribution of nonrandom union of gametes on inbreeding), estimated with the Weir and Cockerham's (1984) method. These computations were also implemented in Fstat. For the LD tests, we used the *G*-based randomization test with 10,000 randomizations, combined over all subsamples for each locus pair, which was shown to be the most powerful combination method (De Meeûs et al., 2009). Since LD tests produce series of non-independent tests, we adjusted the *p*-values obtained with the Benjamini and Yekutieli procedure (Benjamini & Yekutieli, 2001) with R (command p.adjust). Levels of significance (at BY level) were compared using a one sided signed rank test for paired data with rcmdr, the pairing unit being the locus pair.

The nLD/H_T criterion (Manangwa, De Meeûs et al., 2019) was used to study the correlation between the number of times a locus occurred in a significant LD pair (nLD), and its total genetic diversity estimated with Nei's unbiased H_T (Nei & Chesser, 1983). This was tested with a two-sided Spearman rank correlation test with rcmdr. For F_{IS} comparisons between data with or without cohorts, we used a one-sided Wilcoxon signed rank test for paired data with rcmdr, the pairing unit being the locus.

To test for the effect of neighborhoods (quarters), we could only use subsamples from Benin, cohorts 37 and 39 separately, because of temporal issues or missing information. We measured the subdivision index (F_{ST}) in each cohort, and tested its significance with the *G*-based test in Fstat 2.9.4. We then averaged the F_{ST}, and F_{ST}', and combined the two *p*-values with MultiTest, as described above. We also used 95% confidence intervals (95%CI) obtained by 5,000 bootstraps over loci in Fstat 2.9.4.

Tracking amplification problems in complete genotypes within cohorts and quarters

Wright's F_{IS} and F_{ST} were estimated with Weir and Cockerham's unbiased estimators in Fstat. We computed 95%CI of jackknives over subsamples to draw a picture of the variation of these indices across subsamples. We also used 5,000 bootstraps over loci to obtain 95%CI across loci. In order to detect amplification problems, the following criteria were used.

First, the ratio between the standard errors of F_{IS} (StdrdErrFIS) and F_{ST} (StdrdErrFST), obtained by jackknives over loci, RS=StdrdErrFIS/StdrdErrFST was measured. This statistic can be indicative for the presence of null alleles if above 2, as is a positive correlation between F_{IS} and F_{ST} (De Meeûs, 2018). Correlation was tested with a one-sided Spearman's rank correlation test with rcmdr.

Second, short allele dominance (SAD) was tested, for each locus, with the F_{IT} / allele size correlation criterion (Manangwa et al., 2019), with a one-sided (negative correlation) Spearman's rank correlation test. In case of doubt, we also computed the regression between F_{IS} and allele size, weighted by the product $p_T(1-p_T)$, where p_T is the total allele frequency of the allele as provided by Fstat (All_W) (De Meeûs, Humair et al., 2004). In case of negative slope, we halved the *p*-value to obtain a one-sided test result retrieved with rcmdr.

Third, not knowing the reproductive system may make it difficult to undertake a specific stuttering test (De Meeûs & Noûs, 2022). This is why, in case of stuttering suspicion, we chose to directly pool alleles with one repeat difference or less (i.e., imperfect microsatellite loci), taking care that pooling groups always contained at least one allele with frequency above 0.05. Indeed, pooling rare alleles together may result in a fairly frequent artificial allele with an unjustified weight on the results (De Meeûs, Chan et al., 2021).

Population genetics structure analyses

Linkage disequilibrium (LD) was tested between each pair of locus with the *G*-based randomization test with 10,000 shuffling of genotypes between loci, as described above. We computed Weir and Cockerham (Weir & Cockerham, 1984) unbiased estimators of Wright's *F*-statistics: F_{IS} measures the inbreeding of individuals relative to the inbreeding of subsamples; F_{ST} represents the inbreeding of subsamples relative to inbreeding of the total sample; and F_{IT} corresponds to the inbreeding of individuals relative to inbreeding in the total sample. The significant deviation from 0 was tested by randomizing 10,000 times alleles between individuals within each subsample in order to explore deviation from panmixia, individuals between subsamples in order to investigate subdivision, and of alleles between subsamples in order to test for the deviation of F_{IT} from 0. The statistics used were the unbiased estimators of F_{IS} and F_{IT} (panmixia within subsamples and within the total sample, respectively), and the *G*-statistic (subdivision).

Jackknives over populations were used to obtain 95%CI around loci values and 5,000 bootstraps over loci to get the 95%CI globally. Jackknives 95%CI use standard error (SE) of the F's to retrieve the average value + or - $SE \times t_{0.05,n-1}$, where $t_{0.05,n-1}$ is the Student parameter with type 1 error 0.05, and n is the number of items (here, subsamples) used. As this procedure assumes a normal distribution, which *F*-statistics cannot follow, it has only an illustrative purpose. For its part, 95%CI of bootstraps do not assume any distribution and can thus be used for statistical decisions (De Meeûs, McCoy et al., 2007).

We checked the importance of clonal reproduction by studying the behavior of LD, F_{IS}, F_{ST}, F_{IT}, and the number of repeated multi-locus genotypes (MLGs), as described in previous papers (Balloux, Lehmann et al., 2003; Arnaud-Haond, Alberto et al., 2005; De Meeûs, Lehmann et al., 2006; Arnaud-Haond, Duarte et al., 2007; Séré, Kabore et al., 2014). According to these references, in purely clonal populations, all loci with enough polymorphism are expected to be in LD; F_{IS} should be constantly negative with a small variance across loci; F_{ST} cannot outreach 0.5 in the most subdivided populations, in which case $F_{IT}=0$; and finally, many identical multi-locus genotypes (MLGs) should be present. In population where the clonal rate is very big (e.g., c>0.95), but with a small sexual rate, the proportion of loci in significant LD should drop down, a strong variance of F_{IS} should be observed, strongly subdivided populations may display a $F_{ST}>0.5$ and a $F_{IT}>0$, and less numerous MLGs should be observed. When sexual rate becomes more important (c<0.95), all parameters should tend to mimic panmixia, except LD and MLGs that should be more important than what is expected under panmixia, if populations keep substantial levels of clonality. We measured the MLG diversity using $R=(G-1)/(N-1)$, where *G* is the number of different MLGs and *N* is the subsample size (Arnaud-Haond et al., 2005) . In pure clones, *R* should be much smaller than 1, while in panmictic populations of reasonable size, $R \rightarrow 1$.

In pure clones, a direct relationship is also known to link Nei's estimator of local genetic diversity H_S (Nei & Chesser, 1983) and F_{IS}, which may lead to an "expected" value for pure clones, with large number of possible alleles and 'perfect' data (no amplification problems): $H_S \geq 0.5$ and $F_{IS\_exp}=-(1-H_S)/H_S$ (Séré et al., 2014). This can lead to define a superimposition criterion for clonal organisms. Here we defined a slightly different criterion as compared to what was previously published (Séré et al., 2014): $S_{FIS}=1-abs(F_{IS}-F_{IS\_exp})/abs(F_{IS\_exp})$, where "abs" means absolute value. When $S_{FIS} \geq 0.95$, data are considered to fit the null hypothesis (pure clonality, many alleles and no amplification problem). According to Séré et al.'s (2014) simulations, $0.95>S_{FIS}>0.5$ may be explained by amplification problems, while $S_{FIS}<0.5$ probably reflects the occurrence of rare sexual events (e.g., clonal rate 0.99<c<0.999).

We also constructed a Neighbor Joining tree (NJTree) (Saitou & Nei, 1987) in order to check how isolates organize according to their characteristics (i.e., cohort and location). We used the Cavalli-Sforza and Edward's chord distance (Cavalli-Sforza & Edwards, 1967) computed in FreeNA (Chapuis & Estoup, 2007) without null allele correction. This method is expected to provide the most accurate tree topology between Operational Taxonomic Units (UTOs) (here subsamples) (Takezaki & Nei, 1996).

Statistical scripts (simple R commands) are provided in the Material and Methods section.

Results

Specificity of primers

Raw data are available online in file S1 (https://zenodo.org/record/7234790).

Manangwa, O., De Meeûs, T., Grébaut, P., Segard, A., Byamungu, M., Ravel, S. (2019) Detecting Wahlund effects together with amplification problems : cryptic species, null alleles and short allele dominance in Glossina pallidipes populations from Tanzania. *Molecular Ecology Resources*, **19**, 757-772. https://doi.org/10.1111/1755-0998.12989

Meglécz, E., Pech, N., Gilles, A., Dubut, V., Hingamp, P., Trilles, A., Grenier, R., Martin, J.F. (2014) QDD version 3.1: a user-friendly computer program for microsatellite selection and primer design revisited: experimental validation of variables determining genotyping success rate. *Molecular Ecology Resources*, **14**, 1302-1313. https://doi.org/10.1111/1755-0998.12271

Meirmans, P.G., Hedrick, P.W. (2011) Assessing population structure: F_{ST} and related measures. *Molecular Ecology Resources*, **11**, 5-18. https://doi.org/10.1111/j.1755-0998.2010.02927.x

Mills, J.N., Yates, T.L., Childs, J.E., Parmenter, R.R., Ksiazek, T.G., Rollin, P.E., Peters, C.J. (1995) Guidelines for working with rodents potentially infected with hantavirus. *Journal of Mammalogy*, **76**, 716–722. https://doi.org/10.2307/1382742

Nei, M., Chesser, R.K. (1983) Estimation of fixation indices and gene diversities. *Annals of Human Genetics*, **47**, 253-259. https://doi.org/10.1111/j.1469-1809.1983.tb00993.x

Pumhom, P., Morand, S., Tran, A., Jittapalapong, S., Desquesnes, M. (2015) *Trypanosoma* from rodents as potential source of infection in human-shaped landscapes of South-East Asia. *Veterinary Parasitology*, **208**, 174-180. https://doi.org/10.1016/j.vetpar.2014.12.027

R-Core-Team (2020) R: A Language and Environment for Statistical Computing, Version 3.6.3 (2020-02-29) ed. R Foundation for Statistical Computing, Vienna, Austria, http://www.R-project.org.

Rossi, J.P., Kadaouré, I., Godefroid, M., Dobigny, G. (2018) Landscape epidemiology in urban environments: The example of rodent-borne *Trypanosoma* in Niamey, Niger. *Infection Genetics and Evolution*, **63**, 307-315. https://doi.org/10.1016/j.meegid.2017.10.006

Saitou, N., Nei, M. (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, **4**, 406-425. https://doi.org/10.1093/oxfordjournals.molbev.a040454

Ségard A, Roméro A, Ravel S, Truc P, Gauthier D, Gauthier P, Dossou H-J, Sylvestre B, Houéménou G, Morand S, Chaisiri K, Noûs C, De Meeûs T (2022). Development of nine microsatellite loci for *Trypanosoma lewisi*, a potential human pathogen in Western Africa and South-East Asia, and preliminary population genetics analyses, Supplementary File S1. Zenodo. https://doi.org/10.5281/zenodo.7234790

Séré, M., Kabore, J., Jamonneau, V., Belem, A.M.G., Ayala, F.J., De Meeûs, T. (2014) Null allele, allelic dropouts or rare sex detection in clonal organisms : simulations and application to real data sets of pathogenic microbes. *Parasites and Vectors*, **7**, 331. https://doi.org/10.1186/1756-3305-7-331

Sikes, R.S., Gannon, W.L. (2011) Guidelines of the American Society of Mammalogists for the use of wild mammals in research. *Journal of Mammalogy*, **92**, 235-253. https://doi.org/10.1644/10-MAMM-F-355.1

Takezaki, N., Nei, M. (1996) Genetic distances and reconstruction of phylogenetic trees from microsatellite DNA. *Genetics*, **144**, 389-399. https://doi.org/10.1093/genetics/144.1.389

Tatard, C., Garba, M., Gauthier, P., Hima, K., Artige, E., Dossou, D., Gagare, S., Genson, G., Truc, P., Dobigny, G. (2017) Rodent-borne *Trypanosoma* from cities and villages of Niger and Nigeria: A special role for the invasive genus Rattus? *Acta Tropica*, **171**, 151-158. https://doi.org/10.1016/j.actatropica.2017.03.027

Teriokhin, A.T., De Meeûs, T., Guegan, J.F. (2007) On the power of some binomial modifications of the Bonferroni multiple test. *Zhurnal Obshchei Biologii*, **68**, 332-340.

Truc, P., Büscher, P., Cuny, G., Gonzatti, M.I., Jannin, J., Joshi, P., Juyal, P., Lun, Z.R., Mattioli, R., Pays, E., Simarro, P.P., Teixeira, M.M., Touratier, L., Vincendeau, P., Desquesnes, M. (2013) Atypical human infections by animal trypanosomes. *PLoS Neglected Tropical Diseases*, **7**, e2256. https://doi.org/10.1371/journal.pntd.0002256

Truc, P., Jamonneau, V., N'Guessan, P., N'Dri, L., Diallo, P.B., Cuny, G. (1998) *Trypanosoma brucei* ssp. and *T-congolense*: mixed human infection in Cote d'Ivoire. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, **92**, 537-538. https://doi.org/10.1016/S0035-9203(98)90904-X

Wabale, V., Nalage, P., Joshi, A., Bharadwaj, R., Desphande, K., Chowdhary, A. (2015) Human Asian trypanosomiasis due to *Trypanosoma evansi*: a rare case. *The Journal of Advances in Parasitology*, **2**, 65–68. http://dx.doi.org/10.14737/journal.jap/2015/2.3.65.68

Wang, J. (2015) Does G_{ST} underestimate genetic differentiation from marker data? *Molecular Ecology*, **24**, 3546-3558. https://doi.org/10.1111/mec.13204

Weir, B.S., Cockerham, C.C. (1984) Estimating F-statistics for the analysis of population structure. *Evolution*, **38**, 1358-1370. https://doi.org/10.1111/j.1558-5646.1984.tb05657.x

Wright, S. (1965) The interpretation of population structure by F-statistics with special regard to system of mating. *Evolution*, **19**, 395-420. https://doi.org/10.1111/j.1558-5646.1965.tb01731.x

Zhang, X., Li, S.J., Li, Z.Y., He, C.Y., Hide, G., Lai, D.H., Lun, Z.R. (2019) Cell cycle and cleavage events during in vitro cultivation of bloodstream forms of *Trypanosoma lewisi*, a zoonotic pathogen. *Cell Cycle*, **18**, 552-567. https://doi.org/10.1080/15384101.2019.1577651

Appendix

Cycle threshold of a qPCR

During a real time PCR experiment, a positive reaction translates into the accumulation of a fluorescent signal. The cycle threshold (Ct) corresponds to the number of cycles that is required for the fluorescent signal to exceed the background level. The cycle threshold is expected to be inversely proportional to the amount of targeted nucleic acid in the studied sample; in other words, the lower the Ct, the greater the amount of target nucleic acid in the sample.