POPULATION GENETICS OF *TRYPANOSOMA BRUCEI GAMBIENSE*: A CLONE IN THE DARK

GENETIQUE DES POPULATIONS DE *TRYPANOSOMA BRUCEI GAMBIENSE,* UN CLONE DANS LE NOIR

Mathurin Koffi[1,2,3], **Thierry De Meeûs**[1,2,*], Bruno Bucheton[1,2], Philippe Solano[1,2], Mamadou Camara[4], Dramane Kaba[3], Gérard Cuny[2], Francisco J. Ayala[5], and Vincent Jamonneau[1,2]

[1]Centre International de Recherche-Développement sur l'Elevage en zone Subhumide (CIRDES), 01 BP 454 Bobo-Dioulasso 01, Burkina Faso ; [2]Institut de Recherche pour le Développement, Unité Mixte de Recherche IRD-CIRAD 177, Campus International de Baillarguet, 34398 Montpellier Cedex 5, France ; [3]Institut Pierre Richet, Unité de Recherche "Trypanosomoses", 04 BP 293 Abidjan 04, Côte d'Ivoire; [4]Programme National de Lutte contre la Trypanosomose Humaine Africaine, Conakry, Guinée; [5]Department of Ecology and Evolutionary Biology, University of California, Irvine, CA 92697.

Résumé

La Trypanosomose Humaine Africaine (THA) (ou maladie du sommeil) due à *Trypanosoma brucei gambiense* groupe 1) affecte l'Afrique de l'Ouest et du Centre. *T. brucei* s.l. a une grande faculté d'adaptation à différents hôtes et de nombreuses questions sont débattues quant à son régime principal de reproduction, ses capacités de dispersion et l'importance de sa population réelle. Nous avons analysé le polymorphisme de *T. b. gambiense* au niveau de huit loci microsatellites à partir de souches isolées de patients atteints de THA en Côte d'Ivoire et en Guinée en vue de découvrir comment l'information génétique est répartie au sein des isolats, entre isolats, entre foyers et entre échantillons espacés dans le temps.

Les analyses montrent (1) une absence d'échange à l'échelle de l'Afrique de l'Ouest et que même en Guinée, les deux foyers étudiés échangent peu de gènes; (2) que la taille réelle (génétique) des populations de trypanosomes en termes d'hôtes infectés dépasse probablement le nombre estimé à partir des surveillances épidémiologiques ; et (3) que *T. b. gambiense* groupe 1 se reproduit probablement exclusivement par voie asexuée.

Summary

Human African trypanosomiasis (HAT) (or sleeping sickness) due to *Trypanosoma brucei gambiense*) occurs in Western and Central Africa. *T. brucei* s.l. displays a huge diversity of adaptations and host specificities, and questions about its reproductive mode, dispersal abilities, and effective size remain under debate. We have investigated genetic variation at eight microsatellite loci of *T. b. gambiense* strains isolated from HAT patients in Côte d'Ivoire and Guinea, with the aim of knowing how genetic information was partitioned within and between individuals, between foci and between temporally spaced samples. The results indicate that (i) migration of *T. b. gambiense* group 1 strains does not occur at the scale of West Africa, and that even at a finer scale (e.g. within Guinea) migration is restricted; (ii) effective population sizes of trypanosomes as reflected by infected hosts are probably higher than what the epidemiological surveys suggest; and (iii) *T. b. gambiense* group 1 is most likely a strictly clonally reproducing organism.

Key words: *Trypanosoma brucei gambiense*; genetic diversity; clonality; Western Africa; microsatellite markers; genetic differentiation; effective population size

Introduction

The causative agent of Human African trypanosomiasis (HAT) or sleeping sickness, *Trypanosoma brucei* is subdivided into three subspecies (Hoare, 1972): *T. brucei gambiense* (*T. b. gambiense*) is responsible for the chronic form of HAT in Western and Central Africa, *T. b. rhodesiense* is the agent of the acute form of HAT in East Africa, and *T. b. brucei* does not infect humans but causes animal trypanosomiasis (nagana) in cattle. During the last decades, molecular methods have been developed for typing *T. brucei* s.l. stocks in order to study its population structure and taxonomy. Only one group could be

clearly identified as a distinct genetic entity: *T. b. gambiense* group 1, which is considered to be the main causative agent of HAT in Western and Central Africa (Gibson, 1995; Gibson, 2007).

Trypanosoma brucei s.l. displays a huge diversity of adaptations and host specificities and questions about its reproductive mode, dispersal abilities, and effective population size remain under debate. Like most protozoan parasites, *T. brucei* s.l. has been assumed to be clonal (Tibayrenc, 1995; Tibayrenc, 1998; Tibayrenc *et al.*, 1990), although some authors have reported the occurrence of sexual reproduction (Tait, 1980; Jenni *et al.*, 1986; Gibson, 1989; Gibson, 1995; MacLeod *et al.*, 2005a; MacLeod *et al.*, 2005b). The presence or absence of a sexual process will crucially determine the genetics at both, individual and population levels. Estimates of how genetic diversity is portioned within individuals (reproductive system), within and among subpopulations (population structure) may indicate how species track continuously varying environments and adapt to local conditions in the face of gene flow among diverse populations (Criscione *et al.*, 2005). Thus, a better understanding of the reproductive system of such organisms might be crucial for optimizing field control strategies (Tibayrenc *et al.*, 1991; Milgroom, 1996; Taylor *et al.*, 1999; Tibayrenc, 2005).

Recently, microsatellite markers were shown to be polymorphic enough to highlight the existence of genetic diversity within *T. b. gambiense* group 1 (Koffi *et al.*, 2007; Koffi *et al.*, 2009). In the present paper, we revisit the results presented in (Koffi *et al.*, 2009) with particular attention to reproductive modes of *T. brucei gambiense* 1 and other *T. brucei* species and on population size. We conclude on optimal sampling strategies that should further be undertaken to fully understand the population biology, hence epidemiology, of these important pathogens.

Material and Methods

Trypanosome isolates and genotying

Trypanosome isolates (one and more rarely two per patient) were taken from three geographical zones and four sampling dates: in Guinea, Boffa 2002, Dubreka 1998, and Dubreka 2002; in Côte d'Ivoire, Bonon 2000, Bonon 2002, and Bonon 2004 (Koffi *et al.*, 2009) (Figure 1). In Bonon the isolates were 17 in 2000, 14 in 2002 and 17 in 2004. In Guinea, the isolates were 15 in Dubreka 1998, seven in Dubreka 2002, and 20 in Boffa 2002. The study area in Bonon concerns 30,000 inhabitants distributed in 400 km², with an approximate mean prevalence

of 0.004 (Kaba *et al.*, 2006), leading to an estimate of about 120 infected persons (Table 1). In Boffa and Dubreka, these values were extrapolated from medical survey results (Camara *et al.*, 2005) taking into account evaluated population at risk (Table 1), and lead to estimates of 187 and 295 infected persons, in Dubreka and Boffa respectively.

We studied seven microsatellite loci: *M6c8*, *Mt3033* (Biteau *et al.*, 2000), *Micbg1*, *Micbg5*, *Micbg6*, *Misatg4* and *Misatg9* (Koffi *et al.*, 2007). Complete genotypes and multilocus genotypes (MLG) are given in (Koffi *et al.*, 2009). Because *Micbg6* did not vary across all samples (all individuals displayed the same genotype), this locus was removed from the data set in further analyses, except when specified.

Data analysis

The most widely used parameters to infer population structure are the *F*-statistics (Wright, 1951); e.g., (Nagylaki, 1998)). Typically, these parameters are defined for three hierarchical levels. F_{IS} measures the identity (or homozygosity) of alleles within individuals within sub-populations relative to that measured between individuals; it is thus a measure of deviation from local panmixia (random union of gametes producing zygotes). It varies between -1 (single class of heterozygote) as expected in a very small and isolated clonal population (De Meeûs, Balloux, 2005) and +1 (all individuals are homozygous for different alleles), as expected in fully selfing species; and it equals 0 in panmictic populations. F_{ST} measures the identity between individuals within sub-populations, as compared to individuals from other sub-populations within the total population, or the total relative homozygosity due to the Wahlund effect (Wahlund, 1928). It is thus a convenient measure of differentiation between sub-populations that varies between 0 (no structure) and 1 (all populations fixed for one or other allele). These *F*-statistics were estimated by Weir and Cockerham's unbiased estimators (Weir, Cockerham, 1984), with FSTAT version 2.9.3.2 ((Goudet, 2001), updated from (Goudet, 1995)), and their significant deviation from 0 was tested by randomizing alleles between individuals within subsamples and randomizing individuals among sub-samples. Randomizations were set to 10,000 and implemented by FSTAT 2.9.3.2.

To get an encompassing picture of genotypic distribution across space, time, and subspecies, a NJTREE was computed by the MEGA 3.1 software (Kumar et al. 2005, updated from (Kumar *et al.*, 2004)). As recommended (e.g., (Takezaki, Nei, 1996; De Meeûs *et al.*, 2007)), the tree was built according to a Cavalli-Sforza and Edwards chord distance

matrix (Cavalli-Sforza, Edwards, 1967) computed with Genetix 4.05 (Belkhir *et al.*, 2004).

In purely clonal populations, a negative F_{IS}, homogeneous across loci, is expected, while a substantial variance of F_{IS} across loci can be the sign of rare sexual events (De Meeûs *et al.*, 2006). Nevertheless, in clonal organisms F_{IS} is also dependent on mutation rate (the higher mutation rate the higher F_{IS}) (De Meeûs *et al.*, 2006). Consequently, if mutation rate varies from one locus to the other, so will the F_{IS}. Mutation rate of trypanosome microsatellites is unknown but a direct positive relationship is expected between mutation rate and genetic diversity, hence between F_{IS} and, for instance, Nei's unbiased estimate of genetic diversity H_s (Nei, Chesser, 1983). We tested this with a linear regression under R software (R-Development-core-team, 2008).

Inferring clonal sub-population size and migration rate

We used the model developed by Balloux et al. (Balloux *et al.*, 2003). Consider a subdivided monoecious population of diploid individuals, like *T. brucei gambiense* are (Koffi *et al.*, 2009), with non-overlapping generations. Individuals reproduce clonally with probability c and sexually with probability (1-c). Self-fertilization occurs at a rate s. There are n sub-populations, or demes, each composed of N individuals. Migration between the subpopulations follows an island model (Wright, 1951), with a migration rate m. The mutation rate is u for all alleles and therefore the probability that two alleles, identical by descent before mutation, are still identical after mutation is $\gamma=(1-u)^2$. We further assume stable census sizes and no selection and a two-population framework, which we assume being the case in both Côte d'Ivoire and Guinea, with total clonality.

Three probabilities of identity by descent can be defined: Q_i, the probability that two alleles drawn at random from a single individual are identical by descent; Q_s□, the probability that two randomly sampled alleles from two different individuals within a subpopulation are identical by descent; and Q_T□, the probability that two randomly sampled alleles from two individuals in different subpopulations are identical by descent.

The recurrence equations between generations t and t+1 for the different identities by descent are:

$$\begin{cases} Q_{i(t+1)} = \gamma\left\{ cQ_{i(t)} + (1-c)\left[s\left(\frac{1+Q_{i(t)}}{2} \right) + (1-s)Q_{s(t)} \right] \right\} \\ Q_{s(t+1)} = \gamma\left\{ q_s\left[\frac{1}{N}\left(\frac{1+Q_{i(t)}}{2} \right) + \left(1-\frac{1}{N}Q_{s(t)} \right) \right] + (1-q_s)Q_{T(t)} \right\} \\ Q_{T(t+1)} = \gamma\left\{ q_d\left[\frac{1}{N}\left(\frac{1+Q_{i(t)}}{2} \right) + \left(1-\frac{1}{N} \right)Q_{s(t)} \right] + (1-q_d)Q_{T(t)} \right\} \end{cases}$$

with

$$\begin{cases} q_s \cong (1-m)^2 + \frac{m^2}{n-1} \\ q_d = \frac{1-q_s}{n-1} \end{cases}$$

and where q_s is the probability that two individuals taken at random within the same sub-population after migration were born in the same subpopulation and q_d the probability that two individuals sampled after migration in different sub-populations originated from the same subpopulation (Wang, 1997).

Wright's *F*-statistics, can be defined following Cockerham (Cockerham, 1969; Cockerham, 1973) as:

$$\begin{cases} F_{IS} = \frac{Q_i - Q_s}{1-Q_s} \\ F_{ST} = \frac{Q_s - Q_T}{1-Q_T} \\ F_{IT} = \frac{Q_i - Q_T}{1-Q_T} \end{cases}$$

Assuming no selfing (i.e. s=1/N), the systems of equations (1), (2) and (3) lead to:

$$\begin{cases} F_{IS} = \dfrac{\gamma\{q_s - c[\gamma(q_s - q_d) - 1] - 1\}}{2N(1 - c\gamma)[\gamma(q_s - q_d) - 1] - \gamma\{q_s - c[\gamma(q_s - q_d) - 1] - 1\}} \\ F_{ST} = \dfrac{\gamma(1 - c\gamma)(q_s - q_d)}{[2N(1 - c\gamma) - \gamma(1 - c)][1 - \gamma(q_s - q_d)] + \gamma\{q_d[\gamma(c + 1) - 1] - q_s[\gamma(c + 1) - 2]\}} \end{cases}$$

In Côte d'Ivoire there are two foci, an thus two putative subpopulations, Bonon and Sinfra, as is the case for Guinea (Dubreka and Boffa) (Koffi *et al.*, 2009). In a two sub-populations framework with total clonality (n=2, c=1), as it is probably the case in the two areas investigated in the present study, we get:

$$\begin{cases} q_s = (1 - m)^2 + m^2 = 1 - 2m(1 - m) \\ q_d = 1 - (1 - m)^2 - m^2 = 2m(1 - m) = 1 - q_s \end{cases}$$

and combining equations (4) and (5):

$$\begin{cases} F_{IS} = \dfrac{\gamma\{q_s - \gamma(2q_s - 1)\}}{2N(1 - \gamma)[\gamma(2q_s - 1) - 1] - \gamma\{q_s - \gamma(2q_s - 1)\}} \\ F_{ST} = \dfrac{\gamma(1 - \gamma)(2q_s - 1)}{[2N(1 - \gamma)][1 - \gamma(2q_s - 1)] + \gamma\{(1 - q_s)[2\gamma - 1] + 2q_s[1 - \gamma]\}} \end{cases}$$

After neglecting terms in u^2 and u (<<1 or q_s) and simplifications, these equations can be rearranged into:

$$\begin{cases} F_{IS} = -\dfrac{1 - q_s}{(1 - q_s)(1 + 8Nu)} \\ F_{ST} = \dfrac{2u(2q_s - 1)}{(1 - q_s)(8Nu + 1)} \end{cases}$$

From equation (7) it is easy to see that when $q_s \neq 1$ (i.e., m is in]0,1[) F_{IS} becomes independent from migration and can provide an estimate for N in the simple form:

$$N = -\frac{1 + F_{IS}}{8uF_{IS}}$$

If we combine (7) and (8) we can also obtain an estimate for q_s:

$$q_s = \frac{1 + F_{ST}\dfrac{8Nu+1}{2u}}{2 + F_{ST}\dfrac{8Nu+1}{2u}}$$

Because we are in a two populations case, the genetic effect of migration is symmetric around 0.5 (m=0.49 is equivalent to m=0.51). We can thus focus on values below 0.5 for m. From there it is easy to see from (5) that:

$$m = \frac{1}{2}\left[1 - \sqrt{2q_s - 1}\right]$$

and thus combining (9) and (10) gives us access to m as:

$$m = \frac{1}{2}\left[1 - \sqrt{\frac{F_{ST}\dfrac{8Nu+1}{2u}}{2 + F_{ST}\dfrac{8Nu+1}{2u}}}\right]$$

that can be finally combined with (8) to obtain:

$$m = \frac{1}{2}\left[1 - \sqrt{\frac{F_{ST}}{F_{ST} - 4uF_{IS}}}\right]$$

Inferring clonal effective population size

Temporal samples offer the opportunity to estimate effective population sizes (N_e, the size of panmictic adults required to drift at the same rate as the observed population) with the method developed by Waples (Waples, 1989) and implemented in NeEstimator v 1.3 (Peel *et al.*, 2004). For this purpose, we only used the multilocus genotypes data (MLG), which we rendered diploid by duplication of the allele of the single artificial locus obtained. MLG's were chosen because in clonal organisms all loci are linked and heterozygosity excess affects differentiation estimates (De Meeûs *et al.*, 2006). In order to estimate the number of trypanosome generations passed within the time windows (two and four years), we used two drastically different methods. The first method assumes that populations are mainly defined as the infra-populations of cells contained in each individual host. In that case, generation time must be close to the time between two cell divisions. *T.*

brucei cells divide every 5.7 hours (Salmon *et al.*, 2005), which yields 4.2 generations per day and thus 1537 per year. The second method assumes that each host is colonized by a limited number of strains (~1) and that the generation time corresponds more to the time it takes for a human individual newly infected by a trypanosome after a tsetse bite to become infectious for a new tsetse, and for this second tsetse to become infectious to a human individual again. Incubation in human hosts lasts on average 25 days (Gouteux, Artzrouni, 2000), while between 12 and 24 days are required for a newly infected tsetse fly to become infectious for a vertebrate host (Dale *et al.*, 1995; Van den Abbeele *et al.*, 1999). This gives a generation time window of 37 to 49 days for trypanosomes, leading to 7-10 generations per year.

Results

Heterozygosity within sub-samples

Nearly all stocks were heterozygous at each microsatellite locus. Mean F_{IS}=-0.62, the relative heterozygosity measured within subsamples, is strongly negative (strong heterozygote excess as compared to Hardy-Weinberg expectations) with small variance across loci that is almost entirely explained by the genetic diversity maintained at each locus and hence by mutation rates (Figure 2), meaning that individuals are extremely heterozygous at all loci (genome wide heterozygous state) and populations are totally clonal.

In Table 1 it can be seen that no positive relationship exists regarding the surface of the investigated geographical areas, prevalence of infection, or number of infected persons. Moreover, GPS data for Bonon 2000 and 2002 can be used to build groups of trypanosomes from infected patients from different sub-areas in each zone. The F_{IS} computed for the six loci is extremely close (and indeed higher) to the one computed without GPS coordinates so that Wahlund effects can safely be disregarded (Koffi *et al.*, 2009).

Genetic differentiation

In Figure 3 it can be seen that trypanosome strains first differentiate between countries, then between sites (in Guinea), between temporal samples (apparently more pronounced in Dubreka, Guinea, than in Bonon, Côte d'Ivoire), and that the sampling method does not have any impact.

Population sizes.

From the F_{IS} analysis, according to De Meeûs *et al.* criteria (De Meeûs *et al.*, 2006), full clonality can be assumed for *T. brucei gambiense* group 1 for the studied populations. According to (Hellegren, 2000), microsatellite mutation rates mostly range between 10^{-3} and 10^{-4}. We use these two values for estimating clonal effective population sizes with equation (8) of Material and Methods. The results are presented in Figure 4.

If we assume that generation time corresponds to cell divisions, Waples moment-based method (Waples, 1989) gives huge estimates (N_e≈12,000-30,000 cells). During the surveys, it was observed that the expected number of trypanosome cells would approximately range from 7,000,000 in Guinea to 400,000,000 in Bonon. With such values a glance at equation (7) leads to an expected F_{IS}~0, which is far from being the case. Thus the cell is not the individual unit and the individual host does not correspond to a demographic unit (population) for trypanosomes. Figure 4 also presents the results obtained with the trypanosome life-cycle-based method for generation time (37 to 49 days, see Material and Methods). For temporal MLG-based estimates, the values obtained are probably much smaller than the "real" N_e, as indicated by the extremely high upper bounds of the 95% confidence intervals, so that the F_{IS}-based method is probably more accurate, as suggested in general from theoretical analyses of fully clonal populations (De Meeûs, Balloux, 2005).

From Figure 4, the estimated numbers of infected patients seem to almost perfectly match all N_e estimates. This is unexpected if infected patients are to reflect *T. brucei gambiense* group 1 census sizes, which should be at least slightly over N_e.

With u=10^{-4}, F_{IS}-based N_e reaches 297, 760 and 1479 for Boffa, Bonon and Dubréka respectively. These values match well other estimates in Boffa, but clearly surpass the observed number of infected patients in Bonon and Dubreka, for which a reasonable match is reached with u=10^{-3}.

Discussion

According to the De Meeûs *et al.* criterion (De Meeûs *et al.*, 2006), if some sex occurred, even very rarely, a higher F_{IS} with a much stronger variance of F_{IS} across loci would have been observed. We must conclude that the populations studied never sexually recombined in a reasonable length of time. Our results also indicate that within each country, *T. b. gambiense* group 1 populations are relatively small and do

not exchange many migrants. For instance, in Guinea, where two sites were sampled in 2002, with equations (8) and (12) from our model (Material and Methods) and a reasonable mutation rate of $u=10^{-4}$, the effective clonal population size and migration rate respectively are N_{ec}=297 and m=0.001 in Boffa and N_{ec}=1479 and m=0.0008 in Dubreka. Obviously, migration is weak.

At the scale of West Africa (between Côte d'Ivoire and Guinea), our results show that any strain transfer between the two countries is too rare to leave any signature in the investigated microsatellite polymorphism (no MLG in common).

The effective population size results support a complete parasitic-cycle-based generation time and reject a cell-based generation time. Population regulation thus occurs at the scale of a focus. This means that even if multiple infection of the same patient is not rare, at the scale of the focus, it has no demographic influence; probably because the vector will itself always harbors a single genotype (trypanosome individual). This is in line with the absence of recombination found in this study and suggests that tsetse flies represent the main regulatory factor of *T. brucei gambiense* 1. Interestingly enough, the effective population size and migration rate computed for *Glossina palpalis gambiensis* in Dubreka (Solano *et al.*, 2009) was around N_e~1000 and m~0.005 (recomputed here) and are very close to what is found for the parasite assuming a mutation ate of 10^{-4}. As discussed in (Koffi *et al.*, 2009), the clonal population sizes found here strongly support the existence of hidden hosts (animal reservoirs or asymptomatic infected humans) (Jamonneau *et al.*, 2004). These hidden hosts may represent a potential parasite reservoir that could be responsible for the persistence of transmission and re-emergence of sleeping sickness. It would be of interest to sample both healthy humans and animals living next to the HAT cases and identify with microsatellite loci the trypanosomes they may harbor.

Control of the disease at a country scale would probably be efficient in the long term before new strains reinvade the area. Nevertheless, our data also reveal a high degree of local genetic polymorphism, either because of larger population sizes than epidemiological surveys can account for, or because of high mutation rates, which suggests that *T. b. gambiense* may quickly respond to new selective pressures, such as the one imposed by chemical treatment with a new drug.

Finally, the strong heterogeneity found in other sub-species raises a doubt on the unity of each of these taxa. This is particularly

interesting for *T. brucei brucei* that is currently assumed to experience frequent recombination, hidden by an epidemic population structure (MacLeod *et al.*, 2000). But mixing different taxa could also lead to much more confusion (Prugnolle, De Meeûs, 2009) so this issue will require further investigations like differentiation tests as a function of host species.

Acknowledgements

We particularly acknowledge all the technicians from the HAT team of the Institut Pierre Richet (Abidjan, Côte d'Ivoire), and the HAT National Control Programme of Côte d'Ivoire and Guinea for their help in sampling. We thank the Service de Coopération et d'Action Culturelle d'Abidjan (SCAC), the Département de Soutien de Formation (DSF) de l'Institut de Recherche pour le Développement (IRD), the Ministère Français des Affaires Etrangères (Fonds de Solidarité Prioritaire « Recherches en Entomologie, Formation et Stratégies de formation, le cas du paludisme et de la Trypanosomose Humaine Africaine »), and the World Health Organisation (WHO) for their support. We would also like to thank Christian Barnabé who undertook the bootstrap analysis for Figure 3.

References

Balloux F, Lehmann L, De Meeûs T (2003) The population genetics of clonal and partially clonal diploids. *Genetics* **164**, 1635-1644.

Belkhir K, Borsa P, Chikhi L, Raufaste N, Bonhomme F (2004) GENETIX 4.05, logiciel sous Windows TM pour la génétique des populations. Laboratoire Génome, Populations, Interactions, CNRS UMR 5000, Université de Montpellier II, Montpellier (France).

Biteau N, Bringaud F, Gibson W, Truc P, Baltz T (2000) Characterization of Trypanozoon isolates using a repeated coding sequence and microsatellite markers. *Molecular and Biochemical Parasitology* **105**, 185-201.

Camara M, Kaba D, KagbaDouno M*, et al.* (2005) Human African trypanosomiasis in the mangrove forest in Guinea: epidemiological and clinical features in two adjacent outbreak areas. *Médecine Tropicale* **65**, 155-161.

Cavalli-Sforza LL, Edwards AWF (1967) Phylogenetic analysis: model and estimation procedures. *American Journal of Human Genetics* **19**, 233-257.

Cockerham CC (1969) Variance of gene frequencies. *Evolution, International Journal of Organismal Evolution* **23**, 72–84.

Cockerham CC (1973) Analysis of gene frequencies. *Genetics* **74**, 679–700.

Criscione CD, Poulin R, Blouin MS (2005) Molecular ecology of parasites: elucidating ecological and microevolutionary processes. *Molecular Ecology* **14**, 2247-2257.

Dale C, Welburn SC, Maudlin I, Milligan PJM (1995) The kinetics of maturation of trypanosome infections in tsetse. *Parasitology* **111**, 187-191.

De Meeûs T, Balloux F (2005) F-statistics of clonal diploids structured in numerous demes. *Molecular Ecology* **14**, 2695-2702.

De Meeûs T, Lehmann L, Balloux F (2006) Molecular epidemiology of clonal diploids: A quick overview and a short DIY (do it yourself) notice. *Infection Genetics and Evolution* **6**, 163-170.

De Meeûs T, McCoy KD, Prugnolle F*, et al.* (2007) Population genetics and molecular epidemiology or how to "débusquer la bête". *Infection Genetics and Evolution* **7**, 308-332.

Gibson WC (1989) Analysis of a genetic cross between *Trypanosoma brucei rhodesiense* and *T. b. brucei*. *Parasitology International* **99**, 391-402.

Gibson WC (1995) The significance of genetic exchange in trypanosomes. *Parasitology Today* **11**, 465-468.

Gibson WC (2007) Resolution of the species problem in African trypanosomes. *International Journal for Parasitology* **37**, 829-838.

Goudet J (1995) FSTAT (Version 1.2): A computer program to calculate F-statistics. *Journal of Heredity* **86**, 485-486.

Goudet J (2001) FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3). Available from http://www.unil.ch/izea/softwares/fstat.html. Updated from Goudet (1995).

Gouteux JP, Artzrouni M (2000) Persistence and resurgence of sleeping sickness foci. A biomathematical investigation. *Comptes Rendus de l'Académie des Sciences Série III, Sciences de la Vie-Life Sciences* **323**, 351-364.

Hellegren H (2000) Microsatellite mutations in the germline: implications for evolutionary inference. *Trends in Genetics* **16**, 551-558.

Hoare CB (1972) *The trypanosomes of mammals. A zoological monograph* Blackwell, Oxford.

Jamonneau V, Ravel S, Koffi M, *et al.* (2004) Mixed infections of trypanosomes in tsetse and pigs and their epidemiological significance in a sleeping sickness focus of Cote d'Ivoire. *Parasitology* **129**, 693-702.

Jenni L, Marti S, Schweizer J, *et al.* (1986) Hybrid Formation between African Trypanosomes During Cyclical Transmission. *Nature* **322**, 173-175.

Kaba D, Dje NN, Courtin F, *et al.* (2006) The impact of war on the evolution of sleeping sickness in west-central Cote d'Ivoire. *Tropical Medicine & International Health* **11**, 136-143.

Koffi M, De Meeûs T, Bucheton B, *et al.* (2009) Population genetics of *Trypanosoma brucei gambiense*, the agent of sleeping sickness in Western Africa. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 209-214.

Koffi M, Solano P, Barnabé C, *et al.* (2007) Genetic characterisation of *Trypanosoma brucei* ssp by microsatellite typing: new perspectives for the molecular epidemiology of human African trypanosomosis. *Infection Genetics and Evolution* **7**, 675-684.

Kumar S, Tamura K, Nei M (2004) MEGA3: Integrated software for molecular evolutionary genetics analysis and sequence alignment. *Briefings in Bioinformatics* **5**, 150-163.

MacLeod A, Tweedie A, McLellan S, *et al.* (2005a) Allelic segregation and independent assortment in *T. brucei* crosses: Proof that the genetic system is Mendelian and involves meiosis (vol 143, pg 12, 2005). *Molecular and Biochemical Parasitology* **144**, 131-131.

MacLeod A, Tweedie A, McLellan S, *et al.* (2005b) Allelic segregation and independent assortment in *T. brucei* crosses: Proof that the genetic system is Mendelian and involves meiosis. *Molecular and Biochemical Parasitology* **143**, 12-19.

MacLeod A, Tweedie A, Welburn SC, *et al.* (2000) Minisatellite marker analysis of *Trypanosoma brucei*: Reconciliation of clonal, panmictic, and epidemic population genetic structures. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 13442-13447.

Milgroom MG (1996) Recombination and the multilocus structure of fungal populations. *Annual Review of Phytopathology* **34**, 457-477.

Nagylaki T (1998) Fixation indices in subdivided populations. *Genetics* **148**, 1325-1332.

Nei M, Chesser RK (1983) Estimation of fixation indices and gene diversities. *Annals of Human Genetics* **47**, 253-259.

Peel D, Ovenden JR, Peel SL (2004) NeEstimator Version 1.3: software for estimating effective population size. Queensland Government, Department of Primary Industries and Fisheries.

Prugnolle F, De Meeûs T (2009) Apparent high recombination rates in clonal parasitic organisms due to inappropriate sampling design. *Heredity* **In press**.

R-Development-core-team (2008) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, http://www.R-project.org, ISBN 3-900051-07-0.

Salmon D, Paturiaux-Hanocq F, Poelvoorde P, Vanhamme L, Pays E (2005) *Trypanosoma brucei*: growth differences in different mammalian sera are not due to the species-specificity of transferrin. *Experimental Parasitology* **109**, 188-194.

Solano P, Ravel S, Bouyer J*, et al.* (2009) The population structure of *Glossina palpalis gambiensis* from island and continental locations in coastal Guinea. *PLoS Neglected Tropical Diseases* **3**, e392.

Tait A (1980) Evidence for diploïdy and mating in trypanosomes. *Nature (London)* **287**, 536-538.

Takezaki N, Nei M (1996) Genetic distances and reconstruction of phylogenetic trees from microsatellite DNA. *Genetics* **144**, 389-399.

Taylor JW, Geiser DM, Burt A, Koufopanou V (1999) The evolutionary biology and population genetics underlying fungal strain typing. *Clinical Microbiology Reviews* **12**, 126-146.

Tibayrenc M (2005) Bridging the gap between molecular epidemiologists and evolutionists. *Trends in Microbiology* **13**, 575-580.

Tibayrenc M, Kjellberg F, Arnaud J*, et al.* (1991) Are eukaryotic microorganisms clonal or sexual? A population genetics vantage. *Proceedings of the National Academy of Sciences of the United States of America* **88**.

Van den Abbeele J, Claes Y, van Bockstaele D, Le Ray D, Coosemans M (1999) *Trypanosoma brucei* spp. development in the tsetse fly: characterization of the post-mesocyclic stages in the foregut and proboscis. *Parasitology* **118**, 469-478.

Wahlund S (1928) Zusammensetzung von populationen und korrelationsers-chinungen von standpunkt der vererbungslehre aus betrachtet. *Hereditas* **11**, 65-106.

Wang J (1997) Effective size and F-statistics of subdivided populations I. Monoecious species with partial selfing. *Genetics* **146**, 1453-1463.

Waples RS (1989) A generalized approach for estimating effective population size from temporal changes in allele frequency. *Genetics* **121**, 379-391.

Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. *Evolution* **38**, 1358-1370.

Wright S (1951) The genetical structure of populations. *Annals of Eugenics* **15**, 323-354.

Table 1: Data from epidemiological surveys of the investigated areas and estimated F_{IS} obtained with the six most reliable loci in the different *T. brucei gambiense* sub-samples. F_{IS} is a standardized measure of heterozygosity deviation, expexted nul if reproduction is sexual and random, its value is influenced by reproductive mode and/or undetected subdivision within subsamples. Prevalence is the ratio between the number of infected persons (Infected) to the number of persons examined.

Sub-sample	Surface of study (in km^2)	Human population	Prevalence	Infected	F_{IS}
Bonon 2000	400	30000	0.004	120	-0.671
Bonon 2002	400	30000	0.004	120	-0.645
Bonon 2004	400	30000	0.004	120	-0.555
Dubreka 1998	1600	25000	0.0075	187	-0.440
Dubreka 2002	1600	25000	0.0075	187	-0.505
Boffa 2002	2400	25000	0.0118	295	-0.808

Figure Legends

Figure 1: Localization of sampling areas (stars). (Drawing by Fabrice Courtin)

Figure 2: Regrassion of F_{IS} per locus against their genetic diversity H_s. For this analysis only the values obtained for locus *Micbg6* (F_{IS}=-1, H_s=0.5) were used.

Figure 3: Rooted NJTREE of the different isolates combined with reference strains available from the Supplementary Table 2 of (Koffi *et al.*, 2006a) article where the complete information and origins, year and publication references can be found in their Table 3. The tree is based on Cavali-Sforza and Edwards' chord distances computed on the eight loci described in Koffi et al ige. *T. brucei gambiense* 1 reference strains are in purple and all included in one cluster within the sub-tree comprising all strains studied in the present paper (in black). *T. brucei gambiense* 2 reference strains are in red, *T. brucei rhodesiense* in blue and *T. brucei brucei* in green. This particular tree was rooted with strain Feo of *T. brucei brucei* but any other strain of this sub-species or

of *T. brucei rhodesiense* equally illustrates the monophyly of *T. brucei gambiense* 1 and polyphyly of all other *Trypanossoma brucei* types. Bootstrap values of principal nodes (above 750 ‰) are given (obtained with Phylip 3.68, J. Felsestein 2008)

Figure 4: Effective population size (N_e) obtained with the F_{IS}-based method (see Material and Methods Equation 1) ("model"), with $u=10^{-3}$ and $u=10^{-4}$, and with Waples' method from temporally spaced samples (with MLG as a single locus), using trypanosome's life cycle as the generation time with the shortest (sgt=37 days) or largest (lgt=49 days) generation times (see text). Black squares are the means with 95% confidence intervals (CI) (small lines) (averaged over 2000-2002, 2000-2004 and 2002-2004 for Bonon). The dotted line corresponds to the estimated number of infected persons in the different areas according to epidemiological surveys. For Waples' method, CI come from a chi-square distribution with a degrees of freedom (a is the number of alleles, here of different MLG's) (Waples, 1989). For the F_{IS}-based method, confidence intervals correspond to those of F_{IS} obtained by bootstrap over loci.

Figure 1


SENEGAL
MALI
Bamako
GUINEE BISSAU
GUINEA
BURKINA FASO
Boffa
Dubréka
Conakry
SIERRA LEONE
Freetown
COTE D'IVOIRE
Bonon
15° West
LIBERIA
GHANA
Monrovia
5° North
Abidjan
Atlantic ocean
100 0 100 200 Kilometers

Figure 2

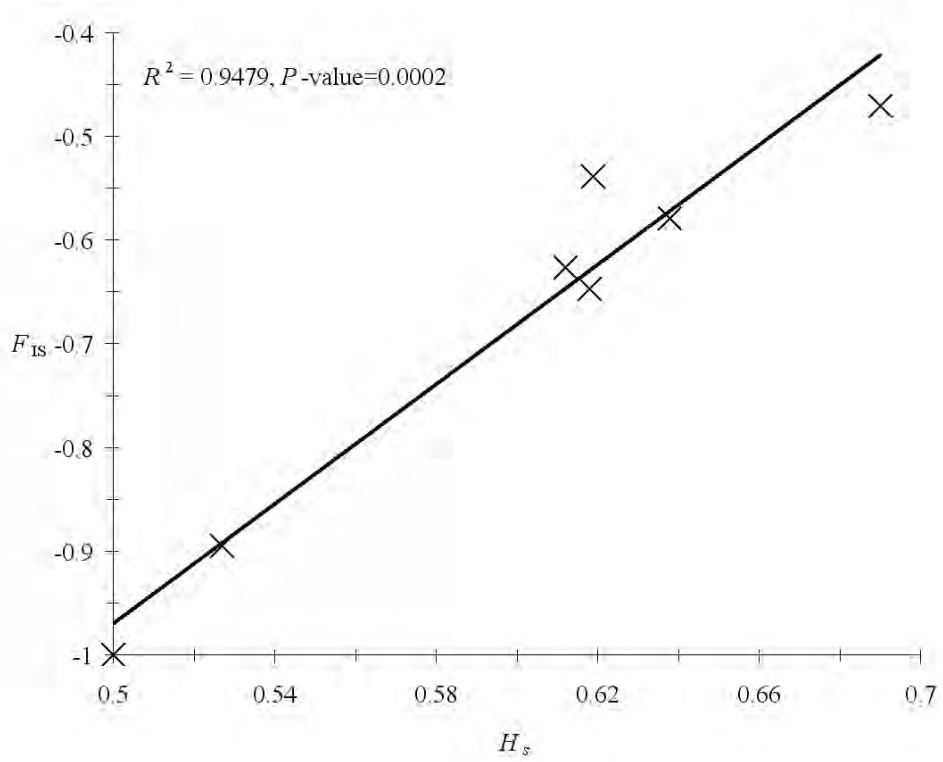

R^2 = 0.9479, P-value=0.0002
-0.4
-0.5
-0.6
-0.7
-0.8
-0.9
-1
F_{IS}
0.5
0.54
0.58
0.62
0.66
0.7
H_S

Figure 3

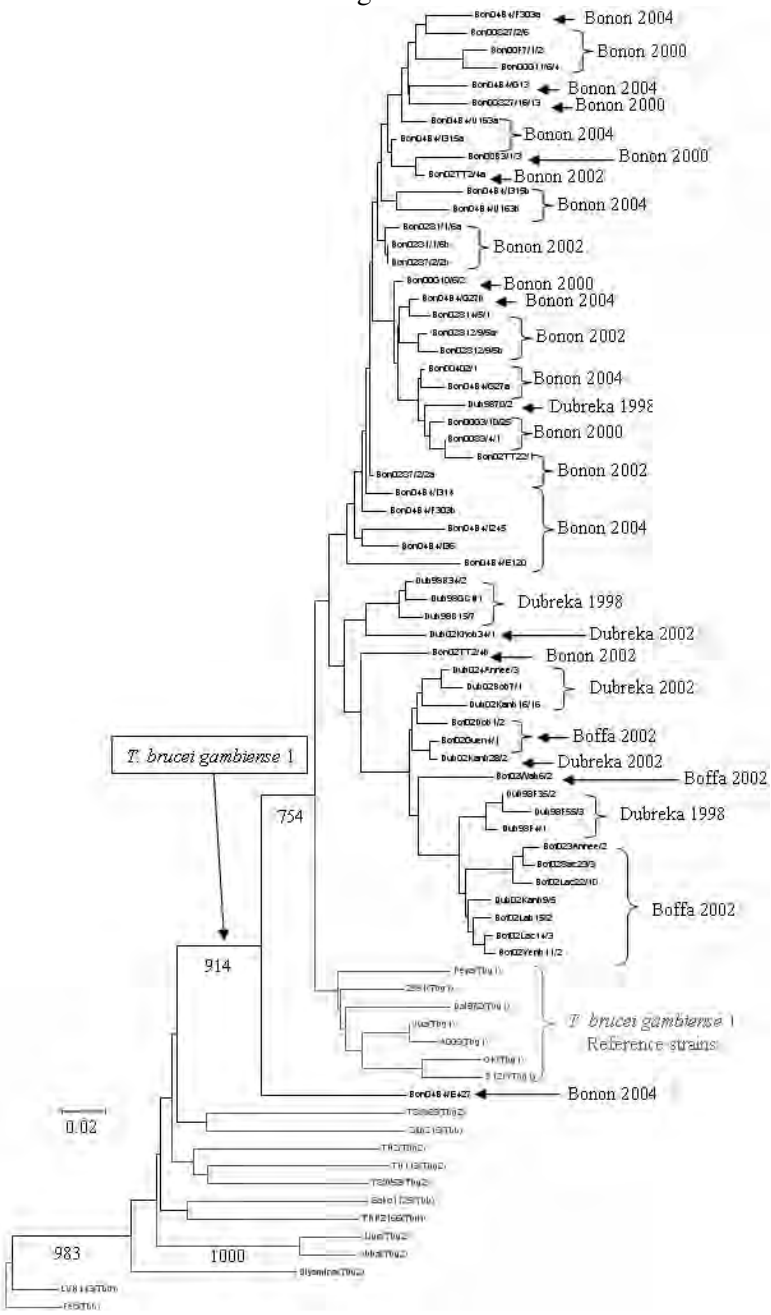
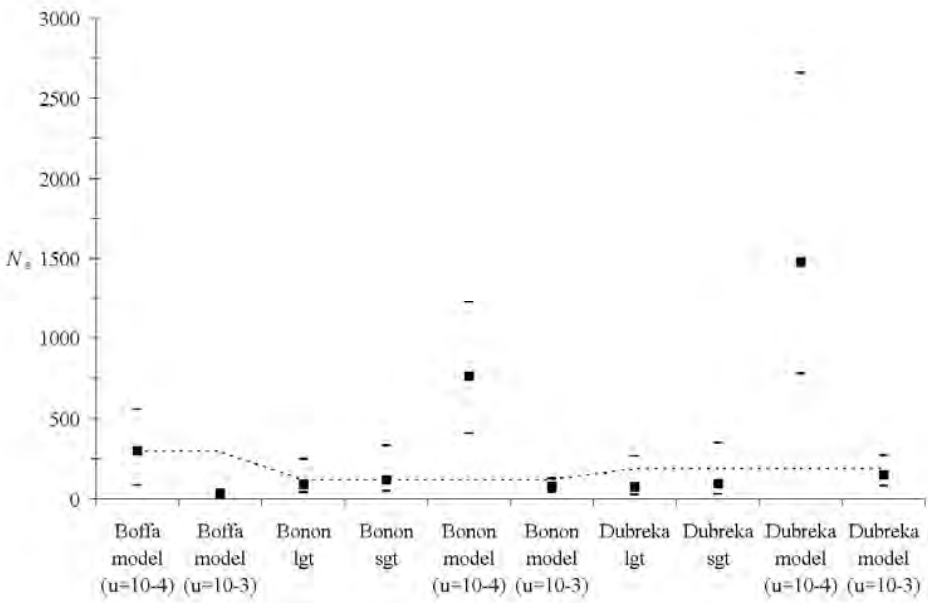

Bonon 2004
Bonon 2000
Bonon 2004
Bonon 2000
Bonon 2004
Bonon 2000
Bonon 2002
Bonon 2004
Bonon 2002
Bonon 2000
Bonon 2004
Bonon 2004
Dubreka 1998
Bonon 2000
Bonon 2002
Bonon 2004
Dubreka 1998
Dubreka 2002
Bonon 2002
Dubreka 2002
Boffa 2002
Dubreka 2002
Boffa 2002
Dubreka 1998
Boffa 2002
T. brucei gambiense 1
754
914
T. brucei gambiense 1
Reference strains
Bonon 2004
0.02
983
1000

Figure 4


3000
2500
2000
N_s 1500
1000
500
0
Boffa model (u=10-4)
Boffa model (u=10-3)
Bonon lgt
Bonon sgt
Bonon model (u=10-4)
Bonon model (u=10-3)
Dubreka lgt
Dubreka sgt
Dubreka model (u=10-4)
Dubreka model (u=10-3)