Population genetics of *Trypanosoma brucei gambiense*, the agent of sleeping sickness in Western Africa

Mathurin Koffi[a,b,c,1,2], Thierry De Meeûs[a,d,1,2], Bruno Bucheton[a,b], Philippe Solano[a,b], Mamadou Camara[e], Dramane Kaba[c], Gérard Cuny[a], Francisco J. Ayala[f,3], and Vincent Jamonneau[a,b]

[a]Institut de Recherche pour le Développement, Unité Mixte de Recherche IRD-CIRAD 177, Campus International de Baillarguet, 34398 Montpellier, Cedex 5, France; [b]Centre International de Recherche-Développement sur l'Elevage en Zone Subhumide, Unité de Recherches sur les Bases Biologiques de la Lutte Intégrée, BP 454 Bobo-Dioulasso 01, Burkina Faso; [d]Génétique et Evolution des Maladies Infectieuses, UMR CNRS/IRD 2724, UR IRD 165, Centre IRD de Montpellier, 34394 Montpellier, Cedex 5, France; [e]Programme National de Lutte Contre la Trypanosomose Humaine Africaine, BP 851 Conakry, Guinea; [c]Institut Pierre Richet, Unité de Recherche ''Trypanosomoses'', BP 293 Abidjan 04, Ivory Coast; and [f]Department of Ecology and Evolutionary Biology, University of California, Irvine, CA 92697

Human African trypanosomiasis, or sleeping sickness caused by *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 8 microsatellite loci of *T. b. gambiense* strains isolated from human African trypanosomiasis patients in the Ivory Coast and Guinea, with the aim of knowing how genetic information was partitioned within and between individuals in both temporal and spatial scales. 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.

clonality | effective population size | genetic differentiation | genetic diversity | microsatellite markers

Human African trypanosomiasis (HAT) or sleeping sickness is the third parasitic disease in subSaharan Africa regarding disability adjusted life years lost (1). The causative agent *Trypanosoma brucei* s.l., transmitted by Tsetse flies, is subdivided into 3 subspecies (2) on the basis of extrinsic criteria (host, clinical features, and geographical distribution), because these trypanosomes are morphologically identical: *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 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 (3, 4).

Trypanosoma brucei s.l. displays a huge diversity of adaptations 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 (5–7), although some investigators have reported the occurrence of sexual reproduction (3, 8–12). 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 (13–14). Thus, a better understanding of the reproductive system of such organisms might be crucial for optimizing field-control strategies (15–18) in a context of the HAT elimination process recently launched by the World Health Organization (19, 20).

Recently, microsatellite markers were shown to be polymorphic enough to highlight the existence of genetic diversity within the very homogeneous *T. b. gambiense* group 1 (21). In the present study, we present a microsatellite-based investigation of genetic polymorphism at different hierarchical levels: individual trypanosomes, within subsamples (identified by each focus), and between subsamples of *T. b. gambiense* group 1 in the Ivory Coast and Guinea (Fig. 1) and between temporally spaced data. We infer the extent of clonal reproduction and population subdivision that our analyses reveal, and discuss future directions of research and sampling strategies that could enhance the understanding of the epidemiology of this disease.

Results

Linkage Disequilibrium Between Loci. Linkage disequilibrium between pairs of loci was tested for the 7 loci varying across subsamples [*Micbg6* excluded, see supporting information (SI) Table S1] over all subsamples (Bonon, Boffa, and Dubreka of different years). There is a global strong linkage disequilibrium between loci as revealed by the impressive proportion of significant associations (18 out of 21) (Table S2), even with the highly conservative sequential Bonferroni level (see *Materials and Methods*) (12 significant tests). Each locus is involved in at least one significant linkage.

Heterozygosity Within Subsamples. Nearly all stocks were heterozygous at each microsatellite locus. One locus, *Trbpa1/2*, displayed an odd behavior and was removed from further analyses (see *Materials and Methods*). There is strong heterozygote excess as compared to Hardy-Weinberg expectations, with small variance

Fig. 1. Localization of sampling areas (*stars*). (Drawing by Fabrice Courtin, Bobo-Dioulasso, Burkina Faso).

Table 1. Data from epidemiological surveys of the investigated areas and estimated F_{IS} obtained with the 6 most reliable loci in the different *T. brucei gambiense* subsamples

Subsample	Surface of study (in km²)	Human population	Prevalence	Infected	F_{IS}
Bonon 2000	400	30,000	0.004	120	−0.671
Bonon 2002	400	30,000	0.004	120	−0.645
Bonon 2004	400	30,000	0.004	120	−0.555
Dubreka 1998	1,600	25,000	0.0075	187	−0.440
Dubreka 2002	1,600	25,000	0.0075	187	−0.505
Boffa 2002	2,400	25,000	0.0118	295	−0.808

F_{IS} is a standardized measure of heterozygosity deviation, expected to be null if reproduction is sexual and random; its value is influenced by reproductive mode or undetected subdivision within subsamples. Prevalence is the ratio between the number of infected persons (Infected) and the number of persons examined.

across loci, so mean $F_{IS} = -0.62$ (Fig. 2). Individuals are extremely heterozygous at all loci (genome-wide heterozygous state).

In Table 1 it can be seen that no positive relationship exists regarding the size of the investigated geographical areas, prevalence of infection, or number of infected persons. Moreover, GPS data for Bonon 2000 and 2002 (Table S3) can be used to b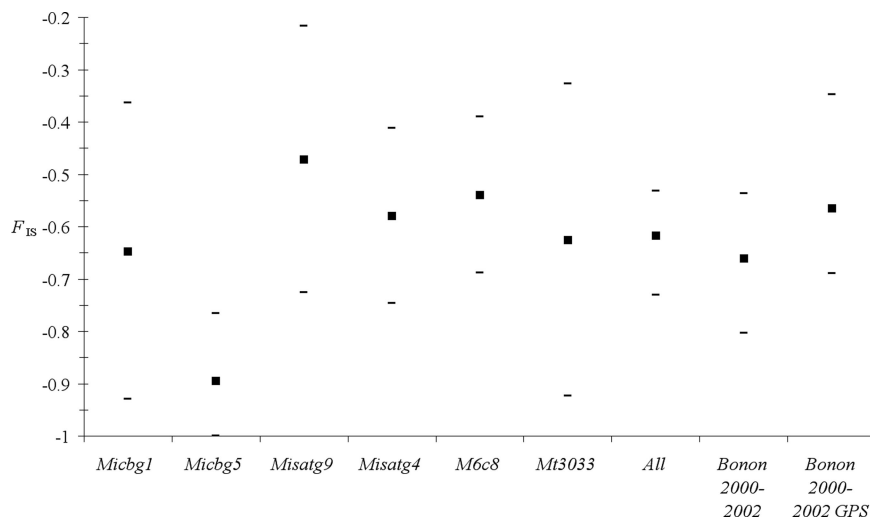uild groups of trypanosomes from infected patients from different subareas in each zone. The F_{IS} computed for the 6 loci is extremely close (and indeed higher) to the one computed without GPS coordinates (see Fig. 2).

Genetic Differentiation. In 2002, differentiation between Bonon (Ivory Coast) and the two Guinean sites was strong ($F_{ST} \approx 0.2$–0.3) and highly significant. It was also highly significant between the two Guinean samples, although to a much lesser extent (0.06, i.e., 3–5 times lower) (Table 2). Given the high degree of polymorphism found in these subsamples ($H_s = 0.62$), these levels of genetic differentiation are fairly high (the maximum possible fixation index is far below 1: $F_{ST\ max} \sim 0.4$). Because *T. brucei gambiense* group 1 is probably strongly clonal, we also used multilocus genotypes (MLGs; treating them as different alleles of a single locus, as defined in Table S1). MLGs yield small values of F_{ST} between countries (≤ 0.09) (see Table 2). But the standardized version of F_{ST} for multiple alleles, F_{ST}' (see *Materials and Methods*) indicates a maximum possible differentiation between Guinea and the Ivory Coast (in fact, there is not any MLG in common), and a fairly strong differentiation between the two Guinean localities (see Table 2).

Fig. 2. F_{IS} per locus and over all 7 polymorphic loci (All) (*Micbg6* and *Trbpa1/2* excluded), averaged over the 6 subsamples. The residual variation across the 6 remaining loci is mainly explained (91%) by the corresponding genetic diversity (in clonal populations a positive relationship is indeed expected, see ref. 22). For each locus, 95% confidence intervals (CI) of the means are estimated with the jackknife method over the populations' standard error. Over all loci, CI was obtained by bootstrap over loci. Mean F_{IS} and 95% bootstrap CI were also computed for Bonon 2000–2002 subsamples (Bonon) and compared to the F_{IS} computed when GPS coordinates of patients are taken into account (i.e., grouping the most proximate isolates into smaller subunits) (Bonon GPS).

Koffi *et al.*

Table 2. Differentiation between *T. brucei gambiense* sub-samples in space (2002) and in time (Bonon and Dubreka) as given by F_{ST} (a standardized measure of differentiation) based on the 6 microsatellite loci kept for the analysis and on the single multilocus haploid genotypes, as defined in the text

		F_{ST}	P-value	F_{ST}'
Spatial				
Guinea/Ivory Coast 2002	Bonon Boffa	0.289	0.0001	0.673
(1,100 km)		*0.090*	*0.0002*	*0.996*
	Bonon Dubreka	0.200	0.0001	0.512
		0.018	*0.1495*	*0.989*
Guinea 2002 (100 km)	Boffa Dubreka	0.059	0.0012	0.140
		0.045	*0.0213*	*0.565*
Temporal				
Dubreka	1998 2002	0.039	0.0327	0.119
		0.101	*0.0081*	*0.996*
Bonon	2000 2002	0.011	0.0210	0.027
		0.110	*0.0015*	*0.881*
	2000 2004	0.018	0.0053	0.049
		0.114	*0.0004*	*1*
	2002 2004	0.004	0.1867	0.009
		0.020	*0.1018*	*0.817*

Geographical distances between sites are given in parentheses. Meirmans' (23) standardized version of F_{ST}, F_{ST}' is also given. Multilocus haploid genotypes are in italics.

Trypanosome stocks were collected in 2000, 2002, and 2004 in Bonon and in 1998 and 2002 in Dubreka. Despite the 2- to 4-year window between subsequent samplings, differentiation is not spectacular within site (see Table 2). Except between 2002 and 2004 in Bonon, all subsamples are only slightly but significantly differentiated. The mean is $F_{ST} \approx 0.004–0.01$ for 2-year and $F_{ST} \approx 0.02$ to 0.04 for 4-year windows. The standardized F_{ST} are not large, either (~0.01–0.03 or ~0.05–0.12, respectively). This suggests that genetic drift is slow and, thus, that the effective population sizes are large. The F_{ST} based on MLGs seem to alter this picture, because F_{ST} ranges from 0.02 to 0.11 and 0.10 to 0.11 for 2-year and 4-year spans, respectively, while F_{ST}' ranges from 0.82 to 0.88 and 1, for 2 and 4 years, respectively. The latter observation implies that during a 4-year span, all MLGs are replaced by others through drift, mutation/migration, and treatment of patients.

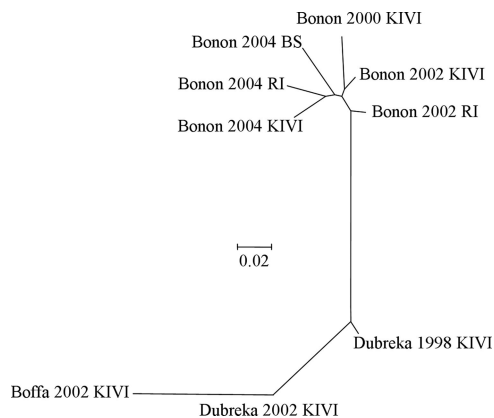Fig. 3 shows that trypanosome strains first differentiate be-

Fig. 3. Unrooted NJTREE representation of genetic distance between the different subsamples of *T. b. brucei* in Guinea (Boffa and Dubreka) and in the Ivory Coast (Bonon) in different years (1998, 2002, and 2004) and with different sampling techniques (KIVI, RI, and BS) using Cavalli-Sforza and Edwards' (24) chord-distance matrix.

Table 3. Trypanosome effective population sizes, estimated with Waples' moment-based method for temporally spaced data, and 95% CI when assuming that generation time corresponds to trypanosome's cell division

	N_e	CI	N_c
Bonon 2000 2002	11,964	4,990–33,858	240,000,000
Bonon 2002 2004	16,924	6,494–51,330	240,000,000
Bonon 2000 2004	26,701	10,844–66,471	240,000,000
Dubreka 1998 2002	15,143	4,429–53,961	7,500,000

CI comes from a chi-square distribution with *a* degrees of freedom, where *a* is the number of alleles, here of different MLG's (25). N_c is the order of magnitude of the number of trypanosome cells present in each focus, using the number of cells per infected patient in the different zones; it is given for comparison (see Table 1 and text for more details).

tween countries, then between sites (in Guinea), between temporal samples (apparently more pronounced in Dubreka and Guinea than in Bonon, Ivory Coast), and that the sampling method does not have any impact.

Effective Clonal Population Size. If we assume that generation time corresponds to cell divisions, Waples' moment-based method (25) gives huge estimates of effective population size ($N_e \approx 12,000–30,000$ cells). During the surveys, it was observed that most patients from Bonon were positive with the miniature anion-exchange/centrifugation technique (mAECT) (26), with a mean of 10 to 20 trypanosomes per mAECT (sometimes more than 100 trypanosomes per mAECT), corresponding approximately to 500 trypanosomes per ml (V.J., personal observation). In Guinea, the mAECT technique is often negative (90%; V.J., personal observation) and patients are generally diagnosed by lymph-node puncture (27). Given the detection threshold of mAECT (28), we can assume a maximum parasitaemia of 10 trypanosomes per ml of blood for most of these patients. Considering 4 liters of blood per patient, this amounts approximately to 2 million and 40,000 trypanosomes per patient in Bonon and Guinea, respectively. Combining these estimates with those from Table 1 yields values very different from N_e estimates (Table 3). With a difference around 10,000-fold in Bonon and 500-fold in Dubreka, values from Table 3 seem incompatible with moment-based estimates. From the F_{IS} analysis, according to De Meeûs *et al.* criteria of constantly strongly negative F_{IS} across strongly linked loci (22), full clonality can be assumed for *T. brucei gambiense* group 1 for the studied populations. According to Hellegren (29), microsatellite mutation rates mostly range between 10^{-3} and 10^{-4}. We use these two values for estimating clonal effective population sizes with equation 1 of *Materials and Methods*. The results are presented in Fig. 4 and fall completely out of the range of values estimated with Waples' method, described in Table 3 (maximum in Dubreka, with $u = 10^{-4}$, $N_e = 1,471$) (see Fig. 4). Indeed, with such huge population sizes, it is probable that a much greater diversity and a much higher F_{IS} would have been obtained. As can be seen from the [*SI Appendix*](), with 10,000 individuals and $u = 10^{-4}$, the expected $F_{IS} \approx -0.11$ and with $n = 2 \times 10^8$, $F_{IS} \approx -6.10^{-6}$.

Fig. 4 shows the results obtained with the trypanosome life cycle-based method for generation time (37 to 49 days, see *Materials and Methods*). With $u = 10^{-4}$, F_{IS}-based N_e reaches 297, 760, and 1,479 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 $u = 10^{-3}$. 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% CIs, so that the F_{IS}-based method is probably more accurate, as suggested in
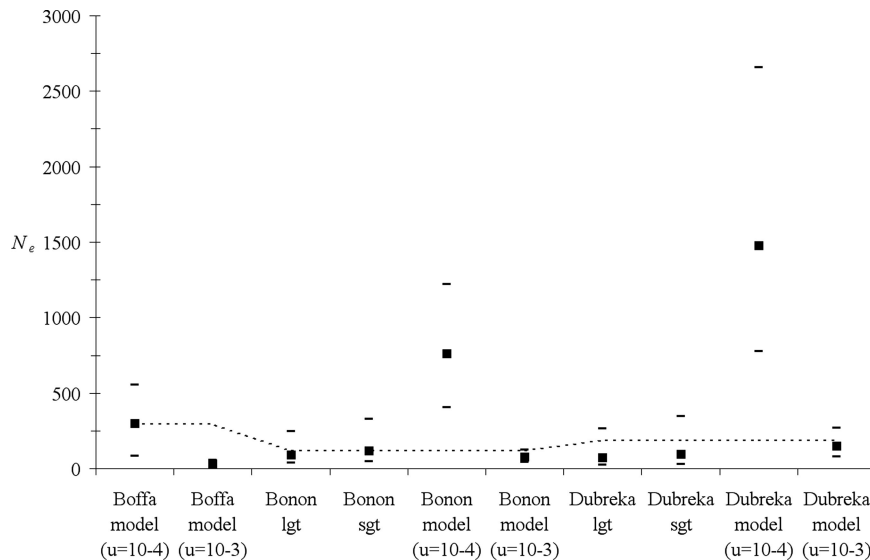
Fig. 4. Effective population size (N_e) obtained with the F_{IS}-based method (see *Materials and Methods* Eq. 1) ("model"), with $u = 10^{-3}$ and $u = 10^{-4}$, and with Waples' method from temporally spaced samples (with MLG as a single locus), using trypanosome's life cycle as the generation time with the shortest (sgt = 37 days) or largest (lgt = 49 days) generation times (see text). Black squares are the means with 95% CIs (*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 comes from a χ^2 distribution with *a* degrees of freedom (*a* is the number of alleles, in this case of different MLG's) (25). For the F_{IS}-based method, CIs correspond to those of F_{IS} obtained by bootstrap over loci.

general from theoretical analyses of fully clonal populations (30). From Fig. 4, the estimated numbers of infected patients seem almost perfectly to match all N_e estimates. This finding is unexpected if infected patients are to reflect *T. brucei gambiense* group 1 census sizes, which should be at least slightly over N_e. Moreover, a mutation rate of $u = 10^{-3}$–10^{-4} was used. Lower mutation rates will inflate N_e to values much higher than the census number of infected patients.

Discussion

According to the De Meeûs *et al.* (22), 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. The microsatellite loci used being located on different chromosomes (21), there is a strong statistical linkage disequilibrium between loci at a genome-wide scale in each subsample, in agreement with a purely clonal reproductive mode. More surprisingly, our results also indicate that within each country, *T. b. gambiense* group 1 populations are small and do not exchange many migrants. For example, in Guinea, where 2 sites were sampled in 2002, with equations 1 and 2 from our model (*Materials 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 $N_{ec} = 1,479$ and $m = 0.0008$ in Dubreka. Obviously, migration is weak. It is possible, however, that the sampling did not target the exact extent of actual *T. b. gambiense* group 1 subpopulations. In this case, some Wahlund effect may have altered our observations. In such a situation, our F_{IS} would have been slightly over-estimated, while the F_{ST} would have been slightly under-estimated. Nevertheless, a positive correlation of F_{IS} with the surface area of the sampling, or with the number of infected persons, was never observed; and the global positioning satellite (GPS) data available in Bonon 2002 and 2004 did not lead to lower estimates of F_{IS}, as expected if some Wahlund effect had affected our analyses (i.e., if each area was composed of several differentiated units). Interestingly, our results seem to tightly converge with those obtained for the closely related *T. brucei*

rhodesiense in Southeastern Africa, with 3 minisatellites (31) and human samples from 3 countries (Kenya, Uganda, and Zambia), for which a mean $F_{IS} = -0.50$ and $F_{ST} = 0.29$ (across countries) could be computed (reanalyzed in ref. 30). Unfortunately, this data set did not allow for more local analyses that could be compared to our Guinean samples.

At the scale of West Africa (between the Ivory Coast 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. In these two regions, HAT is transmitted by two fairly divergent Tsetse species (32). The differentiation between the *T. b. gambiense* from the two countries is maximal. Interestingly, this differentiation is correlated with different vector subspecies: *Glossina palpalis gambiensis* in Guinea and *G. palpalis palpalis* in the Ivory Coast (32, 33). These different pools of trypanosomes may be adapted to different vectors.

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. Effective clonal population sizes estimates are ex-tremely consistent across methods that are based on completely different assumptions and surprisingly fit epidemiological-survey conclusions, at least when mutation rate is assumed $u = 10^{-3}$. This latter point almost certainly represents a fortunate coincidence, because the temporal-based method probably leads to underestimated values. Indeed, the temporal-based method of effective population size was designed for sexual panmictic organisms with several independent loci, and MLG must be fairly sensitive to selective events (that hitchhike all of the genome in clonal organisms). Surprisingly, discrepancies ap-peared to be less pronounced than could have been foreseen. Consequently, and also because mutation rates may probably be lower than 10^{-3}, the observed local incidence of HAT appears to be lower than the corresponding effective clonal population size, probably because many hosts remain unnoticed (animal reservoirs or asymptomatic infected humans, see ref. 34). Un-derestimation of infection prevalence among exposed human populations represents an interpretation that meets the well-

Koffi *et al.*

known debate about aparasitemic seropositive subjects (35, 36). This phenomenon may be the result of failed parasite detection because of weak or fluctuating parasitaemia (37) or because of a phenomenon of control of infection by means of an appropriate immune system response (38). In these two latter cases, subjects who remain untreated may represent a potential parasite reservoir that could be responsible for the persistence of transmission and re-emergence of historical sleeping sickness foci. The data obtained in this study once more suggest that such asymptomatic infected humans may be of great epidemiological interest, unless the role of animal reservoirs can be safely dismissed (34, 39, 40). 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.

Materials and Methods

Trypanosome Isolates. Trypanosome isolates (one, and more rarely two per patient) were taken from 3 geographical zones and 4 sampling dates: in Guinea, Boffa 2002, Dubreka 1998, and Dubreka 2002; in the Ivory Coast, Bonon 2000, Bonon 2002, and Bonon 2004 (see Table S1). In the Ivory Coast, 3 different methods were used to isolate trypanosomes from HAT patients: kit for in vitro isolation of trypanosomes (KIVI), rodent inoculation (RI), and direct blood samples (BS) (see details in ref. 41). A total absence of differentiation between stocks isolated with different techniques (Table S4) leads to the conclusion that, in our study, isolation techniques do not significantly affect microsatellite genotypic frequencies. This factor was thus ignored in our analyses. In Bonon, the isolates were 17 in 2000, 14 in 2002, and 17 in 2004. In Guinea, the isolates were 15 in Dubreka 1998, 7 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 (42), leading to an estimate of about 120 infected persons (see Table 1). In Boffa and Dubreka, these values were extrapolated from medical survey results (27) taking into account evaluated population at risk (see Table 1), and lead to estimates of 187 and 295 infected persons in Dubreka and Boffa, respectively.

We studied 8 microsatellite loci: *M6c8*, *Mt3033* (43), *Trbpa1/2* (44), *Micbg1*, *Micbg5*, *Micbg6*, *Misatg4*, and *Misatg9* (21). Complete genotypes and MLG are given in Table S1. Because *Micbg6* did not vary across samples (all individuals displayed the same genotype), this locus was removed from the data set in further analyses. The strong variance in heterozygosity (as measured by F_{IS}) observed for *Trbpa1/2* across subsamples (Fig. S1), is more likely because of null alleles or selection than to rare events of sex (22, 45). *Trbpa1/2* is the only locus located in an expressed gene (46). It is thus better to remove this locus from further analyses. The absence of individuals with 3 or 4 alleles at any of the 6 remaining loci, and the constant level of heterozygosity across loci and samples, strongly support (if not prove) the diploid status of *T. brucei gambiense*, as already supported by genetic cross studies (47). The monophyly of *T. brucei gambiense* group 1, in particular as compared to *T. brucei gambiense* group 2, was already demonstrated (21). Combining our data with those from ref. 21 confirmed this point (Fig. S2).

Data Analysis. The most widely used parameters to infer population structure are the *F*-statistics (48; e.g., 49). Typically, these parameters are defined for 3 hierarchical levels. F_{IS} measures the identity (or homozygosity) of alleles within individuals within subpopulations 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 (30), 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 subpopulations, as compared to individuals from other subpopulations within the total population, or the total relative homozygosity caused by the Wahlund effect (50). It is thus a measure of differentiation between subpopulations 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 (51), with FSTAT version 2.9.3.2 (Goudet, 2002; updated from ref. 52), and their significant deviation from 0 was tested by randomizing alleles between individuals within subsamples and randomizing individuals among subsamples. Randomizations were set to 10,000 and implemented by FSTAT 2.9.3.2.

For testing linkage disequilibrium between pairs of loci, we used the multisample *G*-based (maximum likelihood ratio) test performed in FSTAT. This test considers each subsample as separated entities but combines the different statistics obtained across them to obtain a single *P*-value. We adjusted the *P*-values with the sequential Bonferroni correction by multiplying the smallest *P*-values by the number of remaining tests (see refs. 53–55).

In clonal diploids, genetic diversity can be very high (45), and this will tend to provide low estimates of F_{ST}. To get a more objective estimate of differentiation, we also computed the maximum possible value for the F_{ST} with Meirmans' method (23), where alleles are recoded so that no 2 subsamples share any allele in common but keep the same genetic diversity. In our case, where only paired F_{ST} were computed, this was made by increasing allele sizes of the second sample by 100. This method provides an estimate of the maximum possible value for F_{ST}, $F_{ST\ max}$, from which a standardized version of F_{ST}, $F_{ST}' = F_{ST}/F_{ST\ max}$ can be computed. Because correlation between loci might bias population differentiation measures and testing, we repeated F_{ST} analyses using the MLG of each individual as a unique haploid locus with as many alleles as defined by the MLGs. To get an encompassing picture of genotypic distribution across space, time, and sampling techniques, an NJTREE was computed by the MEGA 3.1 software (Kumar *et al.* 2005, updated from ref. 56). As recommended (e.g., 53, 57), the unrooted tree was built according to a Cavalli-Sforza and Edwards chord-distance matrix (24) computed with Genetix 4.05 (58).

Inferring Clonal Subpopulation Size. We used the model developed by Balloux *et al.* (45). Consider a subdivided monoecious population of diploid individuals with nonoverlapping generations. Individuals reproduce clonally with probability *c* and sexually with probability (1 − *c*). Self-fertilization occurs at a rate *s*. There are *n* subpopulations, or demes, each composed of *N* individuals. Migration between the subpopulations follows an island model (48), 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. In Appendix 1, it can be seen that in a two-population framework, which we assume being the case in both the Ivory Coast and Guinea, with total clonality, estimates of clonal population size *N*, and migration rate between the 2 populations can be obtained as:

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

and

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

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 (25) and implemented in NeEstimator v 1.3 (59). For this purpose, we only used the MLG data, which we rendered diploid by duplication of the allele of the single artificial locus obtained. MLGs were chosen because in clonal organisms all loci are linked and heterozygosity excess affects differentiation estimates (22). To estimate the number of trypanosome generations passed within the time windows (2 and 4 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 h (60), which yields 4.2 generations per day and thus 1,537 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 (61), while between 12 and 24 days are required for a newly infected Tsetse fly to become infectious for a vertebrate host (62, 63). This gives a generation time window of 37 to 49 days for trypanosomes, leading to 7 to 10 generations per year.

EVOLUTION

1. Cattand P (2001) L'épidémiologie de la trypanosomiase humaine africaine : Une histoire multifactorielle complexe. *Médecine Tropicale* 61:313–322.
2. Hoare CB (1972) *The Trypanosomes of Mammals: A zoological monograph.* Blackwell, Oxford.
3. Gibson WC (1995) The significance of genetic exchange in trypanosomes. *Parasitol Today* 11:465–468.
4. Gibson WC (2007) Resolution of the species problem in African trypanosomes. *Int J Parasitol* 37:829–838.
5. Tibayrenc M (1995) Population genetics of parasitic protozoa and other microorganisms. *Adv Parasitol* 36:47–115.
6. Tibayrenc M (1998) Genetic epidemiology of parasitic protozoa and other infectious agents: the need for an integrated approach. *Int J Parasitol* 28:85–104.
7. Tibayrenc M, Kjellberg F, Ayala FJ (1990) A clonal theory of parasitic protozoa: the population structures of *Entamoeba, Giardia, Leishmania, Naegleria, Plasmodium, Trichomonas,* and *Trypanosoma* and their medical and taxonomical consequences *Proc Natl Acad Sci USA* 87:2414–2418.
8. Gibson WC (1989) Analysis of a genetic cross between *Trypanosoma brucei rhodesiense* and *T. b. brucei. Parasitol Int* 99:391–402.
9. Jenni L, et al. (1986) Hybrid formation between African trypanosomes during cyclical transmission. *Nature (London)* 322:173–175.
10. MacLeod A, et al. (2005) Allelic segregation and independent assortment in *T. brucei* crosses: Proof that the genetic system is Mendelian and involves meiosis (vol 143, pg 12, 2005). *Mol Biochem Parasitol* 144:131–131.
11. MacLeod A, et al. (2005) Allelic segregation and independent assortment in *T. brucei* crosses: Proof that the genetic system is Mendelian and involves meiosis. *Mol Biochem Parasitol* 143:12–19.
12. Tait A (1980) Evidence for diploïdy and mating in trypanosomes. *Nature (London)* 287:536–538.
13. Criscione CD, Poulin R, Blouin MS (2005) Molecular ecology of parasites: elucidating ecological and microevolutionary processes. *Mol Ecol* 14:2247–2257.
14. Hartl DL, Clark AG (1989) *Principles of Population Genetics, Second Edition.* Sinauer Associates, Inc., Sunderland, MA.
15. Milgroom MG (1996) Recombination and the multilocus structure of fungal populations. *Annu Rev Phytopathol* 34:457–477.
16. Taylor JW, Geiser DM, Burt A, Koufopanou V (1999) The evolutionary biology and population genetics underlying fungal strain typing. *Clin Microbiol Rev* 12:126–146.
17. Tibayrenc M (2005) Bridging the gap between molecular epidemiologists and evolutionists. *Trends Microbiol* 13:575–580.
18. Tibayrenc M, et al. (1991) Are eukaryotic microorganisms clonal or sexual? A population genetics vantage. *Proc Natl Acad Sci USA* 88:5129–5133.
19. Jannin J (2005) Sleeping sickness a growing problem? *Br Med J (Clinical Research Ed.)* 331:1242.
20. Simarro P, Jannin J, Cattand P (2008) Eliminating human African trypanosomiasis: Where do we stand and what comes next? *Plos Medicine* 5(2):174–180.
21. Koffi M, et al. (2007) Genetic characterisation of *Trypanosoma brucei* ssp by microsatellite typing: new perspectives for the molecular epidemiology of human African trypanosomosis. *Infect Genet Evol* 7:675–684.
22. De Meeûs T, Lehmann L, Balloux F (2006) Molecular epidemiology of clonal diploids: A quick overview and a short DIY (do it yourself) notice. *Infect Genet Evol* 6:163–170.
23. Meirmans PG (2006) Using the AMOVA framework to estimate a standardized genetic differentiation measure. *Evolution* 60:2399–2402.
24. Cavalli-Sforza LL, Edwards AWF (1967) Phylogenetic analysis: model and estimation procedures. *Am J Hum Genet* 19:233–257.
25. Waples RS (1989) A generalized approach for estimating effective population size from temporal changes in allele frequency. *Genetics* 121:379–391.
26. Lumsden WHR, Kimber CD, Evans DA, Doig SJ (1979) *Trypanosoma brucei*: miniature anion-exchange centrifugation technique for detection of low parasitemias: adaptation for field use. *Trans R Soc of Tropl Med Hyg* 73:312–317.
27. Camara 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.
28. Anonymous (1998) Control and surveillance of African trypanosomiasis. Report of a WHO expert committee. *WHO Tech Rep Ser* 881(1):1–114.
29. Hellegren H (2000) Microsatellite mutations in the germline: implications for evolutionary inference. *Trends Genet* 16:551–558.
30. De Meeûs T, Balloux F (2005) F-statistics of clonal diploids structured in numerous demes. *Mol Ecol* 14:2695–2702.
31. MacLeod A, et al. (2000) Minisatellite marker analysis of *Trypanosoma brucei*: Reconciliation of clonal, panmictic, and epidemic population genetic structures. *Proc Natl Acad Sci USA* 97:13442–13447.
32. Ravel S, et al. (2007) The tsetse fly *Glossina palpalis palpalis* is composed of several genetically differentiated small populations in the sleeping sickness focus of Bonon, Côte d'Ivoire. *Infect Genet Evol* 7:116–125.

33. Camara M, et al. (2006) Genetic and morphometric evidence for population isolation of *Glossina palpalis gambiensis* (Diptera: Glossinidae) on the Loos islands, Guinea. *J Med Entomol* 43:853–860.
34. Jamonneau V, 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.
35. Garcia A, Courtin D, Solan P, Koffi M, Jamonnea V (2006) Human African trypanosomiasis: connecting parasite and host genetics. *Trends Parasitol* 22:405–409.
36. Koffi M, et al. (2006) Aparasitemic serological suspects in *Trypanosoma brucei gambiense* human African trypanosomiasis: A potential human reservoir of parasites? *Acta Tropica* 98:183–188.
37. Dukes P, et al. (1984) A field comparison of seven diagnostic techniques for human African trypanosomiasis in the Luangwa Valley, Zambia. *Tropenmedizin und Parasitologie* 35:141–147.
38. Garcia A, et al. (2000) Follow-up of Card Agglutination Trypanosomiasis Test (CATT) positive but apparently aparasitaemic individuals in Côte d'Ivoire : evidence for a complex and heterogeneous population. *Trop Med Int Health* 5:786–793.
39. Koffi M, et al. (2006) Aparasitemic serological suspects in *Trypanosoma brucei gambiense* human African trypanosomiasis: A potential human reservoir of parasites? *Acta Tropica* 98:183–188.
40. Njiokou F, et al. (2006) Wild fauna as a probable animal reservoir for Trypanosoma brucei gambiense in Cameroon. *Infect Genet Evol* 6:147–153.
41. Jamonneau V, N'Guessan P, N'Dri L, Simarro P, Truc P (2000) Exploration of the distribution of *Trypanosoma brucei* ssp. in West Africa, by multilocus enzyme electrophoresis. *Ann Trop Med Parasitol* 94:643–649.
42. Kaba D, et al. (2006) The impact of war on the evolution of sleeping sickness in west-central Cote d'Ivoire. *Trop Med Int Health* 11:136–143.
43. Biteau N, Bringaud F, Gibson W, Truc P, Baltz T (2000) Characterization of Trypanozoon isolates using a repeated coding sequence and microsatellite markers. *Mol Biochem Parasitol* 105:185–201.
44. Truc P, Ravel S, Jamonneau V, N'Guessan P, Cuny G (2002) Genetic variability within *Trypanosoma brucei gambiense*: evidence for the circulation of different genotypes in human African trypanosomiasis patients in Côte d'Ivoire. *Trans R Soc Trop Med Hyg* 96:52–55.
45. Balloux F, Lehmann L, De Meeûs T (2003) The population genetics of clonal and partially clonal diploids. *Genetics* 164:1635–1644.
46. Roditi I, Furger A, Ruepp S, Schurch N, Butikofer P (1998) Unravelling the procyclin coat of *Trypanosoma brucei. Mol Biochem Parasitol* 91:117–130.
47. MacLeod A, et al. (2005c) The genetic map and comparative analysis with the physical map of *Trypanosoma brucei. Nucleic Acids Res* 33:6688–6693.
48. Wright S (1951) The genetical structure of populations. *Ann Eugenics* 15:323–354.
49. Nagylaki T (1998) Fixation indices in subdivided populations. *Genetics* 148:1325–1332.
50. Wahlund S (1928) Zusammensetzung von populationen und korrelationserschinungen von standpunkt der vererbungslehre aus betrachtet. *Hereditas* 11:65–106.
51. Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. *Evolution* 38:1358–1370.
52. Goudet J (1995) FSTAT (Version 1.2): A computer program to calculate F-statistics. *J Hered* 86:485–486.
53. De Meeûs T, et al. (2007) Population genetics and molecular epidemiology or how to "débusquer la bête". *Infect Genet Evol* 7:308–332.
54. Holm S (1979) A simple sequentially rejective multiple test procedure. *Scand J Stat* 6:65–70.
55. Rice WR (1989) Analyzing tables of statistical tests. *Evolution* 43:223–225.
56. Kumar S, Tamura K, Nei M (2004) MEGA3: Integrated software for molecular evolutionary genetics analysis and sequence alignment. *Brief Bioinform* 5:150–163.
57. Takezaki N, Nei M (1996) Genetic distances and reconstruction of phylogenetic trees from microsatellite DNA. *Genetics* 144:389–399.
58. 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).
59. Peel D, Ovenden JR, Peel SL (2004) NeEstimator (Version 1.3): Software for estimating effective population size. Department of Primary Industries and Fisheries, Queensland Government, Deception Bay, Queensland.
60. 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. *Exp Parasitol* 109:188–194.
61. 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 Sci* 323:351–364.
62. Dale C, Welburn SC, Maudlin I, Milligan PJM (1995) The kinetics of maturation of trypanosome infections in tsetse. *Parasitology* 111:187–191.
63. 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.