The Population Structure of *Glossina palpalis gambiensis* from Island and Continental Locations in Coastal Guinea

Philippe Solano¹, Sophie Ravel², Jeremy Bouyer³, Mamadou Camara⁴, Moïse S. Kagbadouno⁴, Naomi Dyer⁵, Laetitia Gardes², Damien Herault², Martin J. Donnelly⁵, Thierry De Meeûs²,6

1 CIRDES/IRD UMR 177 IRD-CIRAD, Bobo-Dioulasso, Burkina Faso, 2 IRD, UMR 177 IRD-CIRAD, LRCT Campus International de Baillarguet, Montpellier, France, 3 CIRDES/CIRAD UMR 15 CIRAD-INRA, Bobo-Dioulasso, Burkina Faso, 4 PNLTHA, Ministère de la Santé, Conakry, Guinea, 5 Vector Group, Liverpool School of Tropical Medicine, Liverpool, United Kingdom, 6 GEMI, UMR 2724 IRD/CNRS, Montpellier, France

**Abstract**

*Background:* We undertook a population genetics analysis of the tsetse fly *Glossina palpalis gambiensis*, a major vector of sleeping sickness in West Africa, using microsatellite and mitochondrial DNA markers. Our aims were to estimate effective population size and the degree of isolation between coastal sites on the mainland of Guinea and Loos Islands. The sampling locations encompassed Dubrêka, the area with the highest Human African Trypanosomosis (HAT) prevalence in West Africa, mangrove and savannah sites on the mainland, and two islands, Fotoba and Kassa, within the Loos archipelago. These data are discussed with respect to the feasibility and sustainability of control strategies in those sites currently experiencing, or at risk of, sleeping sickness.

*Principal Findings:* We found very low migration rates between sites except between those sampled around the Dubrêka area that seems to contain a widely dispersed and panmictic population. In the Kassa island samples, various effective population size estimates all converged on surprisingly small values (10^<N_e><30) that suggest either a recent bottleneck, and/or other biological or ecological factors such as strong variance in the reproductive success of individuals.

*Conclusion/Significance:* Whatever their origin, the small effective population sizes suggest high levels of inbreeding in tsetse flies within the island samples in marked contrast to the large diffuse deme in Dubrêka zones. We discuss how these genetic results suggest that different tsetse control strategies should be applied on the mainland and islands.

**Introduction**

Mating pattern, population size and migration represent key factors determining the population genetic structure of organisms and shape the evolutionary of species [1–4]. Estimating these parameters is a major objective of population and conservation genetics [5–7]. Molecular markers are useful for estimating these parameters without the need for costly capture-mark-release-recapture (MRR) studies; particularly for organisms such as parasites and their vectors, where techniques such as MRR are difficult, impossible or unethical to apply [2,4]. Furthermore a detailed understanding of parasite and vector population dynamics is crucial for effective sustainable control [2,8–10].

The World Health Organisation recently launched a Human African Trypanosomosis (HAT, or sleeping sickness) elimination programme to counter the recent decline in case detection and treatment, notably in Central Africa [11–12]. However the situation in West Africa and the epidemiology of HAT is less well described. Guinea (especially the coastal area) is believed to be the country most affected by this disease [13]. Guinea has a long history of sleeping sickness, which was particularly prevalent in the years 1930–40 [14]. Current data show prevalences of between 2 and 5% in villages in the coastal mangrove area (Dubreka focus) [15]. This coastal area is composed of mangrove on the coastal margins and savannah inland. Offshore but in close proximity to the Dubreka focus lie the Loos Islands. These islands, physically separated from the mainland about 5000 years ago (D. Bazzo, pers. com.) are known to harbour tsetse flies (*Glossina palpalis gambiensis* (Diptera: Glossinidae), the main vector of HAT in West Africa and the Guinean National Control Programme against HAT recently launched an tsetse elimination programme on the archipelago. To facilitate the work of the elimination programme we used microsatellite and mitochondrial markers to address the following questions which are key to the successful control of tsetse: What is the effective population size in this tsetse species? What is the extent of genetic differentiation between mainland sites, between the islands and the mainland, and between the different islands of Loos archipelago? By answering these questions we hope to improve the design of control strategies in the region, especially with respect to designing and implementing area wide strategies.
Author Summary

Guinea is the country with the highest prevalence of sleeping sickness in West Africa, and we undertook a population genetics analysis there of the most dangerous tsetse fly species of West Africa, Glossina palpalis gambiensis. Our aims were to estimate effective population size and the degree of isolation between coastal sites on the mainland of Guinea (including Dubréka, a highly prevalent sleeping sickness focus) and Loos Islands in order to get the most possible accurate vision of feasibility and sustainability of anti-tsetse strategies of these sites. We found very low migration rates of tsetse between sites except between those situated in the Dubréka area, which seems to contain a widely distributed panmictic tsetse population (i.e. a population where mating occurs at random). Effective population sizes on Loos islands estimated with various techniques all converged to surprisingly small values. These values might be explained by a recent decrease in tsetse numbers on Kassa Island due to bauxite mining activities. But on the other sites, other explanations have to be found, including possible variance in reproductive success. Our genetic results suggest that different control strategies should be advised on the mainland (reduction in tsetse densities, no elimination) compared to the islands (total elimination feasible). This approach could be extended to many areas where vector control of Human and Animal Trypanosomoses is contemplated.

which must target genetically isolated populations if elimination is the objective [16]. Our results suggest that tsetse elimination is a feasible strategy on the Loos islands given both the genetic isolation between island and mainland populations and the small total surface to be controlled, but transmission reduction rather than elimination is more advisable for mainland tsetse populations.

Methods

Study area

The Loos islands are a small archipelago of five islands separated from the mainland of Guinea and the capital Conakry, by 4 km of sea. Three of the islands are inhabited, Kassa, Fotoba and Room in order of decreasing population size, with a total of around 7,000 inhabitants. On Kassa Island, two areas were sampled for tsetse, one in the north and one in the south. The biggest focus of HAT in Guinea, Dubréka, is on the mainland in a mangrove some 30 km distant from the Loos Islands. The area around Dubréka is characterised by coastal mangrove, with anthropic Guinean savannah, and permanently or temporarily inundated areas. Near the town of Dubréka (25,000 inhabitants), people live in villages of between 300 to 2,000 inhabitants. The main economic activities include fishing, salt extraction, and agriculture (palm and mango plantations, rice and food crops). In the Dubréka area, tsetse were sampled in 12 sites from two main areas: Touguioussory (five sites), in the mangrove habitat and accessible only by boat, and Magnokhoun (six sites) which is at the boundary between mangrove and savannah. A savannah area comprising a forest gallery bordering a water course was also sampled: Falessadé (one site), 30 km from the mangrove areas of Magnokhoun (see Figure 1).

Entomological surveys

Tsetse collections were made at each location using Vavoua traps [17]. Collecting cages were changed daily over a period of two to four days, and tsetse were counted and separated by sex. Three legs were removed from each fly and put in individual, labelled, dry eppendorf tubes.

All the continental samples and those from Fotoba Island were taken in 2005. Temporal samples were taken in Loos islands: 2005 and 2006 for Fotoba, 2006 and 2007 for Kassa.

Microsatellite markers

A total of 195 individuals were used for the genetic analyses at microsatellite loci: 7 males (M) and 15 females (F) in Kassa 2006, 14 M and 11 F in Fotoba 2005, 7F and 14M in Fotoba 2006, 18F and 12M in Kassa 2007, 17M and 15F in Magnokhoun 2005, 17M and 17F in Touguioussory 2005, and 11M and 20F in Falessadé 2005. Ten microsatellite loci were analysed: Gpg55.3 [18], pGp24, pGp 13, pGp11, pGp1 [19], C102, B104, B110 (kindly given by A.S. Robinson), GpGAG [20], and A10 kindly provided by G. Caccone. Locus Gpg55.3 has been reported to be located on the X chromosome [21], and given an absence of heterozygotes on a subsample of males (data not shown), B104, B110, pGp13 and ppg11 were also interpreted to be located on the X chromosome. For ease of reference we renamed X linked loci XGpg55.3, XB104, XB110, XPgp13 and XpGp11. Because loci A10 and pGp1 were unavailable before 2007, these loci were only used for 2007 sample (Kassa 2007) and thus only influenced local results (Linkage disequilibrium and FIS analyses).

In each tube containing three legs of the tsetse, 200 µl of 5% Chelex chelating resin was added [22–23]. After incubation at 56°C for one hour, DNA was denatured at 95°C for 30 min. The tubes were then centrifuged at 12,000 g for two min and frozen for later analysis.

The PCR reactions were carried out in a thermocycler (MJ Research, Cambridge, UK) in 10 µl final volume, using 1 µl of the supernatant from the extraction step. After PCR amplification, allele bands were resolved on a 4300 DNA Analysis System (LI-COR, Lincoln, NE) after migration on 96-lane reloadable 6.5% denaturing polyacrylamide gels. This method allows multiplexing of loci by the use of two infrared dyes (RDRye), separated by 100 nm (700 and 800 nm), and read by a two channel detection system that uses two separate lasers and detectors to eliminate errors due to fluorescence overlap. To determine the different allele sizes, a large panel of about 30 size markers was used. These size markers had been previously generated by cloning alleles from individual tsetse flies into pGEM-T Easy Vector (Promega Corporation, Madison, WI, USA). Three clones of each allele were sequenced using the T7 primer and the Big Dye Terminator Cycle Sequencing Ready Reaction Kit (PE Applied Biosystems, Foster City, CA, USA). Sequences were analysed on a PE Applied Biosystems 310 automatic DNA sequencer (PE Applied Biosystems) and the exact size of each cloned allele was determined.

PCR products from these cloned alleles were run in the same acrylamide gel as the samples, allowing the allele size of the samples to be determined accurately.

Data analysis

Linkage disequilibrium between pairs of loci was tested under Fstat 2.9.3.2 [24], updated from [25] by randomising loci (free recombination) with a G (log-likelihood ratio) based test allowing to get, for each pair of loci, a global test across sub-samples. For this analysis, the sub-sample unit was the smallest available one (e.g. the trap in Dubréka). Because this procedure involves multiple testing the P-values obtained were adjusted with a sequential Bonferroni procedure [26] (see [2] and references therein for detailed information). A binomial test was used to
check if the proportion of significant tests was significantly greater than expected based upon a 5% significance level (see [27]).

Wright’s $F$-statistics, the parameters most widely used to describe population structure [28], were initially defined for a three levels hierarchical population structure (individuals, sub-populations and total). In such a structure, three fixation indices or $F$-statistics can be defined: $F_{IS}$ is a measure of the inbreeding of individuals (hence I) resulting from the deviation from panmixia (random union of gametes) within each sub-population (hence S). $F_{ST}$ is a measure of inbreeding of individuals due to the structure of the population (non-random distribution of individuals among sub-populations); $F_{ST}$ also quantifies the differentiation between sub-populations in the total population (hence S and T). $F_{IT}$ is a measure of the inbreeding of individuals resulting both from non-random union of gametes within sub-populations and from population structure (deviation from panmixia of all individuals of the total population, hence I and T). These $F$-statistics are classically estimated by Weir and Cockerham’s unbiased estimators $f$ (for $F_{IS}$), $\theta$ (for $F_{ST}$) and $F$ (for $F_{IT}$) [29]. When appropriate, these statistics were estimated with Fstat 2.9.3.2.

Bilateral sex-biased dispersal tests were done in Fstat 2.9.3.2 with the $F_{ST}$ based test and the mean assignment index (the multilocus probability of belonging to the sampling site) corrected for population effects and its variance ($\bar{m}$ and $\sigma^2_{\bar{m}}$) [30], as recommended in [24]. We used 10000 permutations of individuals within samples [24] and applied the tests in continental samples considering each trap containing more than three flies or all traps from the same village as the sub-population unit (two analyses).

More than three levels (i.e. individuals, sub-populations and total) exist in the tsetse samples near Dubreka. Here, individuals were sampled using traps, in sites that are located in different “districts” (i.e. Magnokhoun, Touguissoury). HierFstat version 0.03–2 [31] is an analytical package written in R [32]. This package computes hierarchical $F$-statistics from any number of hierarchical levels. The significance of $F_{IT/S}$, the homozygosity due to subdivision into different traps within sites, was tested by randomising individuals among traps of the same site. The significance of $F_{S/D}$, which measures the relative homozygosity due to the geographical separation between sites within districts, was tested by randomizing traps (with all individuals it contains)
among the different sites in the same district. Finally, $F_{D/T}$ measures the relative homozygosity due to the subdivision into different districts, and was tested by randomising districts in the total sampling area. A gentle step by step description of how using HierFstat can be found in [2].

The significance of the $F$-statistics was tested by randomization (10000 permutations in each case). The significance of $F_{IS}$ was tested randomising alleles between individuals within sub-samples. The significance of $F_{ST}$ was tested by randomising individuals among sub-samples. These tests were performed with Fstat 2.9.3.2. For $F_{IS}$ the statistic was used directly the $F/\bar{F}_{IS}$ estimator. For $F_{ST}$ (and other differentiation tests), the statistic used was the maximum likelihood ratio $G$ [33].

(Differentiation between the northern and southern samples in Kassa were analysed by paired $F_{ST}$ (G based test) and tested in each year (2006, 2007) with Fstat 2.9.3.2. The two $F_{ST}$ were combined with an unweighted mean and the corresponding $P$-values with Fisher’s procedure [34] as described in [2].

Non random association of alleles within individuals ($F_{IS}>0$) may be due to null alleles. We used Micro-Checker 2.2.3 [35] to detect null alleles and estimate their frequency may be due to null alleles. We used Micro-Checker 2.2.3 [35] to detect null alleles and estimate their frequency (e.g. [2]). We thus used a “corrected” version of this statistics obtained and gaining some confidence on the parameters’ range.

The effective population size, usually noted $N_e$, is a measure of the rate at which a population loses genetic diversity by drift and roughly represent the number of adults that effectively contribute to the next generation (see [2] for a more precise definition and examples). Effective sub-population sizes could be estimated in each sample using various methods: the linkage disequilibrium methods of Bartley et al. (LDa) [41] and of Waples and Do (LDa,Co) [42], the temporal moment based method of Waples (1989) (Temporal) [43] for sites sampled at different times, joint estimation of migration and effective population size with the maximum likelihood (ML), the moment based (Moment) methods of Wang and Whitlock [44] and the method of Vitals and Couvet (2001) (Estim) [6,43–46]. Bartley’s and Waples’ methods were implemented with NeEstimator [47], Waples and Do’s method with LDNe [42]. Wang and Whitlock’s methods were implemented with MLNe v 1.1., and Vitals and Couvet’s method was implemented by Estim 1.2 [45]. For temporal based methods six generations were assumed to separate tsetse flies in one year interval. We also estimated $m$ (migration rate) from $N_e m$ with the formula $N_e m = (1 - F_{ST} / 3 F_{ST})$ that is appropriate for two populations and probably more appropriate here between the two Loos islands in 2006, between North and South in Kassa (2006 and 2007) and between Fotoba island and the mainland samples in 2005. We used $N_e$ from the Temporal method to extract $m$. Implementing all these methods that work under more or less different assumptions allowed the comparison of the values obtained and gaining some confidence on the parameters’ range.

MLNE dataset was obtained using CREATE 1.0 [48]. For LDa method, values obtained for alleles at least as frequent as 0.05 were chosen.

Signatures of bottleneck events were investigated by comparing the expected heterozygosity for a sample ($H_E$) with the heterozygosity that would be expected for a sample taken in a population at mutation/drift equilibrium with the same size and allele number ($H_0$): as allele number decreases faster than heterozygosity, bottlenecks are indicated by $H_E > H_0$ in subsequent generations [49]. This analysis was performed using BOTTLENECK software [50] under an IAM (infinite allele model), a SMM (stepwise mutation model) or a TPM (two phase model), in the latter case we assumed that 70% of mutations consist of one step and 30% consist of multistep change with a variance of 30 (default values).

Significance was assigned using one-tailed Wilcoxon tests [49]. Global $P$-values, overall Fotoba 2005-2006 and overall Kassa-North 2006-2007 samples were obtained with the Fisher procedure [34]. Given it only had three individuals, the Kassa-South 2006 sub-sample was excluded from these analyses. Given our sample sizes and number of loci, bottleneck detection is only possible if it occurred between $t_1 = 0.025 \times 2N_e$ and $t_2 = 2.5 \times 2N_e$ generations ago [49], where $N_e$ represents the post-bottleneck effective population size. We compared these generation times to those believed to have occurred on Kassa Island since the 1960’s (i.e. 276 generations ago) when an important bauxite mining activity is thought to have strongly altered ecological conditions (see http://www.nationsencyclopedia.com/Africa/Guinea.html). This provided a possible $N_e$ included between 55 and 5520 individuals.

Mitochondrial markers

A portion of the 5’ end of the mitochondrial gene cytochrome oxidase 1 was amplified using the primers CI J-2195 TTGATTTTTTGTCATCCAGAAGT [51] and CULR TGAAGCCTTTTATTGGCAACTAAGT. Double distilled water containing 10× PCR buffer (Bioline), dNTP 0.8 mM, primers 0.5 μM each, MgCl₂ 3 mM was incubated with 0.25units of BIOTAq DNA polymerase and approximately 0.5 ng of template DNA in 25 μl reactions. Temperature cycles were 5 min 95°C, 35 cycles of 93°C for 1 min, 55°C for 1 min and 72°C for 2 min, then 72°C for 7 min. PCR products were purified using the Bioline SURECLEAN reagent (BIO-37046) according to the manufacturer’s instructions, and sequenced using an ABI3730XL sequencing machine (Macrogen). Each template was sequenced bi-directionally. Sequence traces were checked using CodonCode Aligner (CodonCode Corporation), and aligned using the ClustalW algorithm implemented in MEGA version 4 with the following settings: gap opening penalty15, gap extension penalty 6.6, IUB weight matrix, transition weight 0.5, delay divergent cut-off 30 [32]. The PCR product size is 850 bp, but for analysis the alignment was trimmed to 723 bp of good quality sequence. The following statistics were calculated using DNAasp: FST_Seq was calculated according to equation 3 in [53] and is comparable to Weir and Cockerham’s $F_{ST}$ estimator for sequence data [29]. $H_{ST}$, an equivalent of Nei’s estimator of $F_{ST}$ ($G_{ST}$) was
calculated according to equation 2–4, and $K_{ST}$* according to equations 7–11 in [54] and is an equivalent of Nei's sequence statistic $\gamma_{ST}$ [55]. A permutation test, in which haplotypes or sequences were randomly assigned to the different localities 10000 times, was used to test the significance of $H_{ST}$ and $K_{ST}$* [54]. Mitochondrial analysis could only be undertaken with 10 individuals from Fotoba 2005, five individuals from Touguissoury and five individuals from Magnokhoun.

Results

The only tsetse species caught was G. p. gambiensis. Entomological surveys gave mean catches of flies per trap per day (FTD) of 10 in Kassa and 1 in Fotoba for Loos islands. On the mainland, mean FTDs were 7.5 in Magnokhoun, 5.5 in Touguissoury, and 11 in Falessadé.

No sex biased dispersal was found (all $P$-values>0.05) in 2005 samples from continental sites. Therefore in all further analyses data from females and males is combined.

Among the 36 tests of linkage disequilibrium between paired loci ( locus XB110 was excluded due to insufficient polymorphism), only two pairs were in significant linkage that did not stay significant after Bonferroni correction ($P_{binomial} = 0.05$).

HierFstat analyses gave no effect for district ($F_{D/T} = 0.005$, $P = 0.21$), site ($F_{S/D} = 0.016$, $P = 0.454$) or traps ($F_{T/S} = -0.006$, $P = 0.90$). Thus, individual tsetse flies from Touguissoury and Magnokhoun (Dubre´ka focus) were considered to belong to the same population for the following analyses. In Kassa, over 2006 and 2007 a substantial differentiation could be seen between the temporal analysis for which the weak differentiation observed in Kassa between 2006 and 2007 was no longer significant. There was no evidence for short allele dominance at this locus. Excluding the six loci with null alleles provided a much smaller deviation of heterozygote frequency from panmictic expectation ($F_{IS} = 0.04$, $P = 0.069$) (Figure 2).

Differentiation was significantly positive between each continental site and Loos islands (2005 samples, $P<0.0001$). In Table 2 are presented the paired $F_{ST}$ mean $\bar{F}$ between paired samples and paired $F_{ST} = F_{ST} / (1 - \bar{F})$. The highest levels of differentiation were found between Loos islands and all the other sites. Geographical differentiation was highly significant for each pair of samples and the smallest (although highly significant) value was observed between continental sub-samples (Falessadé and Dubrêka). Within the Loos islands, there was a high and significant differentiation between Kassa and Fotoba, the two main islands. It is noteworthy that the same analysis undertaken with the four loci without null alleles (XpGp13, XB104, C102 and GpCAG) provided very similar results (not shown), with the exception of the temporal analysis for which the weak differentiation observed in Kassa between 2006 and 2007 was no longer significant. Genotypes at microsatellite loci of all the individuals analysed can be seen in Table S1.

A fragment of the mitochondrial Cytochrome Oydase I gene was amplified by PCR and sequenced (see Table S2). COI sequences were deposited in Genbank (accession numbers FJ387505-FJ387524). The statistics of genetic differentiation based on COI are presented in Table 3. Within Dubrêka, as with nuclear microsatellite markers, Touguissoury 2005 and Magnokhoun 2005 showed an absence of differentiation. Differentiation between Fotoba and Dubrêka was strong and significant (Table 3) and comparable to that observed using microsatellites (Table 2).

Effective population size estimates with different methods are given in Table 4. These results were obtained for all loci, including those with null alleles. Using the four loci without null alleles or even the two autosomal loci with no deviation from Hardy-Weinberg (C102 and GpCAG) greatly widened confidence intervals were obtained after jackknifing over samples, except for overall loci obtained by bootstrap over loci, for loci pGp1 and A10 that were only available in Kassa 2007 samples (two values that provide the range) and for the mean $F_{IS}$ over the four loci where no null alleles were evidenced and for which the range is given by the minimum and maximum values observed. $P$-values obtained after testing that $F_{IS}$ is not significantly above 0 (10000 permutations) are presented between brackets.

doi:10.1371/journal.pntd.0000392.g002

Figure 2. $F_{IS}$ values for each locus and over all loci. Confidence intervals were obtained after jackknifing over samples, except for overall loci obtained by bootstrap over loci, for loci pGp1 and A10 that were only available in Kassa 2007 samples (two values that provide the range) and for the mean $F_{IS}$ over the four loci where no null alleles were evidenced and for which the range is given by the minimum and maximum values observed. $P$-values obtained after testing that $F_{IS}$ is not significantly above 0 (10000 permutations) are presented between brackets.

www.plosntds.org 5 March 2009 | Volume 3 | Issue 3 | e392
Table 1. Results obtained during Microchecker analyses.

| Loci with positive $F_{TS}$ | Observed blanks | Total sample size | Expected blank frequency |
|-----------------------------|-----------------|-------------------|--------------------------|
|                            |                 |                   |                          |
|                            |                 |                   | Brokfield 2              |
|                            |                 |                   | $\lambda$-value          |
|                            |                 |                   | Males                    |
|                            |                 |                   | $\lambda$-value          |
|                            |                 |                   |                           |
| pGp24                      | 54              | 192               | 0.315                    |
| A10                        | 3               | 30                | 0.149                    |
| pGp1                       | 1               | 30                | 0.032                    |
| X55.3                      | 9               | 109               | 0.045                    |
| Xpgp11                     | 19              | 109               | 0.137                    |
| XB110                      | 3               | 109               | 0.033                    |

The six loci that gave positive results are presented with the total number of genotyped individuals for each locus (blanks included). The sum of expected blanks over all samples divided by the sample size gave the expected frequency of blanks and the result of the exact binomial test comparing the observed proportion of blanks to the expected one ($H_0$: there are less blanks than expected with null alales). Expected frequencies were computed with Brookfield’s second method for all loci or directly with the proportion of blank males at X-linked loci as described in the Methods section. Significant test (in bold) means that not enough blank individuals were observed at that locus for null alleles to explain the observed $F_{TS}$.

Table 2. Differentiation between paired samples of tsetse flies from Guinea as measured by $F_{ST}$ (Weir and Cockerham estimator) and $F_{ST}^{'} = F_{ST} / (1 - \Pi_T)$, corrected for polymorphism with Nei’s genetic diversity averaged over the two compared samples.

| Sub-samples       | $F_{ST}$  | $\lambda$-value | $\Pi_T$ | $F_{ST}^{'}$ |
|-------------------|-----------|------------------|---------|-------------|
| Geographical      |           |                  |         |             |
| Falessadé and Dubréka 2005 | 0.0188    | 0.0001           | 0.6991  | 0.0625      |
| Falessadé and Fotoba 2005 | 0.0715    | 0.0001           | 0.6086  | 0.1827      |
| Dubréka and Fotoba 2005 | 0.0971    | 0.0001           | 0.6055  | 0.2461      |
| Fotoba and Kassa North 2006 | 0.0989    | 0.0014           | 0.4963  | 0.1963      |
| Fotoba and Kassa South 2006 | 0.1380    | 0.0158           | 0.4200  | 0.2379      |
| Kassa North and South (2006, 2007) | 0.0854    | 0.0119           | 0.5620  | 0.1949      |
| Temporal          |           |                  |         |             |
| Fotoba 2005 and 2006 | 0.0240    | 0.0441           | 0.4900  | 0.0471      |
| Kassa North 2006 and 2007 | 0.0242    | 0.1655           | 0.5612  | 0.2272      |
| Kassa South 2006 and 2007 | 0.1683    | 0.0271           | 0.4548  | 0.3087      |

Both geographical (contemporaneous sub-samples) and temporal (sympatric sub-samples) are provided and $P$-values are also given (all significant except between Kassa North 2006 and 2007). For Kassa North and South (2006, 2007) $F_{ST}$ was averaged over years and the corresponding $P$-value obtained with Fisher’s procedure.

Table 3. Statistics on comparisons of mitochondrial cytochrome oxidase 1 sequences (723 sites in alignment, no gaps) for 2005 samples of $G. palpalis gambiensis$.

| Population pair          | $F_{ST, Seq}$ | $H_{ST}$ | $K_{ST}$* |
|--------------------------|---------------|----------|-----------|
| Fotoba (n = 10)          | 0.148         | 0.144 (0.009) | 0.109 (0.029) |
| Touguissoury (n = 5)     | 0.043         | 0.114 (0.027) | 0.087 (0.072) |
| Magnokhoun (n = 5)       | -0.113        | 0.000 (1)   | -0.035 (0.829) |
| Fotoba (n = 10)          | 0.124         | 0.101 (0.008) | 0.084 (0.016) |
| Dubréka (n = 10)         |               |           |           |

$P$-values are given between brackets. Significant tests are in bold.

doi:10.1371/journal.pntd.0000392.t003

doi:10.1371/journal.pntd.0000392.t002

doi:10.1371/journal.pntd.0000392.t001

intervals but provided mean estimates of the same order of magnitude as with the complete data set (not shown). Table 3 gives migration rate ($m$) estimates for the three sites (Fotoba, Kassa North and Kassa South) for which it was possible. Except for $LD_{WD}$ and Estim for which most $N_e$ were infinite or undefined, the different methods gave consistent $N_e$ and $m$ estimates, particularly for Fotoba where the four available methods converged to very similar estimates. On the mainland, Dubréka seemed to harbour a large effective population of tsetse flies ($475 < N_e < 2016$) while Falessadé showed a notably small estimate ($25 < N_e < 63$). Loos islands displayed surprisingly small effective population sizes (Table 4, 2 < $N_e < 145$).

Bottleneck signatures were only detected with the IAM model in Kassa North sub-sample with a global $P$-value (Fisher’s procedure) of 0.022. Thus, bottleneck detection in Kassa North is suggestive of a post bottleneck effective population size range (55-5520) that overlaps with what was suggested by other methods, most particularly with temporal based methods (Table 4), but with of course a much higher upper bound.

Discussion

Both microsatellite and mitochondrial DNA markers revealed a strong barrier to gene flow between $G. p. gambiensis$ samples from the Loos islands and continental areas. Even within the Loos islands, the genetic differentiation was significant. For the first time
in tsetse flies within the palpalis group, we provide estimates of effective population sizes. In the Loos islands these estimates were found to be low despite high observed densities of tsetse.

Genetic differentiation between pairs of populations: implications for control

The high and significant genetic differentiation found between the Loos islands and the mainland combined with the small total surface area to be treated (around 15 sq km), has led the Guinean National Control Programme to launch a tsetse elimination campaign (M. CAMARA, pers. com.). We show here that within the Loos islands, the high and significant differentiation found between Fotoba and Kassa suggests a low number of migrants between islands (probably less than one per generation). This result, combined with morphometric data suggest that the elimination can be based on the implementation of an area wide approach using a sequential control strategy on each of the islands in turn [16,40]. On the mainland, all sites in Dubréka could be considered as within a single reproductive unit. These sites were significantly differentiated but not completely isolated, from Falessadé (Savannah site). This suggests that, contrarily to what was observed on Loos islands, tsetse elimination in this mainland area should not yet be chosen as the ideal strategy so far because exchanges of flies between these sites may still occur through the dense hydrographic network present. An elimination programme in the area would then require to create artificial barriers around the area to be treated (made of insecticide impregnated traps for instance, see [56]), which would prevent immigration into this area. Such a programme would require more detailed sampling. Alternatively, tsetse control with the participation of local communities to reduce (but not eliminate) tsetse densities may be advised since this has been shown to be technically feasible [57].

G. palpalis structure within populations

Over all sample sites and loci in our Guinean samples, the strong and significant FS found was reasonably explained by null alleles, it was thus not necessary to invoke a Wahlund effect, in contrast to results obtained in Burkina Faso for the same tsetse species [22,58] or for the closely related subspecies G. p. palpalis in Côte d’Ivoire [59]. In Guinea, this could be explained by the high rainfall (3000 mm /annum) and large numbers of river and stream habitats, combined with good host availability allowing good dispersal conditions and a less restricted distribution. The exception to this would be on Kassa, where a human settlement that led to habitat fragmentation may have caused a slight Wahlund effect (in this island mean $F_{IS} = 0.092$, $P = 0.058$), which may also explain why $N_e$ estimated with LD methods were often lower than those estimated with temporal based methods.

We found surprisingly low estimates of $N_e$. To our knowledge, such estimates were made only once in tsetse, for a savannah species belonging to the morsitans group in East Africa, G. saynertoni [60], which is phylogenetically and ecologically very far from G. palpalis (riverine fly of the palpalis group) [61]. Using mitochondrial markers, Marquez et al., 2006 found very low estimates of population size, and attributed it to a recent bottleneck. In Kassa Island, tsetse seemed structured in fairly isolated populations that may explain why extremely small $N_e$ were found. However, Kassa Island is the place where tsetse densities were the highest (mean of 10 flies / trap / day, up to 100 in some traps). In North Kassa, the possible signature for a Bottleneck that led to habitat fragmentation may have caused a slight Wahlund effect (in this island mean $F_{IS} = 0.092$, $P = 0.058$), which may also explain why $N_e$ estimated with LD methods were often lower than those estimated with temporal based methods.

### Table 4. Effective population size ($N_e$) of the different samples with 95% confidence intervals (CI) obtained with five different methods (when possible).

| Methods         | LD$_0$ | LD$_{w0}$ | Temporal | ML | Estim |
|-----------------|--------|-----------|----------|----|-------|
| Fotoba          | 2016 (313, $\infty$) | 475 (75, $\infty$) |          |    |       |
| Kassa North     | 25 (19, 35) | 63 (21, $\infty$) |          |    |       |
| Kassa South     | 145 (38, $\infty$) | $\approx$ (26, $\infty$) |          |    |       |
| Fotoba          | 35 (6, 383) | 53 (9, $\infty$) |          |    |       |
| Kassa North     | 90 (22, $\infty$) | $\approx$ (18, $\infty$) | 42 (17, 134) | 37 (18, 86) | $\approx$ (2, $\infty$) |
| Fotoba          | 11 (7, 18) | $\approx$ (32, $\infty$) |          |    |       |
| Kassa North     | 22 (14, 46) | $\approx$ (6, $\infty$) |          |    |       |
| Kassa North     | 17 (11, 32) | $\approx$ (6, $\infty$) | 35 (14, 107) | 34 (16, 105) | 0 (0, $\infty$) |
| Kassa South     | 2 (1, 6) | $\approx$ (6, $\infty$) |          |    |       |
| Kassa South     | 2 (2, 2) | $\approx$ (6, $\infty$) |          |    |       |
| Kassa South     | 2 (1, 4) | $\approx$ (6, $\infty$) | 11 (4, 46) | 11 (5, 35) | $\approx$ (0, $\infty$) |

For Fotoba and Kassa, sequential values obtained with LD methods were also averaged. Moment based method from Wang and Whitlock did not provide any $N_e$. $<\approx$ $\infty$.

doi:10.1371/journal.pntd.0000392.t004

### Table 5. Estimation of migration rates with three methods on the three sites where this was possible.

| Methods         | ML | Moment | Estim | $N_m/\overline{N_e}$ | $N_e/\overline{N_e}$Temporal |
|-----------------|----|--------|-------|----------------------|-----------------------------|
| Fotoba ($N_m=1.18$) | 0.024 (0.009, 0.056) | 0.020 | 0.031 | 0.032 |
| Kassa North ($N_m=1.24$) | 0.043 (0.014, 0.104) | 0.070 | 0.036 |
| Kassa South ($N_m=1.06$) | 0.172 (0.052, 0.551) | 0.430 | 0.096 |

$N_m$ estimates, based on two population model, were averaged over estimates from Table 2 and in m extracted using the Temporal estimate of $N_e$ from Table 4. Neither Moment or Estim methods provided confidence intervals.

doi:10.1371/journal.pntd.0000392.t005
Tsetse Population Structure in Coastal Guinea

Table 6. Estimation of the possible range for population sizes ($N_e$) as a function of the number of females sired by the most successful males (other males do not mate) ($n_{f\text{eff}}$) in order to estimate small effective population sizes found in Guinean tsetse flies.

| Sub-sample          | Order of magnitude of $N_e$ | $N_e$ if $n_{f\text{eff}} = 4$ | $n_{f\text{eff}} = 460$ |
|---------------------|-----------------------------|--------------------------------|------------------------|
| Dubréka             | 1000                        | 2467                           | 245333                 |
| Fales saddé         | 40                          | 99                             | 9813                   |
| Fotoba              | 60                          | 148                            | 14720                  |
| Kassa North         | 30                          | 74                             | 7360                   |
| Kassa South         | 10                          | 25                             | 2453                   |

Details for computations can be found in the text and Appendix S1. The Order of magnitude of $N_e$ for each sub-sample comes from Table 4. doi:10.1371/journal.pntd.0000392.t006

... (tsesse are viviparous) (data not shown). Whatever the cause of these small effective population sizes, they are suggestive of significant (very high in certain sites such as in Kassa Island) levels of inbreeding in $G. p. gambiensis$ populations from coastal Guinea.

These results suggest that further analyses should be conducted on the potential of $G. p. gambiensis$ to maintain genetic diversity at local and global scales in Guinea, in particular regarding interactions with the aetiological agent of HAT, Trypanosoma brucei gambiensis, the epidemiology of which is known to vary substantially through West Africa and in Guinea in particular [13,15]. We also hope that, within the context of the Pan African eliminations programmes that have been launched by the WHO [11–12] and the African Union PATTEC (Pan African Tsetse and Trypanosomosis Eradication Campaign, [63]), more studies will be conducted on tsetse population genetics. Indeed, control programmes are beginning to recognise the importance of such data in helping to choose specific control strategies.

Supporting Information

Alternative Language Abstract S1 Translation of the Abstract into French by Philippe Solano. Found at: doi:10.1371/journal.pntd.0000392.s001 (0.03 MB DOC)

Appendix S1 Found at: doi:10.1371/journal.pntd.0000392.s002 (0.03 MB DOC)

Table S1 Found at: doi:10.1371/journal.pntd.0000392.s003 (0.13 MB XLS)

Table S2 Found at: doi:10.1371/journal.pntd.0000392.s004 (0.03 MB DOC)

Acknowledgments

We thank Christine Chevillon and Franck Prugnolle for very useful advice regarding bottleneck detection and $N_e$ estimations. We are very grateful to the PNLTHA and the Ministry of Health of Guinea, as well as CIRDES for provision of excellent conditions of work. Thanks to Dr. Michael Gaunt and two anonymous referees for their help in improving the manuscript.

Author Contributions

Conceived and designed the experiments: PS JB ND MJD TDM. Performed the experiments: SR MC MSK LG DH. Analyzed the data: PS MC MSK LG DH MJD TDM. Contributed reagents/materials/analysis tools: SR JB MC MSK LG DH. Wrote the paper: PS SR JB ND MJD TDM.

References

1. Criscione CD, Blouin MS (2005) Effective sizes of macroparasite populations: a conceptual model. Trends Parasitol 21: 212–217.

2. De Meers T, McCoy KD, Prugnolle F, Chevillon C, Durand P, et al. (2007) Population genetics and molecular epidemiology or how to “debusquer la bête.” Infect Genet Evol 7: 308–332.

3. Koffi BB, De Meers T, Barrett N, Durand P, Amathan C, et al. (2006) Founder effects, inbreeding and effective sizes in the Southern cattle tick: the effect of transmission dynamics and implications for pest management. Mol Ecol 15: 4603–4611.

4. Prugnolle F, De Meers T, Durand P, Sire C, Théron A (2002) Sex-specific genetic structure in Coenagrion mercuriale: evolutionary and epidemiological implications. Mol Ecol 11: 1231–1238.

5. Statkin M (1987) Gene flow and the geographic structure of natural populations. Science 236: 787–792.

6. Vitalis R, Couvet D (2001c) Two-locus identity probabilities and identity disequilibrium in a partially selfing subdivided population. Genet Res 77: 67–81.

7. Watts PC, Roussel F, Saccheri IJ, et al. (2007) Compatible genetic and ecological estimates of dispersal rates in insect (Cyanagora marculata). Odonata: Zygoptera) populations: analysis of “neighbourhood size” using a more precise estimator. Mol Ecol 16: 737–751.

8. Milgroom MG (1996) Recombination and the multilocus structure of fungal populations. Annu Rev Phytopathol 34: 457–477.

9. Taylor JW, Genser DM, Burt A, Koufopanou V (1999) The evolutionary biology and population genetics underlying fungal strain typing. Clin Microbiol Rev 12: 120–146.

10. Thiayenc M (1999) Towards an integrated genetic epidemiology of parasitic protozoa and other pathogens. Annu Rev Genet 33: 449–477.
11. Jannin JG (2005) Sleeping sickness-a growing problem? Br Med J (Clin Res Ed) 331: 1242.

12. Simarro PP, Jannin JG, Cattand P (2008) Eliminating Human African Trypanosomiasis: where do we stand and what comes next? PLoS Med 5: e55. doi:10.1371/journal.pmed.0050053.

13. Courtin F, Jamonneau V, Duvallet G, Garcia A, Coulibaly C, et al. (2008) Sleeping sickness in West Africa (1906–2006): Changes in spatial repartition and lessons from the past. Trop Med Int Hlth 13: 1–11.

14. Brengues J, Challier A, Ouedraogo VK (1964) Contribution à la connaissance de l'épidémiologie de la trypanosomiase humaine: enquête entomologique dans les territoires kivisi et limitrophes (avril-mai 1964: Sierra-Leone, Lébien, Guinée) OCCGE, Boho-Doullasso, Burkina Faso.

15. Camara M, Kabba D, KagboDouam M, Sanon JR, Ouendez F, et al. (2005) Human African trypanosomiasis in the mangrove forest in Guinea: epidemiological and clinical features in two adjacent outbreak areas. Med Trop 65: 155–161.

16. Yvbreyen M, Robinson AS, Hendrichs J (2007) Area wide control of insect pests. Viena, Austria: Springer.

17. Laviaisier C, Grebaut P (1990) Recherches sur les pièges a glosesines (Diptera: Glossinidae). Mise au point d'un modèle économique: le piège "Vavouca". Trop Scand 4: 165–70.

18. Solomon P, Duvallet G, Dumas V, Cuisance D, Cuny G (1997) Microsatellite markers for genetic population studies in Glossina palpalis (Diptera: Glossinidae). Acta Trop 65: 175–180.

19. Lina C, Bouzouzi M, Cheng Q, Robinson AS, Akow S, Zheng L (2001) Microsatellite polymorphism in tsetse flies. J Med Entomol 38: 376–381.

20. Baker MD, Krafsur ES (2001) Identification and properties of microsatellite markers in tsetse flies Glossina morsitans sensu lat. (Diptera: Glossinidae). Mol Ecol Notes 1: 234–236.

21. Gooding RH, Solano P, Ravel S (2004) X-chromosome mapping experiments suggest occurrence of cryptic species in the tsetse fly Glossina morsitans westfalli. Mol Ecol Notes 1: 354–356.

22. Faviere L, Balloux F, Goudet J, Perrin N (1997) Female-biased dispersal in the monogamous mammal Crocidura russula population structure. Evolution 38: 1358–1370.

23. Fisher RA (1970) Statistical Methods for Research Workers. Edinburgh: 14th Edition, Sinauer Associates, Inc.

24. Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004) MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. Mol Ecol Notes 5: 535–538.

25. Brookfield JFY (1996) A simple new method for estimating null allele frequency from heterozygote deficiency. Mol Ecol 5: 453–455.

26. Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of genetic population studies in diploid populations. Mol Biol Evol 1: 354–356.

27. Hedrick PW (2005) A standardized genetic differentiation measure. Evolution 59: 1633–1638.

28. Fisher RA (1970) Statistical Methods for Research Workers. Edinburgh: 14th Edition, Sinauer Associates, Inc.

29. Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004) MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. Mol Ecol Notes 5: 535–538.

30. Brookfield JFY (1996) A simple new method for estimating null allele frequency from heterozygote deficiency. Mol Ecol 5: 453–455.

31. De Meede T, Dujardin JP, Ravel S, Dujardin JP, Morisod JP, et al. (2006) Genetic and morphometric evidence for population isolation of Glossina palpalis gambiensi (Diptera: Glossinidae) on the Loos islands, Guinea. J Med Entomol 43: 853–860.

32. Wildl MC (2003) Estimating effective population size and migration rates from genetic samples over space and time. Genetics 165: 429–446.

33. Vitalis R, Coutant V (2011) ESTIM 1.0: a computer program to infer population parameters from one- and two-locus gene identity probabilities. Mol Ecol Notes 1: 534–536.

34. Vitalis R, Coutant V (2011) Estimation of effective population size and migration rate from one- and two-locus identity measures. Genetics 155: 911–925.

35. Brookfield JFY (1996) A simple new method for estimating null allele frequency from heterozygote deficiency. Mol Ecol 5: 453–455.

36. Wang JL, Windlock MG (2003) Creating a tsetse population size and migration rates from genetic samples over space and time. Genetics 165: 429–446.

37. Waples RS, Do C (2008) LDNE: a program for estimating effective population size from data on linkage disequilibrium. Molecular Ecology Resources; in press.

38. Waples RS (1989) A generalized approach for estimating effective population size from temporary changes in allele frequency. Genetics 121: 379–391.

39. Wang JL, Windlock MG (2003) Creating a tsetse population size and migration rates from genetic samples over space and time. Genetics 165: 429–446.

40. Hall M, Solano P, Donnelly MJ (2008) Molecular phylogenetics of tsetse flies (Diptera: Glossinidae) based on mitochondrial (CO1, 16S, ND2) and nuclear barcoding markers. J Med Entomol 45: 179–189.

41. Hall M, Solano P, Donnelly MJ (2008) Molecular evolutionary genetics analysis (MEGA) software version 4.0. Mol Ecol Evol 24: 1596–1599.

42. Hudson RR, Slatkin M, Maddison WP (1992) Estimation of levels of gene flow from DNA sequence data. Genetics 132: 365–369.

43. Hudson RR, Boss DD, Kaplan NL (1992) A statistical test for detecting geographic subdivision. Mol Ecol Evol 13: 158–151.

44. Nei M (1987) Molecular Evolutionary Genetics. New York: Columbia University Press.

45. Nei M, Tajima F, Kanamori J (1975) Statistical analysis of DNA deletions. Genetics 81: 243–249.

46. Nei M (1987) Molecular Evolutionary Genetics. New York: Columbia University Press.

47. Nei M, Tajima F, Kanamori J (1975) Statistical analysis of DNA deletions. Genetics 81: 243–249.

48. Nei M (1987) Molecular Evolutionary Genetics. New York: Columbia University Press.

49. Cornuet JM, Luikart G (1996) Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. Genetics 144: 1401–1414.

50. Piry S, Luikart G, Cornuet JM (1999) BOTTLENECK: A computer program for detecting recent reductions in the effective population size using allele frequency data. J Hered 90: 502–503.

51. Simons C, Frits F, Bouchenbach A, et al. (1994) Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymers chain reaction primers. Ann Entomol Soc Am 87: 631–701.

52. Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol 24: 1596–1599.

53. Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol 24: 1596–1599.

54. Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol 24: 1596–1599.

55. Nei M (1987) Molecular Evolutionary Genetics. New York: Columbia University Press.

56. Cornuet JM, Luikart G (1996) Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. Genetics 144: 1401–1414.