Population genetics of *Glossina fuscipes fuscipes* from southern Chad

Sophie Ravel[1], Mahamat Hissène Mahamat[2], Adeline Ségard[1], Rafael Argiles-Herrero[3], Jérémy Bouyer[1,3,4]3, Jean-Baptiste Rayaisse[†5]4, Philippe Solano[1], Brahim Guihini Mollo[2], Mallaye ~~Pèka~~[6]Pèka[†5], Justin ~~Darnas~~[6]Darnas[5], Adrien Marie Gaston ~~Belem~~[7]Belem[6], Wilfrid ~~Yoni~~[5]Yoni[4], Camille ~~Noûs~~[8]Noûs[7], Thierry de Meeûs[1]*

[1] Intertryp, IRD, Cirad, Univ Montpellier, Montpellier, France.

[2] Institut de Recherche en Elevage pour le Développement (IRED), Ndjaména, Tchad.

[3] Insect Pest Control Laboratory, Joint Food and Agriculture Organization of the United Nations/International Atomic Energy Agency Program of Nuclear Techniques in Food and Agriculture, A-1400, Vienna, Austria.

~~[4] CIRAD, UMR ASTRE CIRAD-INRA « AnimalS, health, Territories, Risks and Ecosystems», Campus international de Baillarguet, 34398 Montpellier cedex 05, France.~~

[5]-[4] Centre International de Recherche Développement sur l'Elevage en zone Subhumide (Cirdes), Bobo-Dioulasso, Burkina Faso.

[6]-[5] Programme National de Lutte contre la THA (PNLTHA), Ndjaména, Tchad.

[7]-[6] Université Nazi Boni, Bobo-Dioulasso, Burkina Faso.

[8]-[7] Cogitamus laboratory, France, https://www.cogitamus.fr/.

E-mails: sophie.ravel@ird.fr; mahamatout@ymail.com; adeline.segard@ird.fr; J.Bouyer@iaea.org; philippe.solano@ird.fr; Mollomodougou@yahoo.fr; peka.mallaye@gmail.com; justindarnas@gmail.com; belemamg@hotmail.fr; wilfridyoni@gmail.com; camille.nous@cogitamus.fr

* Correspondence: thierry.demeeus@ird.fr (T. De Meeûs)

KeyWords: Tsetse flies, Dispersal, Trypanosomosis, Control.

Abstract

In Subsaharan Africa, t~~T~~setse flies (genus *Glossina*) ~~transmit deadly trypanosomes to human populations and domestic animals in sub-Saharan Africa:~~are vectors of trypanosomes causing Human African ~~Trypanosomosis~~ Trypanosomiasis (HAT) and Animal African Trypanosomosis (AAT). Some foci of HAT persist in Southern Chad, where a program of tsetse control was started against the local vector *Glossina fuscipes fuscipes* in the Mandoul focus in 2014, and in Maro in 2018. Flies were also sampled in 2018 in Timbéri and Dokoutou. We analyzed the population genetics of *G. fuscipes fuscipes* from the four tsetse-infested zones. The trapping samples were characterized by a strong female biased sex-ratio, except in Timbéri and Dokoutou that had high tsetse densities. Apparent density and effective population density appeared smaller in the main foci of Mandoul and Maro and the average dispersal distance (within the spatial scale of each zone) was as large as or larger than the total length of each respective zone. The genetic signature of a population bottleneck was found in the Mandoul and Timbé~~e~~ri area, suggesting a large ancient interconnected metapopulation that underwent genetic subdivision into small, isolated pockets due to adverse environmental conditions. The long-range dispersal and the existence of genetic outliers suggest a possibility of migration from remote sites such as the Central African Republic in the south (although the fly situation remains unknown there) and/or a genetic signature of recent exchanges. Due to likely isolation, an eradication strategy may be considered for sustainable HAT control in Mandoul focus~~ and control can also be advised in the Dokoutou-Timbéeri zone where HAT has not been reported yet but where AAT may cause problems on animal health~~. Another strategy will probably be required in Maro focus, which ~~shows~~ probably experiences much more exchanges with its neighbors.

Introduction

Tsetse flies (genus *Glossina*) transmit *Trypanosoma* spp. to humans and domestic animals in sub-Saharan Africa, causing the devastating diseases Human African Trypanosomosis (HAT) or sleeping sickness, and African Animal Trypanosomosis (AAT) or nagana. There is no vaccine available against these diseases, and treatments are difficult in humans and often compromised in animals due to the development of resistance against the available trypanocidal drugs (Bouyer et al., 2009). The WHO aims at interrupting transmission of *gambiense* HAT due to *Trypanosoma brucei gambiense* by 2030 (Büscher et al., 2018). Despite intensive disease surveillance programs and curative treatments, some HAT foci persist in different countries in Sub-Saharan Africa. In the southern part of Chad, medical surveillance and treatment has been supplemented with control efforts against the main HAT vector *Glossina fuscipes fuscipes* since 2014 in the Mandoul focus (Mahamat et al., 2017) and since 2018 in Maro (Ndung'u et al., 2020). The use of insecticide-impregnated tiny targets has suppressed the tsetse population significantly and resulted subsequently in a 63% decrease in HAT cases in the focus of Mandoul (Mahamat et al., 2017). Nevertheless, to understand and predict the sustainability of such vector control programs, it is necessary to study the biology of the vector populations, in particular the size and connectivity of the different subpopulations and dispersal capacities of the insects that drive reinvasion risks. This can be studied using polymorphic genetic markers as microsatellite loci and population genetics tools (De Meeûs et al., 2007). Such information can then be used to inform and develop the most appropriate tsetse population management strategy, i.e. local eradication can be considered if the tsetse target population is isolated~~.~~ (Solano et al., 2010)~~(e.g. Solano et al., 2010)~~, whereas other situations would spur undertaking alternative control strategies.

Given the humidity and microhabitat requirements for the survival of *G. f. fuscipes*, only the rivers with their riparian vegetation of the extreme South of Chad can sustain populations of this fly. The remaining part of the country has a Sahel vegetation and hence remains too dry for the survival of *G. f. fuscipes*. In this paper, we analyzed the population genetics of several *G. f. fuscipes* populations that are infesting the southern part of Chad. This included Mandoul and Maro, the main HAT foci of the country, but also Timbéeri and Dokoutou, where HAT cases were not reported. Nine microsatellite loci were used for a population genetics analysis of a total sample of 205 tsetse flies to estimate effective population density, dispersal distances and bottleneck signatures. The consequences of these results are discussed in the context of a potential tsetse eradication program in this area.

Material and Methods

Ethical statement

A prior informed consent (PIC) was obtained from the local focal point and a mutually agreed terms (MAT) form was written and approved between Chadian laboratories and French laboratories involved in the study for the use of the genetic diversity found in tsetse flies from Chad.

Origin of the samples

All flies were captured with biconical-Challier-Laveissière traps (Challier & Laveissière, 1973).

Sampling locations are described below and are presented in Figure 1. Details of traps deployed in different sites and dates, and numbers of captured flies are presented in the supplementary Figure S1. Detailed data with genotypes of individuals are available in the supplementary File S1.

Mandoul and Maro are two active HAT foci. These two zones had to be sampled and subjected to control at the same time, which required all logistic means. These zones were thus studied one at a time. Later, Dokoutou and Timbéri, which are not HAT foci, could then be sampled during the same season.

Commenté [SR1]: ???

Commenté [TDM2R1]: corrigé

Number of sampled flies (females, males and total) per location and zone, the cohort they belong to, taking two months as the generation time (De Meeûs et al., 2019) are presented in Table 1.

It is important to note that during the surveys of 2018 that resulted in the sampling of Timbéeri and Dokoutou, no other tsetse flies were caught between Mandoul and these localities despite trap deployment, meaning that the closest known geographic locality to the Mandoul population, infested with tsetse flies, was ~50 km away as the crow flies (J.B. Rayaisse, unpublished and Figure 1).

Figure 1: Location of sampling zones (Dokoutou, Timbéri Mandoul and Maro), and traps (red crosses) for *Glossina fuscipes fuscipes* in Southern Chad. Cohorts and numbers of flies trapped are indicated in Table 1. Mandoul and Maro are active sleeping sickness foci. Main water courses are indicated with dashed blue lines and area subjected to flooding are represented by blue areas. Forest galleries are symbolized in grey.

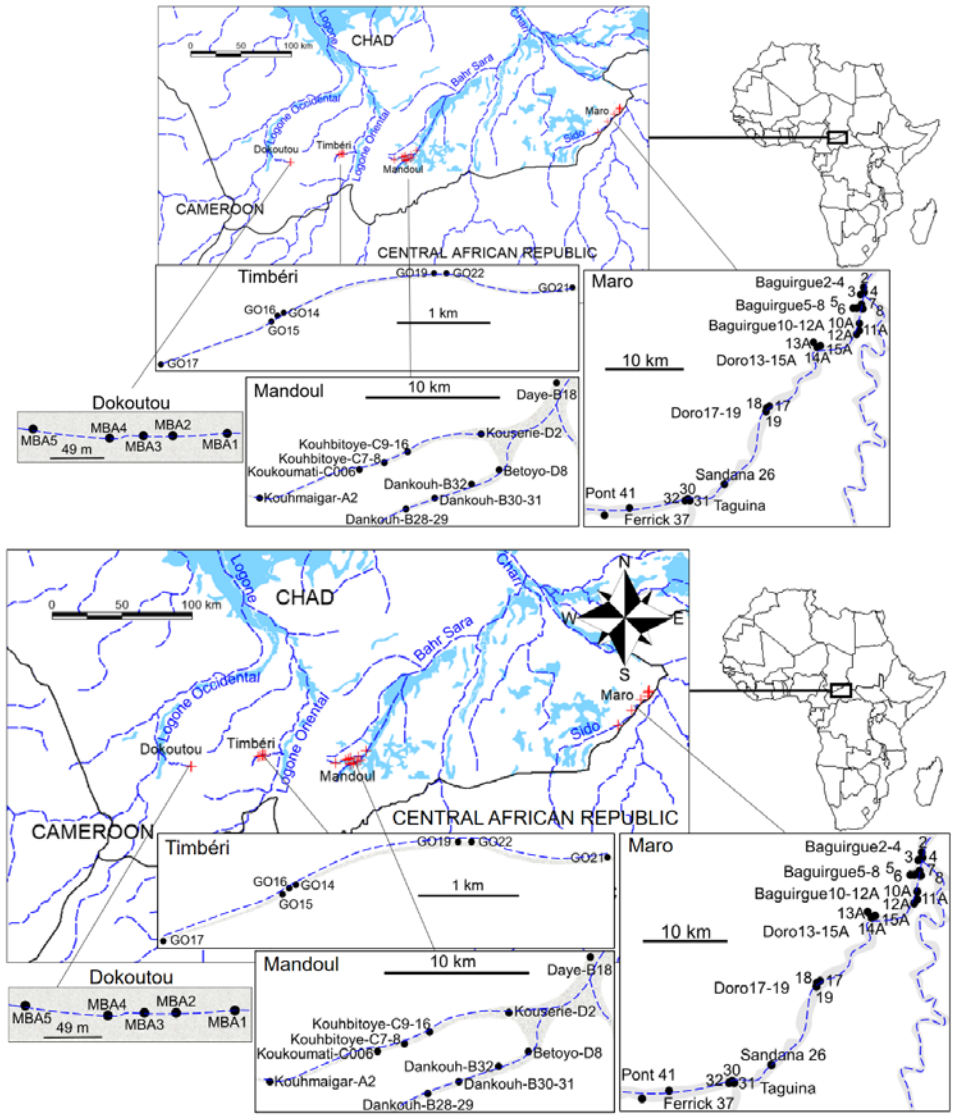


Table 1: ~~Focus~~Zone, cohort, number of females (N_f), males (N_m) and total number (N_t) of *Glossina fuscipes fuscipes* trapped in Southern Chad, and number of genotyped individuals (N_g). Cohorts were defined according to trapping dates, considering two months per *Glossina* generation (Mandoul November 2013 was cohort n°1; Maro April 2017, 42 months later i.e. 21 generations was n°22 and so on). The sex-ratio (SR=N_m/N_f) is also given, with exact *p*-value for significant deviation from even sex-ratio (two-sided exact binomial test).

~~Focus~~Zone	Cohort	N_f	N_m	N_t	N_g	SR	p-value
Mandoul	C1	98	50	148	96	0.5102	<0.0001
Maro	C22	49	18	67	63	0.3673	0.0002
Timbéri	C32	12	10	22	19	0.8333	0.8318
Dokoutou	C32	12	15	27	27	1.25	0.7011
Total		171	93	264	205	0.5439	<0.0001

The significant deviation of the sex-ratio from 1 (even sex-ratio) was tested with a two-sided exact binomial test with R (R-Core-Team, 2020) (command "binom.test"). The significant variations of the sex-ratio from one ~~site~~ zone to another were tested with Fisher's exact tests under the R-commander (rcmdr) package (Fox, 2005; Fox, 2007) for R. Densities of trapped flies (D_t) were computed for each ~~focus~~ zone as the total number of flies captured (N_t, as defined in Table 1), divided by the surface of the polygon defined by the traps with at least one fly (S_p). Except for Dokoutou, this surface was computed with Karney's algorithm (Karney, 2013) with the package geosphere (command areaPolygon) (Hijmans et al., 2019) for R (see appendix 1). For Dokoutou, traps were deployed in a very short portion (213 m long) of the forest gallery. The attractive cone of a trap is known to be much bigger than that, i.e. with a radius of 200 m (Bouyer et al., 2015). We thus considered that the surface of this site was defined by the length of the sampling zone (i.e. 213 m) plus twice the radius of the attractive cone (i.e. 2×200), hence 613 m, and a width corresponding to twice this radius, hence 400 m. This led to a surface of 0.2452 km², which approximatively corresponds to the surface occupied by the dense vegetation found in this area.

Surfaces of zones where then 32.11, 226.74, 0.2452 and 1.37 km² for Mandoul, Maro, Dokoutou and Timbéri respectively.

The correlation between densities of captured flies (D_c) and sex-ratio (*SR*) was tested with a two-sided Spearman's rank correlation test under rcmdr.

Microsatellite markers

A total of nine di-nucleotidic microsatellite loci were used (GFF3, GFF4, GFF8, GFF12, GFF16, GFF18, GFF21, GFF23, GFF27) with primers designed from a previously built microsatellite bank of *G. f. fuscipes* (Ravel et al., 2020). All the markers selected were autosomal (i.e. not on the X chromosome).

Genotyping

Legs from these flies were received in our lab in Montpellier. Three legs from each of G. *f. fuscipes* individuals were subjected to chelex treatment as previously described (Ravel et al., 2007) in order to obtain DNA for further microsatellite genotyping.

After PCR amplification of microsatellite loci, allele bands were routinely resolved on ABI 3500XL sequencer. This method allows multiplexing by the use of four different dyes. Allele calling was done using GeneMapper 4.1 software and the size standard GS600LIZ short run. A total of 205 individuals were genotyped (Table 1).

Structure of the data

Data were sorted according to the cohort (n°1, 22, and 32), considering two months per generation, as routinely described in previous publications (e.g. see File S1 in (De Meeûs et al., 2019)), traps ~~'49 -~~(49 traps in total), then according to the sub-site as defined in the Figure 1 (gathering traps that were less than 400 m apart) (see ~~also~~ Figure S1), then sites (12 sites: Baguirgue, Betoyo, Dankouh, Daye, Dokoutou, Doro, Kouhbitoye, Kouhmaigar, Koukoumati, Kouserie, Taguina and Timbéri), and zones (Mandoul, Maro, Timbéri and Doukoutou) (Figure 1). Raw data are available in the supplementary file S1.

Except for analyses undertaken with HierFstat and sex biased dispersal (see below), all genetic data were typed in the Create (Coombs et al., 2008) format and converted by this software into the needed formats.

Defining the relevant hierarchical levels of population structure

Different hierarchical levels of population structure could be considered in Chadian tsetse flies. In Mandoul, and Maro, we defined the Total sample, Sites, Subsites and Traps, with their corresponding *F*'s: F_{SiteT}, $F_{SubsiteSite}$, and $F_{TrapSubsite}$. For the Timbéri and Dokoutou, we could define the levels Total sample, Zone, Subsite and Trap, with the corresponding F_{ZoneT}, $F_{SubsiteZone}$ and $F_{TrapSubsite}$. To measure and test the significance of these hierarchical levels, we have used the algorithms implemented in HierFstat Package

(Goudet, 2005) for R. Hierarchical *F*-statistics estimate followed Yang's algorithm (Yang, 1998) and their significance was tested with 1000 randomizations of individuals between traps within subsites, of traps between subsites within sites or zones, and of subsites between sites or zones, to test the significant departure from 0 of $F_{TrapSubsite}$, $F_{SubsiteSite}$ or $F_{SubsiteZone}$, and F_{SiteT} or F_{ZoneT} respectively.

Because of the asynchrony of these samples, this needed to be undertaken in Mandoul (cohort 1), Maro (cohort 22), and Timbéri-Dokoutou (cohort 32) separately (three independent analyses).

More explanations and comments on hierarchical *F*-statistics can be found in (De Meeûs & Goudet, 2007).

Testing the quality of genetic markers and sampling

We first studied the statistical independence of loci with the *G*-based test for linkage disequilibrium (LD) across traps implemented in Fstat 2.9.4 (Goudet, 2003), updated from (Goudet, 1995), with 10000 randomizations. This procedure is indeed the most powerful way to combine tests across subsamples (De Meeûs et al., 2009). There are as many non-independent tests as there are locus pairs (here 36 pairs). The 36 tests series were adjusted with the Benjamini and Yekutieli (BY) false discovery rate (FDR) procedure for non-independent tests series (Benjamini & Yekutieli, 2001) with R.

Deviation from local panmixia, absence of subdivision and deviation from global panmixia were measured by Wright's F_{IS}, F_{ST} and F_{IT} respectively (Wright, 1965). Interested readers can found more extensive definitions in (De Meeûs et al., 2007). These were estimated with Weir and Cockerham's unbiased estimators (Weir & Cockerham, 1984) and their significance tested with 10000 randomizations of alleles between individuals within subsamples (for panmixia), of individuals between subsamples (for subdivision), and of alleles between individuals across the whole sample (global panmixia) with Fstat. For these tests, the statistics used were the F_{IS} estimator, *G* (Goudet et al., 1996) and F_{IT} estimator respectively. Default testing is unilateral (heterozygote deficit) for F_{IS} and F_{IT}. The bilateral *p*-value was obtained by doubling the *p*-value if it was below 0.5, or doubling 1-*p*-value otherwise. When needed, we compared F_{IS} and F_{IT} with a one-sided ($F_{IS} < F_{IT}$) (unless specified otherwise) Wilcoxon signed rank test for paired data with rcmdr. In that case, the pairing unit was the locus.

Jackknife over subsamples provided a standard error for *F*-statistics. This allowed computing 95% confidence intervals (95%CI) of *F*-statistics as described in (De Meeûs et al., 2007) to measure locus variation across subsamples. As it uses the student *t*

distribution (assuming normality, which is obviously not the case here), these 95%CI had only an illustrative purpose. The 95%CI of *F*-statistics were also obtained with 5000 bootstraps over loci, as described in (De Meeûs et al., 2007). This procedure assumes no particular distribution and thus have a statistical utility. We also computed standard error of F_{IS} and F_{ST} from jackknives over loci, StdrdErrFIS and StdrdErrFST to be used for null allele detection (seen Appendix 3).

In case of significant heterozygote deficit, we have looked for short allele dominance (SAD), stuttering, null alleles and Wahlund effects as described in previous studies (see Appendix 3).

LD tests, *F*-statistic estimates and testing, jackknives and bootstraps were undertaken with Fstat 2.9.4 (Goudet, 2003) updated from (Goudet, 1995).

~~*Defining the relevant hierarchical levels of population structure*~~

~~Different hierarchical levels of population structure could be considered in Chadian tsetse flies. In Mandoul, and Maro, we defined the Total sample, Sites, Subsites and Traps, with their corresponding *F*'s: F_{SiteT}, $F_{SubsiteSIte}$, and $F_{TrapSubsite}$. For the Timbéri and Dokoutou, we could define the levels Total sample, Zone, Subsite and Trap, with the corresponding F_{ZoneT}, $F_{SubsiteZone}$ and $F_{TrapSubsite}$. To measure and test the significance of those hierarchical levels, we have used the algorithms implemented in HierFstat Package (Goudet, 2005) for R. Hierarchical *F*-statistics estimate followed Yang's algorithm (Yang, 1998) and their significance was tested with 1000 randomizations of individuals between traps within subsites, of traps between subsites within sites or zones, and of subsites between sites or zones, to test the significant departure from 0 of $F_{TrapSubsite}$, $F_{SubsiteSIte}$ or $F_{SubsiteZone}$, and F_{SiteT} or F_{ZoneT} respectively.~~

~~Because of the asynchrony of these samples, this needed to be undertaken in Mandoul (cohort 1), Maro (cohort 22), and Timbéri-Dokoutou (cohort 32) separately (three independent analyses).~~

~~More explanations and comments on hierarchical *F*-statistics can be found in (De Meeûs & Goudet, 2007).~~

~~*Global level of subdivision*~~*Population genetics structure regarding reproduction*

Due to the temporal isolation between Mandoul, Maro and the Dokoutou-Timbéri complex, these three samples were studied in specific paragraphs.

In some instance, we compared F_{IS} and F_{IT} with a one-sided Wilcoxon signed rank test for paired data (the pairing object being the locus), under rcmdr, with H1 (alternative

hypothesis): $F_{IS}<F_{IT}$. We also used the same approach to compare F_{IS} within subsamples (traps or subsites) with F_{IS_pooled} after the pooling of all tsetse flies into a single sample.

After correction for stuttering (when appropriate), null alleles, or more exactly missing data (N_{Blanks}) explained most of F_{IS} (or F_{IT}) variations. We then used the intercept of the regression F_{IS} (or F_{IT}) ~ N_{Blanks} as an estimate of the basic F_{IS} of the population in absence (or quasi-absence) of null alleles.

Global subdivision

Because of the presence of null alleles, F_{ST} was estimated with the ENA correction with FreeNA (Chapuis & Estoup, 2007), for which we recoded missing data as homozygous for null alleles (coded 999, as recommended). We labelled this new estimate as F_{ST_FreeNA}. Confidence intervals of these estimates were computed after 5000 bootstraps over loci.

~~Because~~ For microsatellite loci, because of high mutation rates and excesses of polymorphism that results from it, the maximum possible value is lower than unity for F_{ST} (F_{ST}-max<1) ~~in microsatellite loci~~ (Hedrick, 2005b). To correct this estimate for excess of polymorphism, we can divide the actual estimator by the maximum possible value given the polymorphism observed within subsamples (Meirmans, 2006), or use G_{ST}"=[n(H_T-H_S]/[(nH_T-H_S)(1-H_S)] (Meirmans & Hedrick, 2011). Wang's criterion (Wang, 2015) allows determining which of the two approaches is more appropriate. If the correlation between Nei's G_{ST} and H_S is strongly negative, then F_{ST} based standardizations are more accurate, otherwise G_{ST}" should be used. This was tested with a one-sided Spearman's rank correlation test under rcmdr. We computed the standardized estimator of F_{ST} using Recodedata (Meirmans, 2006) to compute a maximum possible $F_{ST\_~~FreeNA~~\_max}$. We then obtained the standardized F_{ST_FreeNA}'= F_{ST_FreeNA}/$F_{ST\_~~FreeNA~~\_max}$. In that case, we obtained 95%CI with 5000 bootstraps over loci. These standardized subdivision measures could then be used to compute the effective number of immigrants within subpopulations as N_em=(1-F_{ST}')/(4F_{ST}'), where F_{ST}' stands for G_{ST}" or F_{ST_FreeNA}' (depending on Wang's criterion), or N_em=(1-F_{ST}')/(8F_{ST}'), in the special case of two subpopulations (i.e. Timbéri and Dokoutou ~~sites~~) (e.g. (De Meeûs, 2012), page 50).

Isolation by distance

Except for the Timbéri-Dokoutou sites for which captures were done the same year, isolation by distance was tested inside each zone separately. It was measured and tested with Rousset's model of regression in two dimensions $F_{ST\_R}=a+b\times\ln(D_{Geo})$ (Rousset,

Mis en forme : Indice

Mis en forme : Non Exposant/ Indice

Mis en forme : Police :Italique

1997). In this equation, $F_{ST\_R}=F_{ST}/(1-F_{ST})$ is Rousset's genetic distance between two subsamples (traps), *a* and *b* are the intercept and the slope of the regression respectively, and $\ln(D_{Geo})$ is the natural logarithm of the geographic distance between the two traps. Geographic distances were computed with the command distGeo of the package geosphere of R (see Appendix 1). The significance of the regression was tested by 5000 bootstraps over loci that provided a 95%CI of the slope. Because null alleles were present, we recoded all blank genotypes as homozygous profiles for allele 999 and used the ENA correction as recommended (Chapuis & Estoup, 2007) to compute $F_{ST\text{-}FreeNA}$. This was undertaken with FreeNA (Chapuis & Estoup, 2007). In case of significance, the neighborhood size and number of immigrants coming from neighbors and entering a subpopulation at each generation (in two dimensions) was computed as $Nb=4\pi D_e\overline{\sigma^2}=1/b$, and $N_em=1/(2\pi b)$ respectively (Rousset, 1997; Watts et al., 2007). In these formulae, D_e is the effective population density, $\overline{\sigma^2}$ is the average of squared axial distances between adults and their parents, and *b* is the slope of Rousset's regression model for isolation by distance (Rousset, 1997).

Code de champ modifié

Some subsamples harbored too few individuals that could not be taken into account in isolation by distance between traps or even subsites. We thus also undertook isolation by distance between individuals with Genepop 4.7.0 (Rousset, 2008), with the parameter $\hat{e}$ (Watts et al., 2007) for the genetic distance if not specified otherwise (i.e. when $Nb>50$), and 1000000 randomizations for the Mantel test. Note that in that case, no correction for null alleles was possible. In case of non-significance with previous procedures, we also undertook a Mantel test using the Cavalli-Sforza and Edwards' chord distance $D_{CSE\text{-}FreeNA}$ (Cavalli-Sforza & Edwards, 1967), computed with the INA correction for null alleles (Chapuis & Estoup, 2007) with FreeNA, and 10000 randomizations with the "Mantelize it" menu of Fstat. This genetic distance can indeed prove more powerful in case of weak signals (Séré et al., 2017). Mantel test in Fstat is two sided. We thus computed the one-sided *p*-value as half the *p*-value obtained for a positive correlation or (1-*p*-value)/2 otherwise.

Effective population sizes

For these computations, subsample units used were defined by the results obtained with HierFstat. In case of suspicion of a weak population subdivision, like in Mandoul and Maro foci, we also used the whole corresponding zone as a single unit. Effective population sizes were estimated with four different methods. The first method was the linkage disequilibrium (LD) method (Waples, 2006) adjusted for missing data (Peel et al.,

2013), and the second method was the coancestry method (Nomura, 2008). These two methods were both implemented with NeEstimator version 2.1 (Do et al., 2014). The third method was the within and between loci correlations (Vitalis & Couvet, 2001b) computed with Estim 1.2 (Vitalis, 2002) updated from (Vitalis & Couvet, 2001a). The fourth method was the heterozygote excess method from Balloux (Balloux, 2004). For the LD method, we retained only data with minimum allele frequency 0.05 as recommended in NeEstimator manual. We averaged N_e across usable values (excluding "infinite" results). We also retained minimum and maximum values across the four methods used. We finally computed the grand average and average minimum and maximum N_e across methods.

Effective population densities

All the four zones investigated are quite isolated from each other's: in time, by at least 10 generations, and in space, by at least 50 km for all, except between Dokoutou and Timbéri, which are spatially isolated from each other's by 50 km, but are contemporaneous.

Computing the effective population density in a given zone X (D_{e_X}) needs a knowledge of the relevant surface S_X, on which computing the total effective population size on that surface N_{e_X}, so that $D_{e\_X}=N_{e\_X}/S_X$.

We adapted the estimate of total effective population sizes to what was observed in each zone.

When no or weak population subdivision occurred, then each subsample was considered as a representative of the total zone and the average N_e was used as N_{e_X}. This is what we observed within all four zones.

When a significant subdivision occurred, as between Dokoutou and Timbéeri, several quatities were computed. ~~In Mandoul, the average effective population size over subsites was multiplied by the number of subsites to obtain the total $N_{e\text{-}T}$, we also used the estimates obtained over all the focus. In Maro, the average N_e across traps, or computed over all the focus was used to estimate $N_{e\text{-}T}$.~~ For Dokoutou and Timbéri, separately, we used the global N_e of each zone. The effective population densities were thus computed as $N_{e\text{-}T}/S$, where S is the surface of the zone, as computed above~~with the command "areaPolygon" of the package geosphere of R~~. No other population of tsetse flies were met between these two zones. ~~For~~ Consequently, for the effective population density across Dokoutou and Timbéri, we summed the two N_e obtained in each of the two zones to obtain $N_{e\text{-}DokoutouTimbériDT}$~~, and, the surface $S_{DokoutouTimbéri}$ was assumed to correspond to the disc defined by the distance between these two zones (D_{geo}) as the diameter of this~~

Mis en forme : Pas de paragraphes solidaires

disc, i.e. $S_{DokoutouTimbéri}=\pi\times(D_{geo}/2)^2$. When considering isolation by distance across traps of both zones, we computed this surface using the GPS coordinates of all traps of both zones with the package geosphere for R (command areaPolygon) (S_{DT_Area}). The effective population density was then obtained as $D_{e\text{-}DokoutouTimbéri}=N_{e\text{-}DokoutouTimbéri}/S_{DokoutouTimbéri\_X}$, where X stands for Disc or Area.

Isolation by distance

Except for the Timbéri-Dokoutou sites for which captures were done the same year, isolation by distance was tested inside each zone separately. It was measured and tested with Rousset's model of regression in two dimensions $F_{ST\_R}=a+b\times\ln(D_{Geo})$ (Rousset, 1997). In this equation, $F_{ST\_R}=F_{ST}/(1-F_{ST})$ is Rousset's genetic distance between two subsamples (traps), a and b are the intercept and the slope of the regression respectively, and $\ln(D_{Geo})$ is the natural logarithm of the geographic distance between the two traps. Geographic distances were computed with the command distGeo of the package geosphere of R (see Appendix 1). The significance of the regression was tested by 5000 bootstraps over loci that provided a 95%CI of the slope. Because null alleles were present, we recoded all blank genotypes as homozygous profiles for allele 999 and used the ENA correction as recommended (Chapuis & Estoup, 2007) to compute F_{ST_FreeNA}. This was undertaken with FreeNA (Chapuis & Estoup, 2007). In case of significance, the neighborhood size and number of immigrants coming from neighbors and entering a subpopulation at each generation (in two dimensions) was computed as $Nb=4\pi D_e\overline{\sigma^2}=1/b$, and $N_em=1/(2\pi b)$ respectively (Rousset, 1997; Watts et al., 2007). In these formulae, D_e is the effective population density, $\overline{\sigma^2}$ is the average of squared axial distances between adults and their parents, and b is the slope of Rousset's regression model for isolation by distance (Rousset, 1997).

Code de champ modifié

Some subsamples harbored too few individuals that could not be taken into account in isolation by distance between traps or even subsites. We thus also undertook isolation by distance between individuals with Genepop 4.7.0 (Rousset, 2008), with the parameter $\hat{e}$ (Watts et al., 2007) for the genetic distance if not specified otherwise (i.e. when $Nb>50$), and 1000000 randomizations for the Mantel test. Note that in that case, no correction for null alleles was possible. In case of non-significance with previous procedures, we also undertook a Mantel test using the Cavalli-Sforza and Edwards' chord distance D_{CSE_FreeNA} (Cavalli-Sforza & Edwards, 1967), computed with the INA correction for null alleles (Chapuis & Estoup, 2007) with FreeNA, and 10000 randomizations with the "Mantelize it" menu of Fstat. This genetic distance can indeed prove more powerful in case of weak

~~signals (Séré et al., 2017). Mantel test in Fstat is two-sided. We thus computed the one-sided *p* value as half the *p* value obtained for a positive correlation or (1-*p* value)/2 otherwise.~~

Dispersal distances

The average distance between adults and their parents was extracted with the equation (e.g. (De Meeûs et al., 2019)):

$$\delta \approx 2\sqrt{\frac{1}{4\pi b D_e}}$$

In this equation, *b* is the slope of Rousset's regression for isolation by distance and D_e is the average effective population density. This quantity is only accurate when dispersal distances follow a symmetrical distribution with a strong kurtosis. In any other case, like skewed distributions (right or left), or platykurtic distributions, δ will be slightly overestimated. Since there is also a lack of accuracy for D_e, δ corresponded more to an order of magnitude than a precise estimate of dispersal distance.

In the special case of Dokoutou-Timbéeri meta-zone, we had the opportunity to compute this distance using quasi-independent methods. The first method used the F_{ST}' based estimate of *m* (immigration rate) between the two zones, the average distance between these (D_{DT}) to get δ_m=m×D_{DT}. The second method used the slope b_{All} of isolation by distance between traps across the two zones and the S_{DT_Area} based estimate of $D_{e\text{-}DT}$ to obtain δ_{b_All} with the formula above. The third used the slope b_{Within} of isolation by distance within each zone and the corresponding surface defined above for each zone, and computed δ_{b_Within}. We also used individual, trap, and subsite based isolation by distance parameters to obtain various estimates of δ. This allowed checking for the consistency between the different values obtained. Individual-based isolation by distance does not correct for null alleles and thus is expected to produce overestimated and more variable slopes.

~~We also used the census of captured flies in traps to compute census population densities of captured flies (not the census population size) within each zone.~~

Factorial components analysis (FCA), DAPC and NJTree analyses

In order to visualize how the genetic information of the different individuals distribute relative to each other's, we have undertaken a factorial correspondence analysis (FCA) (She et al., 1987), where the values of inertia along each principal axis can be seen as F_{ST}

Mis en forme : Non Exposant/ Indice

Mis en forme : Police :Italique

combinations of different loci (Guinand, 1996). This analysis was undertaken with Genetix (Belkhir et al., 2004). Significance of the axes was assessed with the broken stick criterion (Frontier, 1976). We have also undertaken a DAPC analysis (Jombart et al., 2010), with the adegenet package (Jombart, 2008) for R. We finally computed a neighbor joining tree (NJTree) (Saitou & Nei, 1987) between sites, based on a Matrix of Cavalli-Sforza and Edwards chord distance (Cavalli-Sforza & Edwards, 1967), D_{CSE} as recommended (Takezaki & Nei, 1996). The matrix was computed with the INA correction of FreeNA to correct for null alleles, with missing data recoded as homozygotes for allele 999 as recommended (Chapuis & Estoup, 2007), and the NJTree built with MEGA 7 (Kumar et al., 2016). To test for the respective effects of geographic and temporal distances on genetic distances of this tree, we also undertook a partial Mantel test (Manly, 1997) with Fstat 2.9.4, based on the absolute regression coefficients and 10000 randomizations. In Fstat, *p*-values are two sided, but here we expected a positive correlation. One-sided *p*-values were thus obtained by halving *p*-values of positive correlations, and computing 1-(*p*-value)/2 otherwise.

Sex specific genetic structure

To test for the existence of a sex specific genetic structure, we used the biased dispersal menu of Fstat. We studied this in the four samples separately (namely in Mandoul C1, Maro C22, and Timbéri-Dokoutou C32). To gain in power and have enough males and females per subsample, we considered the subsites, as defined earlier, as subpopulation units. We used the corrected average assignment index *mAIc*, the variance of this index *vAIc* and Weir and Cockerham's unbiased estimate of F_{ST}, as recommended (Goudet et al., 2002; Prugnolle & De Meeûs, 2002) with 10000 permutations of gender status within subsamples. Significant male biased dispersal was seldom found in tsetse flies: once in *G. palpalis palpalis* in Cameroon (Mélachio et al., 2011), and twice for *G. tachinoides* in Burkina-Faso (Kone et al., 2011; Ravel et al., 2013). We thus used one-sided tests for male biased dispersal with the alternative hypotheses (subscript F and M designing female and male parameters respectively): $mAIc_F > mAIc_M$; $vAIc_F < vAIc_M$; and $F_{STF} > F_{STM}$. Here, correction for null alleles was not possible, and alleles needed to be recoded with two digits. For each parameter, there are three tests (the three cohorts: C1, C22, C32). For each parameter tested, we combined the *p*-values obtained across cohorts with the generalized binomial procedure (Teriokhin et al., 2007) computed with MultiTest v1.2 (De Meeûs et al., 2009) and following the rules described in the user guide: using $k'=k/2$ if $k>3$, and $k'=k$ otherwise, where k is the number of tests to be combined and k' is

the subset of smallest *p*-values to be considered. More explanations can be found elsewhere (De Meeûs, 2014).

Bottleneck detection

We used the algorithm developed by Cornuet and Luickart (Cornuet & Luikart, 1996) to detect the signature of a recent bottleneck in the different subsamples. We used the unilateral Wilcoxon test as recommended by the authors. As suggested ((De Meeûs, 2012), pages 104-105) , we studied IAM, TPM with default values (i.e. 70% of SMM and a variance of 30), and SMM models of mutation. A bottleneck signature likely occurred when the test is highly significant with IAM, and significant at least with TPM. Alternatively, a slightly significant bottleneck signature only observed with IAM more probably reflects small effective subpopulations sizes. We used Bottleneck v 1.2.02 (Piry et al., 1999) to undertake these tests in each cohort separately. The *p*-values obtained were combined across subsamples with the generalized binomial procedure, to get a global picture. We also used the Figure 3 in (Cornuet & Luikart, 1996) to extrapolate the probable post and pre bottleneck effective population sizes ($N_{e\text{-post}}$ and $N_{e\text{-pre}}$ respectively), using the probable $\tau = g/(2N_{e\text{-post}})$ and $\alpha = N_{e\text{-pre}}/N_{e\text{-post}}$, where g is the number of generations after the bottleneck event, and given the number of loci (here 9≈10), their genetic diversity (H_S) and sample size (N_{sample}) used.

Results

Sex-ratio within and between foci

There was a global and highly significant biased sex-ratio in favor of females (Table 1). This sex-ratio significantly varied between the different zones (*p*-value=0.0469). Densities of flies trapped in Mandoul, Maro, Dokoutou and Timbéri were 4.6, 0.3, 110.1, and 16, flies/km², respectively. Variation of effective population density across sites was strongly positively correlated with densities of capture (ρ=1), but marginally not significantly so (*p*-value=0.0833, two-sided). However, with four points, this *p*-value was in fact the minimum possible one.

Defining the relevant hierarchical levels of population structure

The results of this approach are presented in Table 2. Scripts and detailed results are presented in Appendix 2. It can be seen that population genetic structure did not occur at the same scale for the different sites/foci. In Mandoul, only the subsites displayed a significant effect. In Maro, only traps mattered. In Timbéeri and Dokoutou, the zone

mattered most, but not significantly so. Nevertheless, when only levels Trap and Zone were kept, $F_{ZoneTotal}$=0.0867 with *p*-value=0.002. Moreover, signals were quite small in Mandoul and Maro (Table 2). This will need to be further explored.

Table 2: Results of the hierarchical *F*-statistics with HierFstat of the different samples for *Glossina fuscipes fuscipes* from Chad. The effect of subsites was measured within each site in Mandoul and Maro and within each zone for Timbéri and Dokoutou. For each sample, most important level is in bold.

Effect	Sample	Mandoul	Maro	Timbéri-Dokoutou
Zone	F_{ZoneT}	NA	NA	**0.075**
	p-value			0.196
Sites	F_{SiteT}	0.000	-0.007	NA
	p-value	0.303	0.567	
Subsites	$F_{SubsiteSite/Zone}$	**0.018**	0.000	0.020
	p-value	0.025	0.656	0.660
Traps	$F_{TrapSubsite}$	-0.027	**0.005**	-0.020
	p-value	0.720	0.031	0.961

Testing the quality of genetic markers and sampling

Detailed analyses were quite fastidious and are presented in Appendix 3.

No signature of any linkage disequilibrium could be detected and all loci were considered as statistically independent in all zones.

No SAD signature could be found in any of the four zones. Null alleles were present in all samples at several loci and corrected accordingly. Stuttering was found at several loci in Maro, Timbéeri and Dokoutou and correction applied as described in Appendix 3.

There was no evidence of any Wahlund effect in any of the four zones.

Population genetics structure regarding reproduction of tsetse flies from Mandoul

Due to null alleles, the global F_{IS}=0.128 in 95%CI=[0.039, 0.243], was significantly different from 0 (*p*-value<0.0002). Population structure was weak, with a small and marginally not significant F_{ST}=0.005 in 95%CI=[-0.007, 0.016] (*p*-value=0.0722). Interestingly, F_{IT}=0.132 in 95%CI=[0.047, 0.244] was not significantly different from the F_{IS} (*p*-value=0.2129). It is thus possible that the whole focus behaves as a single pangamic population. Now, considering the whole focus as a single population, only two locus pairs

appeared in significant LD (p-values=0.0084 and 0.0344), none of which remained significant after BY adjustment (all p-values=1), and the F_{IS}=0.13 in 95%CI=[0.045, 0.238], was not significantly bigger than within subsites (p-value=0.3594) (no statistically detectable Wahlund effect). Again, missing data explained very well the positive F_{IS} (ρ=0.6836, p-value=0.0212, R^2=0.5733), with a residual $F_{IS\text{-}res}$=-0.0493.

Using F_{IS} or F_{IT} regressions against number of missing genotypes (Appendix 3), the intercept was used to estimate the residual values in absence of null alleles, which were F_{IS_res}=-0.0547 and F_{IT_res}=-0.0474.

Global subdivision in Mandoul

With FreeNA, the corrected subdivision measure was bigger than the uncorrected one: F_{ST_FreeNA}=0.0192 in 95%CI=[0.0084, 0.0295].

The correlation between G_{ST} and H_S was strongly negative (ρ=-0.7833, p-value=0.0086). Recodedata (Meirmans, 2006) provided $F_{ST\_FreeNA\text{-}max}$=0.2691 in 95%CI=[0.2086, 0.3410]. Consequently, F_{ST_FreeNA}'=0.0713 in 95%CI=[0.0405, 0.0866]. Some subdivision was observed, but given the correspondence between F_{IS} and F_{IT} it was at best weak.

Isolation by distance in Mandoul

Isolation by distance between subsites provided a very small and not significant slope b=0.0088 in 95%CI=[-0.0303, 0.0407]. The $\hat{e}$-based isolation by distance between individuals did not provide a different conclusion: b=0.0016 in 95%CI=[-0.0039, 0.0082] (Mantel test p-value=0.3178) ~~(Nb=607>>50)~~. When using D_{CSE}, the Mantel test provided a highly significant correlation (p-value=0.0003), with a very small coefficient of determination (R^2=0.0776). Isolation by distance thus occurred, but with a very weak signal. This would be in line for the existence of a nearly pangamic unit at the focus scale for Mandoul. Parameters' estimate from isolation by distance between subsites yielded a neighborhood Nb=114 individuals and an effective number of immigrants from neighbor sites N_em=18 individuals per generation. For isolation by distance between individuals, the neighborhood obtained was Nb=607 individuals and N_em=97 individuals per generation.

~~Now, considering the whole focus as a single population, only two locus pairs appeared in significant LD (p-values=0.0084 and 0.0344), none of which remained significant after BY adjustment (all p-values=1), and the F_{IS}=0.13 in 95%CI=[0.045, 0.238], was not significantly bigger than within subsites (p-value=0.3594) (no Wahlund effect). Again, missing data explained very well the positive F_{IS} (ρ=0.6836, p-value=0.0212,~~

Mis en forme : Police :Italique

Mis en forme : Police :Italique

~~R^2=0.5733), with a residual $F_{IS\ res}$=-0.0493. Hence, pooling all individuals into a single unit did not produce any Wahlund effect. This would be in line for the existence of a nearly pangamic unit at the focus scale for Mandoul.~~

Effective population size in Mandoul

Effective population sizes were computed within each subsite containing enough individuals (i.e. at least 7 individuals) or within the whole focus as a single population. Only two subsites provided usable values with the LD method (DankouhB30-31 and DankouhB28-29) and the coancestry method (DankouhB32 and DankouhB28-29), and only one with Estim (Betoyo). For Balloux' method, we used the residual F_{IS}-r computed with the missing genotype/F_{IS} regression. The average ~~effective subsite size~~ was N_e=50 in minimax=[9, 153] individuals. When we considered the whole focus as a single population, N_e=141 in minimax=[10, 272]. This is obviously not different from subsite-based estimate, though much more variable due to a lack of replicates. We thus kept within subsites averaged values.

Effective population densities in Mandoul

The ~~average surface of subsites was $S_{subsites}$=3304 m², and it~~surface of Mandoul was $S_{Mandoul}$=32 km² ~~for the whole focus~~. This led to ~~very different effective population densities: $D_{e\text{-}subsites}$=15111 in minimax=[2784, 46288] individuals/km², and~~ $D_{e\text{-}Mandoul}$=~~4.4~~1.6 in minimax=[0.3, ~~8.5~~47.6] individuals/km². ~~More than 15000 individuals/km² did not correspond to field observations and the whole zone obviously represents a more accurate scale.~~

Dispersal distances in Mandoul

Using $D_{e\text{-}Mandoul}$, we obtained two different effective dispersal distances: $\delta_{subsites}$=~~2870~~ 4823 m in minimax=[~~2067~~871, ~~10675~~11237] m/generation, for the subsite based isolation by distance regression; and $\delta_{individuals}$=~~6634~~ 11149 in minimax=[20144778, ~~24676~~25976] m/generation for the individual based isolation by distance regression. The two methods provided largely overlapping values. For information, the two most distant traps that captured at least one fly were 24 km distant from each other's in that focus.

Population genetics structure regarding reproduction of tsetse flies from Maro ~~after corrections for stuttering~~

Mis en forme : Police :Italique

Mis en forme : Police :Italique

Mis en forme : Police :Italique

After correction for stuttering at loci Gff3, 12, 16, 18 and Gff27 (Appendix 3), there was a non-significant and weak heterozygote excess within traps (F_{IS}=-0.001 in 95%CI=[-0.045, 0.036], *p*-value=0.9268). Null alleles affected weakly the data, with p_{null}=0.0585, and nine missing genotypes for Gff4 and much less for other loci.

Global subdivision in Maro

Subdivision was very small and not significant: F_{ST}=0.003 in 95%CI=[-0.01, 0.019] (*p*-value=0.135). This suggested again that tsetse flies from Maro almost behaved as a single population. Indeed, when pooling all individuals into one single unit, we observed only one significant LD locus pair (not significant after BY correction), and a F_{IS}=0.001 in 95%CI=[-0.035, 0.034], that was not significantly greater than the initial one (*p*-value=0.2852~~, one-sided test~~). Nevertheless, with FreeNA estimates, $F_{ST\text{-}FreeNA}$=0.0182 in 95%CI=[0.0017, 0.0419] was significantly above 0. The correlation between H_S and G_{ST} was not significantly negative (ρ=0.1333, *p*-value=0.646, one sided test), nevertheless, G_{ST}''=0.0479 (without 95%CI) was almost the same as the value obtained with Meirmans' method: $F_{ST\text{-}FreeNA}$'=0.0434 in 95%CI=[0.0069, 0.0716]. There was thus a possibility for a feeble subdivision signature with a global number of effective immigrants (using Meirmans estimates) N_em=5.06 on average and overall the focus.

Isolation by distance in Maro

Isolation by distance between traps, using F_{ST} estimates with the ENA correction computed with FreeNA, and after recoding missing data as null homozygotes, was not significant with the bootstrap 95%CI of the slope of Rousset's regression: *b*=0.0074 in 95%CI=[-0.00024, 0.0169]. However, the Mantel test based on geographic distances and $D_{CSE\text{-}FreeeNA}$ was highly significant (one sided *p*-value=0.0002). Finally, isolation by distance between individuals with Genepop (and no correction for null alleles), yielded a negative slope. So, at best, isolation by distance was weak and dispersal distances were probably substantial, and may be close or equal to the maximum length of the zone defined by Maro (32.4 km).

Effective population size of Maro

Effective population sizes computations did not output many values within traps: one with the LD method, two with the coancestry method, and five with Balloux's method (i.e. the five loci with a heterozygous excesses). It averaged N_{e_traps}=55 in minimax=[17, 118]. For the focus taken as a whole, only coancestry (one value) and Balloux's methods

Mis en forme : Police :Italique

Mis en forme : Police :Italique

(five values) provided usable values. The average was $N_{e\text{-focus}}$=28 in minimax=[20, 36], which was quite convergent with the previous values, confirming that the right scale ~~may be~~was the focus. We thus kept the trap-based estimate.

Effective population density in Maro

The area occupied by traps with at least one fly corresponded to a surface ~~S_{focus}~~S_{Maro}≈227 km². This yielded to very small effective population densities in the focus $D_{e\text{-focus}}$=0.24 in minimax=[0.08, 0.52] individuals per km² ~~for traps, and $D_{e\text{-focus}}$=0.12 for subsites and whole focus estimates, in minimax=[0.09, 0.16] individuals per km²~~.

~~Isolation by distance between traps, using F_{ST} estimates with the ENA correction computed with FreeNA, and after recoding missing data as null homozygotes, was not significant with the bootstrap 95%CI of the slope of Rousset's regression: b=0.0074 in 95%CI=[-0.00024, 0.0169]. However, the Mantel test based on geographic distances and $D_{CSE\text{-}FreeeNA}$ was highly significant (one sided p-value=0.0002). Finally, isolation by distance between individuals with Genepop (and no correction for null alleles), yielded a negative slope. So, at best, isolation by distance was weak and dispersal distances were probably substantial, and may be close or equal to the maximum length of the zone defined by Maro.~~

Dispersal distances in Maro

~~Using traps based estimate of density, t~~The average dispersal distance was δ_{traps}=13.7 km per generation, in minimax=[9.3, 24.6]~~. Using whole zone estimate of density provided δ_{Maro}=19.2 km per generation in minimax=[10.8, Infinity], where infinity probably means the maximum distance observed between two traps δ_{max}=32.5 km~~.

Population genetics structure regarding reproduction of tsetse flies from Dokoutou and Timbéri ~~and Dokoutou~~

After correction for stuttering at loci Gff8, 12 and Gff18 (Appendix 3), there was still a small but not significant heterozygote deficit (F_{IS}=0.031, in 95%CI=[-0.045, 0.144], p-value=0.2906) (panmictic populations), with some evidence of rare null alleles at some loci but with a complete disconnection with the too rare missing genotypes frequencies. We thus chose not to recode these missing genotypes for FreeNA computations.

Mis en forme : Police :Italique

Mis en forme : Police :Italique

Mis en forme : Retrait : Gauche : 0 cm, Suspendu : 1.25 cm

Global subdivision between Dokoutou and Timbéri

Subdivision between the two zones was highly significant (F_{ST}=0.08 in 95%CI=[0.055, 0.101], *p*-value<0.0001). Corrected F_{ST} was a little smaller (F_{ST_FreeNA}=0.0745 in 95%CI=[0.05, 0.0938]). The correlation between G_{ST} and H_S was positive. We thus used H_S=0.651, and H_T=0.679 to compute G_{ST}"=0.227. Interestingly, recoded $F_{ST\text{-}FreeNA\text{-}max}$=0.3274 provided the same value for $F_{ST\text{-}FreeNA}$'=0.2276 in 95%CI=[0.154, 0.2864] as for G_{ST}". We thus chose Meirmans' method, to keep 95%CIs. This allowed the computation of an effective number of immigrants N_em=0.4 in 95%CI=[0.3, 0.7] individuals per generation (with two subpopulations), exchanged between the two zones (e.g. ~ one individual every six months).

Isolation by distance within and between Dokoutou and Timbéri

Isolation by distance was explored first using traps as subsample units, with $F_{ST\text{-}FreeNA}$, but without recoding missing data, as these did not correspond to actual null homozygotes. With all traps of the two foci, isolation by distance was significant with a slope $b_{DT\text{-}traps}$=0.0144 in 95%CI=[0.001, 0.0208], a neighborhood size Nb=69 individuals in 95%CI=[48, 1031], and an effective number of immigrants from neighboring traps N_em=11 individuals per generation in 95%CI=[8, 164].

Within each site (separately), isolation by distance between traps provided a negative slope in Dokoutou for the average and the 95%CI~~, and f~~. For Timbéri, only the upper limit was positive ($b_{Timberi\text{-}Traps\text{-}u}$=0.033), with a corresponding lower Nb_l=30 individuals and N_em_l=5 individuals per generation. However, the low number of traps and the existence of traps with a single (unusable) fly led us to test isolation by distance between traps with a D_{CSE} based Mantel test. The result was significant (one sided *p*-value=0.0019). Isolation by distance between individuals, using parameter $\hat{e}$, gave similar results in Dokoutou (all slopes were negative), and Timbéri (only the upper limit was positive, $b_{Timberi\text{-}Ind\text{-}u}$=0.0044). For Timbéri, the corresponding lower Nb_l=226 individuals and N_em_l=36 individuals per generation. Using subsites, we observed a significant isolation by distance in Timbéeri with $b_{Timbéri\text{-}subsites}$=0.0095 in 95%CI=[0.0042, 0.0147], Nb=105 in 95%CI=[69, 238], N_em=17 in 95%CI=[11, 38].

Effective population size of Dokoutou and Timbéri

We could not get many usable values for N_e, especially for Dokoutou, which only provided infinite results, except with Balloux's method. Nevertheless, we used the rare cases where a lower limit of 95%CI was available ~~were used~~ as a lower limit~~s~~ to N_e in that

Mis en forme : Paragraphes solidaires
Mis en forme : Police :Italique
Mis en forme : Police :Italique
Mis en forme : Police :Italique

zone. These lower limits all suggested higher values in Dokoutou than in Timbéri (Table 3). We used these lowest values to obtain "minimum" averages of effective population densities. Doing so, actually considerably extended the range of possible N_e's in both zones.

Over the two zones, average N_e=38 in minimax=[6, 105]. Nevertheless, a~~A~~s the two zones are quite isolated from each other, the total (combined) effective population size can be assumed to correspond to the sum of the effective population sizes in ~~each~~Dokoutou and Timbéri. Hence $N_{e\text{-}Tot}$=76 in minimax=[12, 209].

Effective population densities in Dokoutou and Timbéri

Mis en forme : Police :Italique

As seen above, surfaces of these two zones were 0.2452 and 1.37 km² for Dokoutou and Timbéri respectively. Timbéri displayed an important effective population density D_e=20 individuals/km² in minimax=[0, 67], while Dokoutou appeared as extremely dense with more than 200 individuals/km² in minimax=[49, 478] (Table 3). ~~Over the two zones, N_e=38 in minimax=[6, 105]. Using the number of immigrants computed above, the immigration rate was m=0.0111 on average, and varied between 0.0029 and 0.1132 (minimum and maximum values). The average distance between traps of the two foci was $D_{Timberi\text{-}Dokoutou}$=50 km. We could thus estimate a rough proxy for the average dispersal distance ($m \times D_{Timberi\text{-}Dokoutou}$) δ_{TD}=557 m per generation, with a variation between 149 and 5662 meters, which looked much smaller than what was observed in the other two zones (Mandoul and Maro).~~

Table 3: Effective population sizes (N_e) of *Glossina fuscipes fuscipes* in Dokoutou and Timbéri (Chad), with different methods, and 95%CI (between brackets) when available, and averaged across methods; and minimum and maximum values observed. The surface (S) of Timbéri, in km², was computed with geosphere for R and as described in the Material and Methods section for Dokoutou. Averaged values of N_e were used to compute effective population densities (D_e) with N_e/S and minimum and maximum values observed across methods. The lowest value of 95% confidence intervals was used to compute averages when nothing else was available.

		Focus	
	Method	Dokoutou	Timbéeri
N_e	LD	Infinite [117.3, Infinite]	92 [23, Infinite]
	Coancestry	Infinite	13 [6, 22]
	Estim	Infinite [18, Infinite]	Infinite [0, Infinite]
	Balloux	12	4
	Average	49 [12, 117]	27 [0, 92]
S (km²)		0.2453	1.3712
D_e (individuals/km²)		200 [49, 478]	20 [0, 67]

~~Isolation by distance was explored first using traps as subsample units, with $F_{ST\text{-}FreeNA}$, but without recoding missing data. With all traps of the two foci, isolation by distance was significant with a slope $b_{TD\text{-}traps}$=0.0144 in 95%CI=[0.001, 0.0208], a neighborhood size Nb=69 individuals in 95%CI=[48, 1031], and an effective number of immigrants from neighboring traps N_em=11 individuals per generation in 95%CI=[8, 164].~~

~~As the two zones are quite isolated from each other, the total effective population size can be assumed to correspond to the sum of the effective population sizes in each. Hence $N_{e\text{-}Tot}$=76 in minimax=[12, 209]. Considering that the space between Timbéri and Dokoutou was empty (which seemed accurate), t~~

The total surface occupied by all traps of both foci was ~~S_{TD}~~S_{DT_Area}=3392662 m² ~~(computed with the areaPolygon function)~~. This led to an effective population density $D_{e\_TD\ trapsDT}$=22 individuals/km² in minimax=[4, 62]~~, and an estimate of dispersal $\delta_{TD\text{-}traps}$=993 m per generation in 95%CI=[826, 3824] and minimax=[498, 9580], which appeared very close to the $F_{ST\text{-}FreeNa}$' based estimation given above. Within each site (separately), isolation by distance between traps provided a negative slope in Dokoutou for the average and the 95%CI, and for Timbéri only the upper limit was positive ($b_{Timberi}$~~

Mis en forme : Police :Italique

~~$_b$=0.033), with a corresponding lower Nb_l=30 individuals and $N_e m_l$=5 individuals per generation. For Timbéri, D_e=20 individuals/km² in minimax=[0, 67], which would give $\delta_{Timberi\text{-}Traps\text{-}l}$=699 m/generation in minimax=[13, infinity], where infinity may correspond to the maximum distance between two traps (4876 m). This is also in the range estimated before, though this average represents only a 7th of the width of this (almost) panmictic site. However, the low number of traps and the existence of traps with a single (unusable) fly led us to test isolation by distance between traps with a D_{CSE}-based Mantel test. The result was significant (one-sided p-value=0.0019). Isolation by distance between individuals, using parameter $\hat{a}$, gave similar results in Dokoutou (all slopes were negative), and Timbéri (only the upper limit was positive, $b_{Timberi\text{-}a}$=0.0044). For Timbéri, the corresponding lower Nb_l=226 individuals and $N_e m_l$=36 individuals per generation and $\delta_{Timberi\ Traps\text{-}l}$=1909 m/generation in minimax=[5, infinity]. These values lied again into the window of values computed above. Isolation by distance between subsites was only possible in Timbéri. It gave a significant result with $b_{Timbéri\text{-}subsites}$=0.0095 in 95%CI=[0.0042, 0.0147], Nb=105 in 95%CI=[69, 238], $N_e m$=17 in 95%CI=[11, 38], and $\delta_{Timberi\ subsites}$=1304 m per generation in 95%CI=[1048, 1961] with a minimax=[568, infinity]. These values were not significantly different from those measured between traps across the two zones (as can be seen above). Hence, whatever the scale of study, F_{ST}-based between the two populations, isolation by distance over all or within Timbéri alone, or between individuals, dispersal distance was almost the same: $\delta_{average}$=1092 m/generation in minimax=[247, 5974].~~

Dispersal distances within and between Dokoutou and Timbéri

Using the number of immigrants between Dokoutou and Timbéri and averaged N_e computed above, the immigration rate was m=0.0111 on average, and varied between 0.0029 and 0.1132 (minimum and maximum values). The average distance between traps of the two foci was D_{DT}=50 km. We could thus estimate a rough proxy for the average dispersal distance ($m \times D_{DT}$) δ_m=557 m per generation, with a variation between 149 and 5662 meters, which looked much smaller than what was observed in the other two zones (Mandoul and Maro).

Between traps, over both zones, we computed an estimate of dispersal $\delta_{b\ All}$=993 m per generation in 95%CI=[826, 3824] and minimax=[498, 9580], which appeared very close to δ_m.

In Timbéri, still between traps, $\delta_{b\ Within}$=699 m/generation in minimax=[13, infinity], where infinity may correspond to the maximum distance between two traps in that zone

Mis en forme : Police :Italique

Mis en forme : Police :Italique

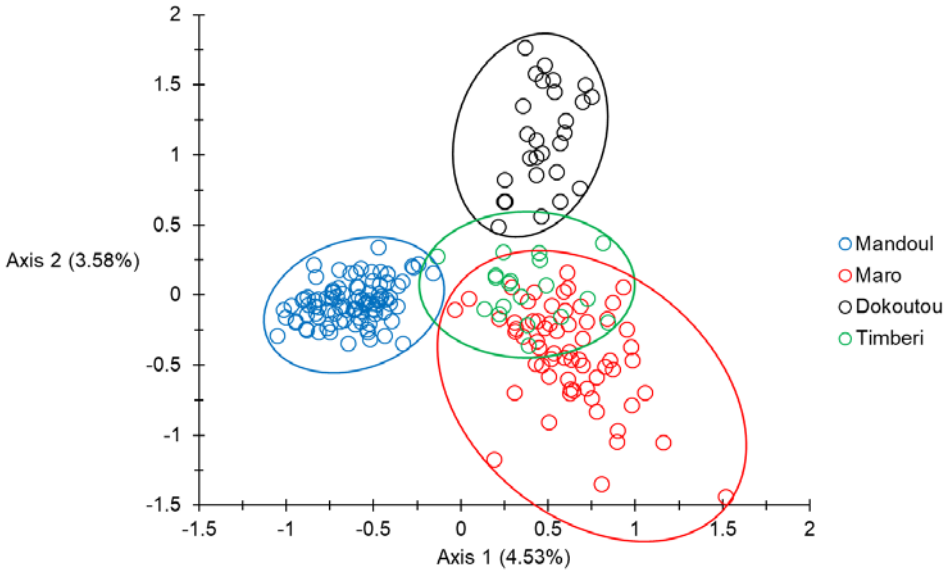
(i.e. 4876 m). This is also in the range estimated before. ~~and~~Still in Timbéri, but between individuals $\delta_{\text{Timberi-Traps-Individuals}}$=1909 m/generation in minimax=[5, infinity]. These values lied again into the window of values computed above. ~~I~~Finally, isolation by distance between subsites was only possible in Timbéri. and $\delta_{\text{Timberi-subsites}}$=1304 m per generation in 95%CI=[1048, 1961] with a minimax=[568, infinity].

All t~~T~~hese values were not significantly different from ~~those measured between traps across the two zones (as can be seen above)~~each others. Hence, whatever the scale of study, F_{ST} based between the two populations, isolation by distance over all or within Timbéri alone, ~~or~~ between subsites, traps or individuals, dispersal distance was almost the same: δ_{average}~~≈~~=1092 m/generation in minimax~~≈~~=[247, 5974].

Factorial components analysis (FCA), DAPC and NJTRee analyses

The results of the FCA analysis is presented in Figure 2. The two first axes were significant according to the broken stick criterion (highest expected percentages of inertia I_{E1}=3.77, and I_{E2}=3.09; observed ones I_{O1}=4.53 and I_{O2}=3.38 respectively). Axis 1 separated rather well Mandoul individuals from individuals from other samples, except for a few individuals that seemed close to Timbéri or Maro. The second axis separated Dokoutou, except for a few individuals that appeared close to individuals from Timbéri or Maro. Most other flies from Timbéri seemed to belong to the same pool defined by Maro individuals. Maro appeared very heterogeneous, which suggested substantial immigration from nearby (genetically close) or even remote (genetically distant) sites. Some outliers also suggested recruitment of flies from zones that were not sampled. It is difficult to clearly see the contribution of spatial and temporal distances to that picture. Spatially, Maro appeared as the most remote zone, while temporally, Mandoul is by far the most isolated one.

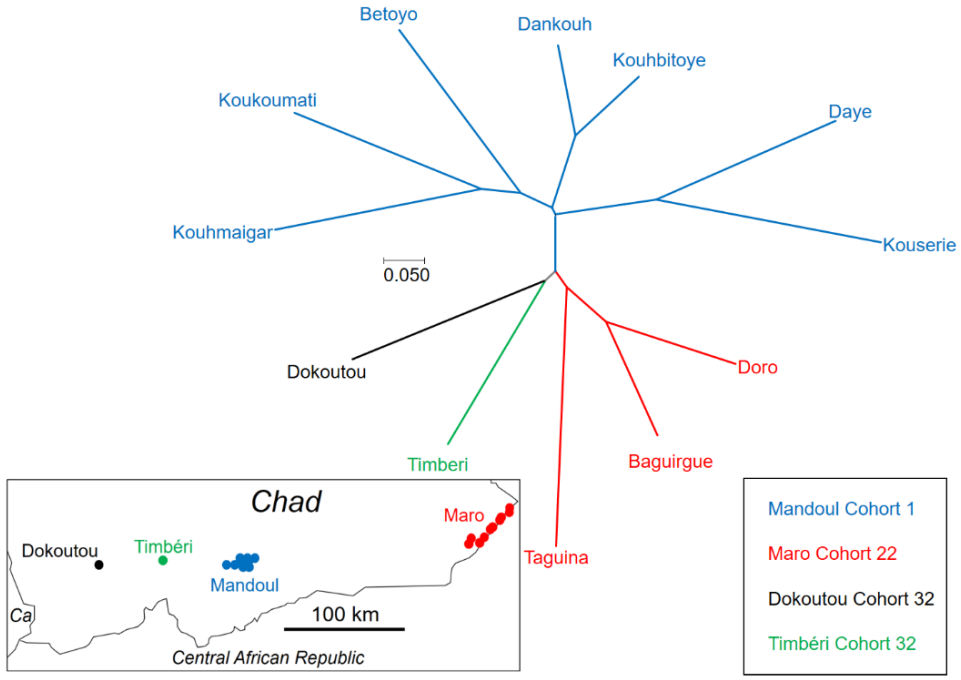
Figure 2: Presentation of the two dimensions projection of individuals of *Glossina fuscipes fuscipes* from different zones (with different colors) from Southern Chad according to the first two axes of a Factorial correspondence analysis. Percent of inertia are indicated. Both Axes 1 and 2 were significant. Mandoul flies belong to cohort 1, Maro to cohort 22 and Dokoutou and Timbéeri to cohort 32.



The DAPC analysis offered a very confused picture that was impossible to interpret biologically. This analysis is presented and discussed in Appendix 4.

The NJTree brought some more light (Figure 3) as temporal distances apparently affected more the distribution of branch lengths than geographic distances. Indeed, Maro and Dokoutou, which were the two most remote zones, were relatively close in the tree and only 10 generations apart, while Mandoul sites, which were geographically closer to Timbéri, but temporally very distant (31 generations), appeared as the most remote lineage of the tree. This was confirmed by the partial Mantel test that provided a higher partial correlation of D_{CSE} with temporal distances ($r_{Temporal}$=0.3175, p-value<0.0001) than with geographic distances ($r_{Geographic}$=0.2108, p-value=0.0041).

Figure 3: Neighbor-joining Tree based on Cavalli-Sforza and Edward's chord distance between the different sites of Southern Chad for *Glossina fuscipes fuscipes*. Zones and cohorts are indicated with the same colors as for Figure~~s~~ 2.



Sex specific genetic structure

Subsamples with only one gender or one individual were removed for these analyses to avoid error messages. Measures were contradictory depending on the statistic or the cohort used (Table 4). Globally, no test was significant (*p*-values>0.19), even if there was some tendencies toward male biased dispersal.

Table 4: Results of the sex specific genetic structure analyses undertaken in the different cohorts available, for the different statistics used. Significance (*p*-values) and their combination with the generalized binomial procedure (All) are also given. All tests were one-sided (alternative hypothesis H1: males disperse more). Values indicating the "most dispersive gender" are in bold. C1: Mandoul; C22: Maro; and C32: Dokoutou-Timbéri. Note that with three tests, the maximum possible combined *p*-value (All) was 0.125.

Parameter tested		C1	C22	C32	All
mAI_c	Females	0.1862	0.1100	**-0.3838**	-0.0282
	Males	**-0.4276**	**-0.2781**	0.3838	**-0.1073**
	p-value	0.1672	0.2925	0.8069	>0.1250
vAI_c	Females	5.6234	**6.5356**	**11.2584**	**7.8058**
	Males	**8.6431**	5.9403	5.5601	6.7145
	p-value	0.0932	0.4934	0.9363	>0.1250
F_{ST}	Females	0.0095	-0.0106	0.0792	0.0260
	Males	**0.0056**	**-0.0275**	**0.0655**	**0.0145**
	p-value	0.4971	0.3219	0.3081	0.1229

Bottleneck detection

For these analyses, following the previous results, we considered Mandoul, Maro, Dokoutou, and Timbéri as four different subpopulations. The results of these analyses are presented in Table 5. Globally, we found a rather convincing evidence of a bottleneck signature. Locally, only Mandoul and Timbéri presented a moderately and a strongly (respectively) significant signature of bottleneck.

Table 5: Results of the Bottleneck analysis for different samples, and for different models of mutations (IAM, TPM, and SMM). For each model of mutation, *p*-values were combined with the generalized binomial test (All), with the adapted optimal number of tests considered (*k*'=2), following rules defined for this procedure (see text).

Sample	IAM	TPM	SMM
Mandoul	0.0019	0.0644	0.1797
Maro	0.9356	0.9932	1
Dokoutou	0.1016	0.5898	0.999
Timbéri	0.001	0.0049	0.4102
All	<0.0001	0.0228	0.5425

Discussion

Although null alleles explained most heterozygote deficits, there was a tendency for stuttering at several loci. Stuttering was quite variable across the different zones: no evidence in Mandoul, five loci were probably affected in Maro, and three loci in Dokoutou and Timbéri. Fortunately, no SAD was evidenced in any of these samples. Stuttering and null alleles issues were taken care of before further analyses and inferences were made. Nevertheless, finding a way to avoid more efficiently amplification problems remains a progress that would be very welcome for the study of tsetse flies.

The strongly female biased sex-ratio observed in the least dense zones is difficult to understand. As can be seen in Table 6, densities of trapped flies were strongly correlated with effective population densities (ρ=1, *p*-value=0.0417, one-sided), which gives some reliability to density estimates and its correlation with *SR*. The data suggested that populations with very low densities contain much more females than males, whereas the sex ratio becomes more balanced in areas with higher population densities. It might also be that males from low-density populations respond less to biconical traps than females, a phenomenon that would tend to disappear in the sites with higher population densities (Table 6). Sites with high tsetse population densities may correlate with higher resource availability (more hosts) where females, with higher energy requirements, do not need to fly a lot to find a host for feeding. Alternatively, females need to spend more time flying in zones with scattered hosts on which to feed, and hence, would be more easily trapped, while male with smaller energy needs would fly less and not be so much exposed to trapping signals as females. Another non-exclusive hypothesis would relate to the density of suitable spots for larviposition. Pregnant females are known to be highly selective

before choosing a site where to larviposit (Gimonneau et al., 2021). In zones with higher densities of suitable larviposition spots, females do not need to search far away for larvipositing their larva, while in zones with less suitable larviposition spots, females would spend more time searching for suitable sites and hence, have a higher probability of being captured in biconical traps. Males can mate with virgin females that emerge from pupae in the larviposition sites soon after their imaginal molt, or when feeding on a host. This is however unlikely to influence trap catches, as tsetse responses to traps are feeding responses and not mating responses. If density negatively correlates with female dispersal distances, our observations may also be related to other disturbing results (De Meeûs et al., 2019). Although the above may seem highly speculative, it opens new routes for specific field and experimental investigations to better understand the density-dependent effects on the ecology of tsetse flies.

Table 6: Synthesis of numbers and densities of *Glossina fuscipes fuscipes* captured in traps (N_t and D_t), of effective population sizes and densities (N_e and D_e), and of Sex-ratio in the different zones of Southern Chad.

	Mandoul	Maro	Dokoutou	Timbéri
S (km²)	32	227	0.2	1.4
N_t	148	67	27	22
N_e	141	28	49	27
D_t (/km²)	4.6	0.3	49.1	16
D_e (/km²)	4.4	0.1	110.1	20
Sex ratio	0.51	0.37	1.25	0.83

Effective population densities in the Mandoul and Maro sites, which are active HAT foci, were similar to those at the lower limit found in the tsetse literature (De Meeûs et al., 2019) (Table 6). In those sites, the convergence between effective population densities and density of trapped flies was high, with $D_e<D_t$ for the smallest values, and the reverse for the highest ones. This may be due to the fact that the proportion of trapped flies as compared with the real population size decreased as the density increased. If this was true, the fly density in Dokoutou and Timbéri, the sites with the highest fly density and where $D_e>D_t$, should have maintained many tsetse flies after the first sampling campaign. Only a second future sampling campaign could test this prediction.

At the scale of each different site, dispersal distances were among the highest recorded for tsetse flies (De Meeûs et al., 2019), in particular for the Mandoul and Maro

910 sites, where an almost free movement across the whole range within each of these foci
911 was apparent, i.e. 24 km and 32 km, respectively. In Dokoutou, only 213 m wide, or
912 Timbéri, 5 km wide, effective dispersal distances were as large as, or larger than the size
913 of these areas. Dokoutou and Timbéri were separated by an average distance of 50 km,
914 but a genetic signature of a moderate exchange of immigrants was obvious between the
915 two sites: i.e. between one to two individuals every three generations (i.e. six months). We
916 observed a tight convergence of dispersal distances estimated from the F_{ST} computed
917 either between the two zones, or from isolation by distance between traps between the two
918 zones, or between individuals, traps or subsites in Timbéri alone. This brings confidence to
919 these estimates. In the literature, a maximum dispersal distance of 25 km in 24 days was
920 reported during a mark-release-recapture study with a wild female *Glossina tachinoides*
921 (Cuisance et al., 1985). Twenty-four days is less than half a generation. This distance was
922 covered in riparian forest bordering a river and not across rivers. Nevertheless, the riverine
923 tsetse species *Glossina palpalis gambiensis* has shown to be able to cross watersheds
924 between different river basins, even when the habitat was less favourable (Vreysen et al.,
925 2013). Although it might be a rare event, covering such a distance between Dokoutou and
926 Timbéri rivers in three generations should not be totally ruled out, especially during
927 favorable periods (rainy season), and using indirect trajectories, in particular via the
928 Southern and more favorable part of the country. Alternatively, we can use equation 9.13a
929 (p 502) of (Hedrick, 2005a) to explain the moderate genetic divergence observed between
930 Dokoutou and Timberi, in absence of any gene flow, i.e. g=-2N_eLN(1-G_{ST}"), where g is the
931 number of tsetse generations, N_e is the average effective population size across the two
932 zones, and G_{ST}" is the standardized F_{ST} estimate of Meirmans and Hedrick (Meirmans &
933 Hedrick, 2011). In that case, the two zones were completely isolated from each other only
934 3.3 years before sampling in minimax=[0.5, 9], for a two-months generation time (4.9 in
935 minimax=[0.8, 13] for three months generation time). Although this is theoretically
936 possible, such an abrupt and very recent environmental split is quite hard to envision. The
937 Mandoul control campaign, including the exploration of the surroundings, took place in
938 November 2013, i.e. five years before the sampling in Dokoutou-Timbéri, and there is no
939 evidence of an environmental continuity between Timbéri and Dokoutou that was followed
940 by an abrupt interruption. In addition, historical imagery of Google Earth Pro also does not
941 show any evidence of such an abrupt split in land cover between 2012 and 2018
942 (Supplementary File S2). Instead, the vegetation gap between the two zones was already
943 visible in 2013, and a very progressive and slow decline of "green areas" is obvious
944 between 2013 and 2018, with an apparent very slight acceleration in 2017. Moreover, if so,

it is hard to understand the convergence of dispersal distances estimates using different models, between Dokoutou and Timbéeri, or within Timbéeri alone. Rare gene exchanges (between one and two alleles every six months) between spots separated by 50 km of unsuitable landscapes as the crow flies, even if questionable, seems a reasonable interpretation of our population genetics results.

Very rare gene exchange may also hold for Mandoul and the CAR border (40 km), with several river courses in between. This was also suggested by the FCA analysis, where some individuals (or part of their genetic inheritance) may have been exchanged between the different zones. Such migration events would be extremely hard to observe, unless people deploy prohibitively large, intense and perennial trapping campaigns between the different zones investigated. On the other hand, the rarity of such an incident, renders the probability of reinvasion of eradicated zones very unlikely, since it would need the immigration of one fertilized female or one female and one male, at least. Trypanosome prevalence in humans was estimated as $P \approx 0.02$ before the control begun in Mandoul and Maro, and around 6% of tsetse flies were found positive for *T brucei sp* (Ibrahim et al., 2021). If we consider that trypanosome prevalence could reach values much lower than that as a result of medical and entomological campaigns, the probability of reinvasion with infected tsetse can reasonably be estimated as null.

The south border with Central African Republic (CAR) is located close to Maro and has not been investigated entomologically. It may represent many potential unexplored, and possibly tsetse rich environments and thus potential sources for reinvasion with tsetse flies. This may explain the great genetic heterogeneity of tsetse flies from Maro, and this focus will therefore need special attention.

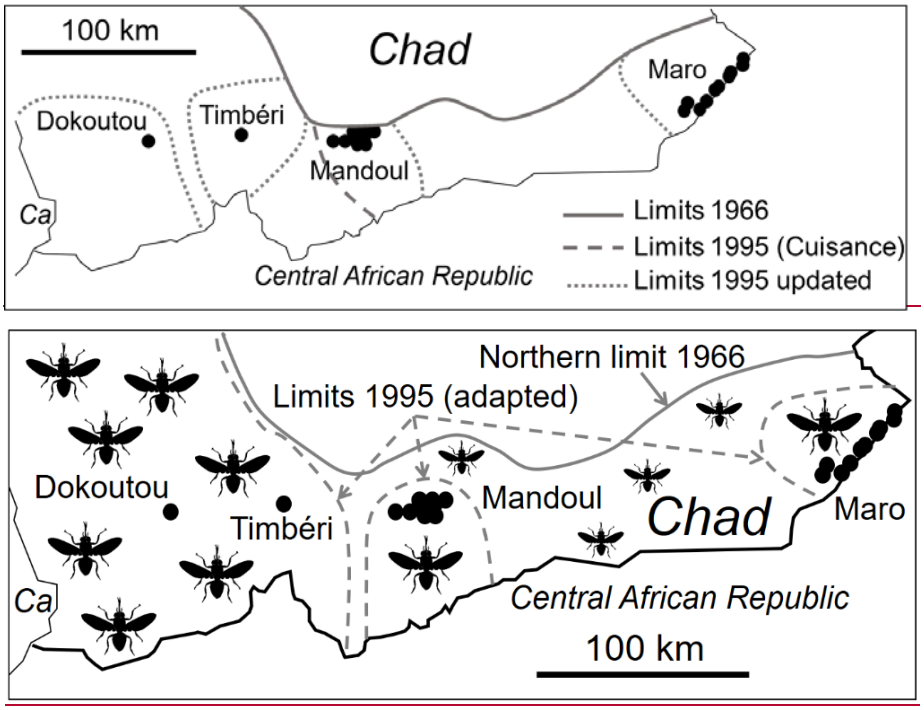
Significant male-biased dispersal has rarely been found in tsetse flies, i.e. once with *G. palpalis* in Cameroon (Mélachio et al., 2011), and twice with *G. tachinoides* in Burkina Faso (Kone et al., 2011; Ravel et al., 2013). Nevertheless, the lack of such research in the literature makes it difficult to draw any solid conclusion whether male tsetse flies disperse more than females. Although there was a tendency with *G. f. fuscipes* from Chad, it was not significant. If females tend to disperse less, they may be less available to trapping devices. The higher proportions of females found in traps, at least in Mandoul and Maro (the least dense zones), were not in line with this interpretation. Mark-release studies have found evidence for female-biased dispersal in some instances (Hargrove & Vale, 1979; Vale et al., 1984; Vreysen et al., 2013), but this is in contrast with the almost absence of genetic signatures. Again, this would require further specific investigations to be fully

understood, but whether females disperse more or less than males may be relevant for control programs.

A moderate and strong bottleneck signature was found in Mandoul and Timbéri, respectively. Previous reports have indicated the geographical retraction of the distribution of tsetse flies in southern Chad (Gruvel, 1966; Cuisance, 1995) mainly due to periods of drought and human activities that have dramatically reduced and fragmented suitable and interconnected habitats into small and isolated subpopulations of tsetse flies around the 1990s. For some reasons, the signature of such events would have been kept in Mandoul and Timbéri but not in Maro or Dokoutou. For Maro, frequent immigration from southern tsetse fly populations may easily have removed any bottleneck signature and Dokoutou was probably too small a sample to detect any bottleneck signature (type error 2). The extreme high population density found in that zone may also have limited the effect of such isolation.

We may use Cornuet and Luickart's (Cornuet & Luikart, 1996) model as explained above and in an earlier paper (De Meeûs et al., 2010) to extrapolate some informative parameters. With 9 loci, subsample sizes of 96 for Mandoul and 19 for Timbéri, and genetic diversity H_S=0.643 and 0.659, respectively, the detection of a bottleneck would have been possible with various scenarios. Nevertheless, given the actual population sizes currently observed in the two populations, it seems that the most likely combination of parameters for both zones and both models (IAM and TPM) may have been τ=1 and α=1000 (i.e. a drastic bottleneck). If so, with 108 and 138 generations since 1995 for Mandoul and Timbéri, respectively, these parameter combinations lead to $N_{e\text{-post}}$=54 for Mandoul, and $N_{e\text{-post}}$=69 for Timbéri, for the effective population sizes after the bottleneck. These values correspond to the range of values of N_e we computed for these two zones. We also computed $N_{e\text{-pre}}$=54000 according to Mandoul parameters, and $N_{e\text{-pre}}$=69000 for Timbéri, before the bottleneck. Such values, if they corresponded to anything, would probably match the global and interconnected big populations that inhabited the area before 1995. This seems to match for Mandoul, and hence Maro, that appeared as probable isolated pockets in 1995 (Figure 4). However, in 1995, Timbéri and Dokoutou were still apparently connected (Figure 4). So maybe the fragmentation occurred later between these two zones, or the 1995 investigations were not accurate enough at that time to detect a hiatus between Dokoutou and Timbéri. No matter the real scenario, populations of *G. fuscipes fuscipes* seem to have strongly declined from very high population densities to the very low densities observed during this work, at least in Mandoul and to a lesser extent in Maro.

Figure 4: Map of sample locations (dots) of *Glossina fuscipes fuscipes* of the present study with the Northern limit described by Gruvels of distributions of the species in 1966 (Gruvel, 1966) (grey line, small and big shadow flies), and speculated limits in 1995 (grey dashed lines, big shadow flies), combining different surveys: ancient (Cuisance, 1995) and more recentadapted according to recent surveys (Signaboubo et al., 2021), including the present one. (*Ca*: Cameroon).



Conclusion

Population genetics confirmed the field observations of a strong subdivision between tsetse populations in Southern Chad, together with very low population densities. Therefore, the probability of reinvasion from neighboring zones are (very) small, at least in Mandoul, Timbéri and Dokoutou. In addition, efficient barriers might be deployed permanently to prevent reinvasion from the southern areas. This was particularly obvious for the Maro focus that appeared to present the higher reinvasion risks. Tsetse eradication may thus be considered as a sustainable option for HAT elimination in Mandoul focus.

~~Control can also be advised in the Dokoutou-Timbéri zone where HAT has not been reported yet but where AAT may cause problems on animal Health.~~ For the Maro HAT focus, another strategy based on continuous tsetse suppression will probably be needed.

Acknowledgements

This study was financed by the International Atomic Energy Agency (IAEA), Austria. The authors thank Giuliano Cecchi (FAO) for providing hydrographic base maps of Southern Chad and two anonymous referees whose suggestions and remarks helped to considerably improve the present manuscript. Authors would like to dedicate this paper to our colleague and friend Dr Jean-Baptiste Rayaisse (1967–2020). Jean-Baptiste was a hard worker, very efficient researcher, deeply committed to trypanosomosis control, and with a great sense of humour. All the community misses him immensely. We will also miss Pèka Mallaye, who passed away the 13/08/2022, who was of great help as the coordinator of the Chadian PNLTHA.

Author's contributions

All authors read, amended and/or approved the final manuscript, except JBR who could not check the last versions.

Conceptualization: Mallaye Peka, Jean-Baptiste Rayaisse, Adrien Marie Gaston Belem, Philippe Solano, Jérémy Bouyer, Camille Noûs.

Sampling and field work: Mahamat Hissene Mahamat, Mallaye Peka, Jean-Baptiste Rayaisse, Mahamat, Justin Darnas, Brahim Guihini Mollo, Wilfrid Yoni, Jérémy Bouyer, Rafael Argiles-Herrero.

Genotyping, genotype interpretation and corrections: Sophie Ravel, Adeline Ségard.

Data analyses: Thierry de Meeûs.

Maps and design of figures: Jérémy Bouyer and Thierry de Meeûs.

Writing of the original draft: Sophie Ravel, Thierry de Meeûs.

Supervision: Mallaye Peka, Jean-Baptiste Rayaisse, Philippe Solano, Thierry de Meeûs, Sophie Ravel.

Data availability

Raw data are available in supplementary file S1 "Gff-TchadDataSupFile1.xlsx". Position of traps, dates of sampling and cohort of flies are available in the supplementary Figure S1 "GffChadCaptureMapsFigS1.tif". Land cover images from Google Earth Pro between Timbéri and Dokoutou for the years 2012-2018 are presented in the

supplementary file "DokTimb2012-2018GoogleEarthSupFileS2.pptx ". Example of scripts to compute geographic distances and surfaces with the package geosphere is available in Appendix 1. HierFstat scripts and results are available in the Appendix 2. Detailed analyses of quality testing of data are in Appendix 3 .Data for the DAPC analysis (package adegenet) are in the file "GffChadSpatialTrapsDAPC.txt", and the corresponding script and results in Appendix 4.

Conflict of interest disclosure

The authors declare that they have no financial conflict of interest with the content of this article. Philippe Solano and Adrien Marie Gaston Belem are recommenders of PCI Infections. Thierry de Meeûs is one of the PCI Infections administrators.

References

Balloux, F. (2004) Heterozygote excess in small populations and the heterozygote-excess effective population size. *Evolution*, **58**, 1891-1900.

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).

Benjamini, Y., Yekutieli, D. (2001) The control of the false discovery rate in multiple testing under dependency. *The Annals of Statistics*, **29**, 1165–1188.

Bouyer, J., Desquesnes, M., Yoni, W., Chamisa, A., Guerrini, L. (2015) Attracting and trapping insect vectors, Technical guide GeosAf, 2. s.n., s.l., France, p. 23.

Bouyer, J., Solano, P., de la Rocque, S., Desquesnes, M., Cuisance, D., Itard, J., Frézil, J.-L., Authié, É. (2009) Trypanosomoses: control methods, in: Lefèvre, P.C., Blancou, J., Chermette, R., Uilenberg, G. (Eds.), *Infectious and Parasitic Diseases of Livestock*. Lavoisier (Tec & Doc), Paris, pp. 1936–1943.

Mis en forme : Français (France)

1096 Brookfield, J.F.Y. (1996) A simple new method for estimating null allele frequency from
1097 heterozygote deficiency. *Molecular Ecology*, **5**, 453-455.
1098 Büscher, P., Bart, J.M., Boelaert, M., Bucheton, B., Cecchi, G., Chitnis, N., Courtin, D.,
1099 Figueiredo, L.M., Franco, J.R., Grébaut, P., Hasker, E., Ilboudo, H., Jamonneau, V., Koffi,
1100 M., Lejon, V., MacLeod, A., Masumu, J., Matovu, E., Mattioli, R., Noyes, H., Picado, A.,
1101 Rock, K.S., Rotureau, B., Simo, G., Thévenon, S., Trindade, S., Truc, P., Van Reet, N.
1102 (2018) Do cryptic reservoirs threaten gambiense-sleeping sickness elimination? *Trends in*
1103 *Parasitology*, **34**, 197-207. 10.1016/j.pt.2017.11.008
1104 Cavalli-Sforza, L.L., Edwards, A.W.F. (1967) Phylogenetic analysis: model and estimation
1105 procedures. *American Journal of Human Genetics*, **19**, 233-257.
1106 Challier, A., Laveissière, C. (1973) Un nouveau piège pour la capture des glossines
1107 (*Glossina*: Diptera, Muscidae): description et assais sur le terrain. *Cahier de l'ORSTOM,*
1108 *Série Entomologie Médicale et Parasitologie*, **11**, 251-262.
1109 Chapuis, M.P., Estoup, A. (2007) Microsatellite null alleles and estimation of population
1110 differentiation. *Molecular Biology and Evolution*, **24**, 621-631. 10.1093/molbev/msl191
1111 Coombs, J.A., Letcher, B.H., Nislow, K.H. (2008) CREATE: a software to create input files
1112 from diploid genotypic data for 52 genetic software programs. *Molecular Ecology*
1113 *Resources*, **8**, 578–580.
1114 Cornuet, J.M., Luikart, G. (1996) Description and power analysis of two tests for detecting
1115 recent population bottlenecks from allele frequency data. *Genetics*, **144**, 2001-2014.
1116 Cuisance, D. (1995) *Réactualisation de la situation des tsé-tsés et des trypanosomoses*
1117 *animales au Tchad. Enquète réalisée du 9 février au 18 mars 1995*, Maisons-Alfort :
1118 CIRAD-EMVT ed. Maisons-Alfort : CIRAD-EMVT, Maison-Alfort.
1119 Cuisance, D., Février, J., Déjardin, J., Filledier, J. (1985) Dispersion linéaire de *Glossina*
1120 *palpalis gambiensis* et *G. tachinoides* dans une galerie forestière en zone soudano-

Mis en forme : Français (France)

Mis en forme : Français (France)

guinéenne (Burkina Faso). *Revue d'Elevage et de Médecine Vétérinaire des Pays Tropicaux*, **38**, 153-172.

De Meeûs, T. (2012) *Initiation à la génétique des populations naturelles: Applications aux parasites et à leurs vecteurs*. IRD Editions, Marseille.

De Meeûs, T. (2014) Statistical decision from *k* test series with particular focus on population genetics tools: a DIY notice. *Infection Genetics and Evolution*, **22**, 91-93.

De Meeûs, T. (2018) Revisiting F_{IS}, F_{ST}, Wahlund effects, and Null alleles. *Journal of Heredity*, **109**, 446-456. 10.1093/jhered/esx106

De Meeûs, T., Chan, C.T., Ludwig, J.M., Tsao, J.I., Patel, J., Bhagatwala, J., Beati, L. (2021) Deceptive combined effects of short allele dominance and stuttering: an example with *Ixodes scapularis*, the main vector of Lyme disease in the U.S.A. *Peer Community Journal*, **1**, e40. https://doi.org/10.24072/pcjournal.34

De Meeûs, T., Goudet, J. (2007) A step-by-step tutorial to use HierFstat to analyse populations hierarchically structured at multiple levels. *Infection Genetics and Evolution*, **7**, 731-735.

De Meeûs, T., Guégan, J.F., Teriokhin, A.T. (2009) MultiTest V.1.2, a program to binomially combine independent tests and performance comparison with other related methods on proportional data. *BMC Bioinformatics*, **10**, 443.

De Meeûs, T., Humair, P.F., Grunau, C., Delaye, C., Renaud, F. (2004) Non-Mendelian transmission of alleles at microsatellite loci: an example in *Ixodes ricinus*, the vector of Lyme disease. *International Journal for Parasitology*, **34**, 943-950.

De Meeûs, T., Koffi, B.B., Barré, N., de Garine-Wichatitsky, M., Chevillon, C. (2010) Swift sympatric adaptation of a species of cattle tick to a new deer host in New-Caledonia. *Infection Genetics and Evolution*, **10**, 976-983.

De Meeûs, T., McCoy, K.D., Prugnolle, F., Chevillon, C., Durand, P., Hurtrez-Boussès, S., Renaud, F. (2007) Population genetics and molecular epidemiology or how to "débusquer la bête". *Infection Genetics and Evolution*, **7**, 308-332.

De Meeûs, T., Ravel, S., Solano, P., Bouyer, J. (2019) Negative density dependent dispersal in tsetse flies: a risk for control campaigns? *Trends in Parasitology*, **35**, 615-621.

Do, C., Waples, R.S., Peel, D., Macbeth, G.M., Tillett, B.J., Ovenden, J.R. (2014) NeEstimator v2: re-implementation of software for the estimation of contemporary effective population size (N_e) from genetic data. *Molecular Ecology Resources*, **14**, 209-214. 10.1111/1755-0998.12157

Fox, J. (2005) The R commander: a basic statistics graphical user interface to R. *Journal of Statistical Software*, **14**, 1–42.

Fox, J. (2007) Extending the R commander by "plug in" packages. *R News*, **7**, 46–52.

Frontier, S. (1976) Etude de la décroissance des valeurs propres dans une analyse en composantes principales: comparaison avec le modèle du bâton brisé. *Journal of Experimental Marine Biology and Ecology*, **25**, 67-75.

Gimonneau, G., Ouedraogo, R., Salou, E., Rayaisse, J.-B., Buatois, B., Solano, P., Dormont, L., Roux, O., Bouyer, J. (2021) Larviposition site selection mediated by volatile semiochemicals in *Glossina palpalis gambiensis*. *Ecological Entomology*, **46**, 301-309. DOI: 10.1111/een.12962

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

Goudet, J. (2003) Fstat (ver. 2.9.4), a program to estimate and test population genetics parameters. Available at http://www.t-de-meeus.fr/Programs/Fstat294.zip, Updated from Goudet (1995).

Goudet, J. (2005) HIERFSTAT, a package for R to compute and test hierarchical F-statistics. *Molecular Ecology Notes*, **5**, 184-186.

Mis en forme : Français (France)

1171 Goudet, J., Perrin, N., Waser, P. (2002) Tests for sex-biased dispersal using bi-parentally
1172 inherited genetic markers. *Molecular Ecology*, **11**, 1103–1114.
1173 Goudet, J., Raymond, M., De Meeûs, T., Rousset, F. (1996) Testing differentiation in
1174 diploid populations. *Genetics*, **144**, 1933-1940.
1175 Gruvel, J. (1966) Les glossines vectrices des trpanosomiases au Tchad. *Revue d'Elevage*
1176 *et de Médecine Vétérinaire des Pays Tropicaux*, **19**, 169-212.
1177 Guinand, B. (1996) Use of a multivariate model using allele frequency distributions to
1178 analyse patterns of genetic differentiation among populations. *Biological Journal of the*
1179 *Linnean Society*, **58**, 173-195.
1180 Hargrove, J.W., Vale, G.A. (1979) Aspects of the feasibility of employing odor-baited traps
1181 for controlling tsetse flies (Diptera, Glossinidae). *Bulletin of Entomological Research*, **69**,
1182 283-290. Doi 10.1017/S0007485300017752
1183 Hedrick, P.W. (2005a) *Genetics of Populations, Third Edition*. Jones and Bartlett
1184 Publishers, Sudbury, Massachusetts.
1185 Hedrick, P.W. (2005b) A standardized genetic differentiation measure. *Evolution*, **59**,
1186 1633-1638.
1187 Hijmans, R.J., Williams, E., Vennes, C. (2019) Package 'geosphere': Spherical
1188 Trigonometry. CRAN, pp. Spherical trigonometry for geographic applications. That is,
1189 compute distances and related measures for angular (longitude/latitude) locations.
1190 Ibrahim, M.A.M., Weber, J.S., Ngomtcho, S.C.H., Signaboubo, D., Berger, P., Hassane,
1191 H.M., Kelm, S. (2021) Diversity of trypanosomes in humans and cattle in the HAT foci
1192 Mandoul and Maro, Southern Chad-A matter of concern for zoonotic potential? *PLoS*
1193 *Neglected Tropical Diseases*, **15**, e0009323. https://doi.org/10.1371/journal.pntd.0009323
1194 Jombart, T. (2008) adegenet: a R package for the multivariate analysis of genetic markers.
1195 *Bioinformatics*, **24**, 1403-1405. 10.1093/bioinformatics/btn129

Mis en forme : Français (France)

1196 Jombart, T., Devillard, S., Balloux, F. (2010) Discriminant analysis of principal
1197 components: a new method for the analysis of genetically structured populations. *BMC*
1198 *Genetics*, **11**, 94. Artn 94
1199 10.1186/1471-2156-11-94
1200 Karney, C.F.F. (2013) Algorithms for geodesics. *Journal of Geodesy*, **87**, 43-55.
1201 10.1007/s00190-012-0578-z
1202 Kone, N., Bouyer, J., Ravel, S., Vreysen, M.J.B., Domagni, K.T., Causse, S., Solano, P.,
1203 De Meeûs, T. (2011) Contrasting population structures of two vectors of African
1204 trypanosomoses in Burkina Faso: consequences for control. *PLoS Neglected Tropical*
1205 *Diseases*, **5**, e1217. e1217
1206 10.1371/journal.pntd.0001217
1207 Kumar, S., Stecher, G., Tamura, K. (2016) MEGA7: Molecular evolutionary genetics
1208 analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution*, **33**, 1870-1874.
1209 10.1093/molbev/msw054
1210 Mahamat, M.H., Peka, M., Rayaisse, J.B., Rock, K.S., Toko, M.A., Darnas, J., Brahim,
1211 G.M., Alkatib, A.B., Yoni, W., Tirados, I., Courtin, F., Brand, S.P.C., Nersy, C., Alfaroukh,
1212 I.O., Torr, S.J., Lehane, M.J., Solano, P. (2017) Adding tsetse control to medical activities
1213 contributes to decreasing transmission of sleeping sickness in the Mandoul focus (Chad).
1214 *PLoS Neglected Tropical Diseases*, **11**, e0005792. ARTN e0005792
1215 10.1371/journal.pntd.0005792
1216 Manangwa, O., De Meeûs, T., Grébaut, P., Segard, A., Byamungu, M., Ravel, S. (2019)
1217 Detecting Wahlund effects together with amplification problems : cryptic species, null
1218 alleles and short allele dominance in Glossina pallidipes populations from Tanzania.
1219 *Molecular Ecology Resources*, **19**, 757-772. 10.1111/1755-0998.12989
1220 Manly, B.J.F. (1997) *Randomization and Monte Carlo Methods in Biology: Second Edition*.
1221 Chapman & Hall, London.

Meirmans, P.G. (2006) Using the amova framework to estimate a standardized genetic differentiation measure. *Evolution*, **60**, 2399–2402.

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

Mélachio, T., Tito, Tanekou, Simo, G., Ravel, S., De Meeûs, T., Causse, S., Solano, P., Lutumba, P., Asonganyi, T., Njiokou, F. (2011) Population genetics of *Glossina palpalis palpalis* from central African sleeping sickness foci. *Parasites and Vectors*, **4**, 140.

Ndung'u, J.M., Boulangé, A., Picado, A., Mugenyi, A., Mortensen, A., Hope, A., Mollo, B.G., Bucheton, B., Wamboga, C., Waiswa, C., Kaba, D., Matovu, E., Courtin, F., Garrod, G., Gimonneau, G., Bingham, G.V., Hassane, H.M., Tirados, I., Saldanha, I., Kabore, J., Rayaisse, J.B., Bart, J.M., Lingley, J., Esterhuizen, J., Longbottom, J., Pulford, J., Kouakou, L., Sanogo, L., Cunningham, L., Camara, M., Koffi, M., Stanton, M., Lehane, M., Kagbadouno, M.S., Camara, O., Bessell, P., Mallaye, P., Solano, P., Selby, R., Dunkley, S., Torr, S., Biéler, S., Lejon, V., Jamonneau, V., Yoni, W., Katz, Z. (2020) Trypa-NO! contributes to the elimination of gambiense human African trypanosomiasis by combining tsetse control with "screen, diagnose and treat" using innovative tools and strategies. *PLoS Neglected Tropical Diseases*, **14**, e0008738. ARTN e0008738 10.1371/journal.pntd.0008738

Nomura, T. (2008) Estimation of effective number of breeders from molecular coancestry of single cohort sample. *Evolutionary Applications*, **1**, 462-474.

Pearson, K. (1903) Assortative mating in man: a cooperative study. *Biometrika*, **2**, 481-498. https://doi.org/10.2307%2F2331510

Peel, D., Waples, R.S., Macbeth, G.M., Do, C., Ovenden, J.R. (2013) Accounting for missing data in the estimation of contemporary genetic effective population size (N_e). *Molecular Ecology Resources*, **13**, 243-253.

1247 Piry, S., Luikart, G., Cornuet, J.M. (1999) BOTTLENECK: A computer program for
1248 detecting recent reductions in the effective population size using allele frequency data.
1249 *Journal of Heredity*, **90**, 502-503.
1250 Prugnolle, F., De Meeûs, T. (2002) Inferring sex-biased dispersal from population genetic
1251 tools: a review. *Heredity*, **88**, 161-165.
1252 Prugnolle, F., De Meeûs, T. (2010) Apparent high recombination rates in clonal parasitic
1253 organisms due to inappropriate sampling design. *Heredity*, **104**, 135-140.
1254 R-Core-Team (2020) R: A Language and Environment for Statistical Computing, Version
1255 3.6.3 (2020-02-29) ed. R Foundation for Statistical Computing, Vienna, Austria,
1256 http://www.R-project.org.
1257 Ravel, S., Rayaisse, J.-B., Courtin, F., Solano, P., De Meeûs, T. (2013) Genetic signature
1258 of a recent southern range shift in *Glossina tachinoides* in East Burkina Faso. *Infection*
1259 *Genetics and Evolution*, **18**, 309-314.
1260 Ravel, S., Sere, M., Manangwa, O., Kagbadouno, M., Mahamat, M.H., Shereni, W.,
1261 Okeyo, W.A., Argiles-Herrero, R., De Meeûs, T. (2020) Developing and quality testing of
1262 microsatellite loci for four species of *Glossina*. *Infection Genetics and Evolution*, **85**,
1263 104515. 10.1016/j.meegid.2020.104515
1264 Rousset, F. (1997) Genetic differentiation and estimation of gene flow from *F*-statistics
1265 under isolation by distance. *Genetics*, **145**, 1219-1228.
1266 Rousset, F. (2008) GENEPOP ' 007: a complete re-implementation of the GENEPOP
1267 software for Windows and Linux. *Molecular Ecology Resources*, **8**, 103-106.
1268 Saitou, N., Nei, M. (1987) The neighbor-joining method: a new method for reconstructing
1269 phylogenetic trees. *Molecular Biology and Evolution*, **4**, 406-425.
1270 Séré, M., Thévenon, S., Belem, A.M.G., De Meeûs, T. (2017) Comparison of different
1271 genetic distances to test isolation by distance between populations. *Heredity*, **119**, 55-63.
1272 10.1038/hdy.2017.26

She, J.X., Autem, M., Kotulas, G., Pasteur, N., Bonhomme, F. (1987) Multivariate analysis of genetic exchanges between *Solea aegyptiaca* and *Solea senegalensis* (Teleosts, Soleidae). *Biological Journal of the Linnean Society*, **32**, 357-371. DOI 10.1111/j.1095-8312.1987.tb00437.x

Signaboubo, D., Payne, V.K., Moussa, I.M.A., Hassane, H.M., Berger, P., Kelm, S., Simo, G. (2021) Diversity of tsetse flies and trypanosome species circulating in the area of Lake Iro in southeastern Chad. *Parasites & Vectors*, **14**, 293. 10.1186/s13071-021-04782-7

Solano, P., Ravel, S., De Meeûs, T. (2010) How can tsetse population genetics contribute to African trypanosomiasis control? *Trends in Parasitology*, **26**, 255-263.

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

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

Thomas, F., Renaud, F., Derothe, J.M., Lambert, A., De Meeûs, T., Cezilly, F. (1995) Assortative pairing in *Gammarus-Insensibilis* (Amphipoda) Infected by a Trematode parasite. *Oecologia*, **104**, 259-264.

Traynor, K.S., Mondet, F., de Miranda, J.R., Techer, M., Kowallik, V., Oddie, M.A.Y., Chantawannakul, P., McAfee, A. (2020) *Varroa destructor*: a complex parasite, crippling honey bees worldwide. *Trends in Parasitology*, **36**, 592-606. 10.1016/j.pt.2020.04.004

Vale, G.A., Hursey, B.S., Hargrove, J.W., Torr, S.J., Allsopp, R. (1984) The use of small plots to study populations of tsetse (Diptera, Glossinidae): difficulties associated with population dispersal. *Insect Science and its Application*, **5**, 403-410. Doi 10.1017/S1742758400008730

Van Oosterhout, C., Hutchinson, W.F., Wills, D.P.M., Shipley, P. (2004) MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes*, **4**, 535-538.

1299 Vitalis, R. (2002) Estim 1.2-2: a computer program to infer population parameters from
1300 one- and two-locus gene identity probabilities, Available at http://www.t-de-
1301 meeus.fr/ProgMeeusGB.html.
1302 Vitalis, R., Couvet, D. (2001a) ESTIM 1.0: a computer program to infer population
1303 parameters from one- and two-locus gene identity probabilities. *Molecular Ecology Notes*,
1304 **1**, 354-356.
1305 Vitalis, R., Couvet, D. (2001b) Estimation of effective population size and migration rate
1306 from one- and two-locus identity measures. *Genetics*, **157**, 911-925.
1307 Vreysen, M.J.B., Balenghien, T., Saleh, K.M., Maiga, S., Koudougou, Z., Cecchi, G.,
1308 Bouyer, J. (2013) Release-Recapture Studies Confirm Dispersal of *Glossina palpalis*
1309 *gambiensis* between River Basins in Mali. *PLoS Neglected Tropical Diseases*, **7**, e2022.
1310 ARTN e2022
1311 10.1371/journal.pntd.0002022
1312 Wahlund, S. (1928) Zusammensetzung von populationen und korrelationsers-chinungen
1313 von standpunkt der vererbungslehre aus betrachtet. *Hereditas*, **11**, 65-106.
1314 Wang, J. (2015) Does G_{ST} underestimate genetic differentiation from marker data?
1315 *Molecular Ecology*, **24**, 3546-3558. 10.1111/mec.13204
1316 Waples, R.S. (2006) A bias correction for estimates of effective population size based on
1317 linkage disequilibrium at unlinked gene loci. *Conservation Genetics*, **7**, 167-184.
1318 Watts, P.C., Rousset, F., Saccheri, I.J., Leblois, R., Kemp, S.J., Thompson, D.J. (2007)
1319 Compatible genetic and ecological estimates of dispersal rates in insect (*Coenagrion*
1320 *mercuriale*: Odonata: Zygoptera) populations: analysis of 'neighbourhood size' using a
1321 more precise estimator. *Molecular Ecology*, **16**, 737-751.
1322 Weir, B.S., Cockerham, C.C. (1984) Estimating F-statistics for the analysis of population
1323 structure. *Evolution*, **38**, 1358-1370.

Werren, J.H. (1980) Sex ratio adaptations to local mate competition in a parasitic wasp. *Science*, **208**, 1157-1159. 10.1126/science.208.4448.1157

Wright, S. (1965) The interpretation of population structure by F-statistics with special regard to system of mating. *Evolution*, **19**, 395-420.

Yang, R.C. (1998) Estimating hierarchical *F*-statistics. *Evolution*, **52**, 950-956.

APPENDICES

Appendix 1: Example of scripts to compute geographic distances or surfaces with the R package geosphere

```
# to compute geographic distance (in meters) with GPPS coordinate in decimal
# degrees: long1 and lat1, and long2 and lat2 for the coordinates of points 1
# and 2 respectively.

distGeo(c(long1,lat1),c(long2, lat2))

#With two files with two columns (longitude and latitude), the first file
#containing the GPS coordinates of the first point of site pairs, and the second
#file containing the corresponding GPS coordinates of the second point of site
#pairs.

LongLat1 <- read.table("Long1Lat1.txt", header=TRUE, stringsAsFactors=TRUE,
sep="\t", na.strings="NA", dec=".", strip.white=TRUE)
LongLat2 <- read.table("Long2Lat2.txt", header=TRUE, stringsAsFactors=TRUE,
sep="\t", na.strings="NA", dec=".", strip.white=TRUE)
distGeo(LongLat1,LongLat2)

# To compute the area of a polygon in angular coordinates (longitude/latitude)
#on an ellipsoid.
#Dataset has two columns : Longitude and Latitude
Dataset <- read.table("MyData.txt", header=TRUE, stringsAsFactors=TRUE,
sep="\t", na.strings="NA", dec=".", strip.white=TRUE)
attach(Dataset)
areaPolygon(data.frame(Longitude,Latitude))
```

Appendix 2: Script and results for the HierFstat analysis

For Mandoul

```
> data<-read.table("MandoulHier.txt",header=TRUE)
> attach(data)
> loci<-
data.frame(Locus1,Locus2,Locus3,Locus4,Locus5,Locus6,Locus7,Locus8,Locus9)
> levels<-data.frame(Site,Subsite,Trap)
> varcomp.glob(levels,loci)
$F
                Site    Subsite         Trap       Ind
Total   0.0004076288 0.01821699 -0.008157967 0.1326307
Site    0.0000000000 0.01781663 -0.008569089 0.1322770
```

Mis en forme : Français (France)

```
Subsite 0.0000000000 0.00000000 -0.026864347 0.1165367
Trap    0.0000000000 0.00000000  0.000000000 0.1396494
> test.within(loci,test=Trap,within=Subsite,nperm=1000)
$p.val
[1] 0.72
> test.between.within(loci,within=Site,rand.unit=Trap,test=Subsite,nperm=1000)
$p.val
[1] 0.025

> test.between(loci,rand.unit=Subsite,test=Site,nperm=1000)
$p.val
[1] 0.303
```

For Maro

```
> data<-read.table("MaroT0Hier.txt",header=TRUE)
> attach(data)
> loci<-
data.frame(Locus1,Locus2,Locus3,Locus4,Locus5,Locus6,Locus7,Locus8,Locus9)
> levels<-data.frame(Site,Subsite,Trap)
> varcomp.glob(levels,loci)
$F
                Site      Subsite         Trap        Ind
Total   -0.006634284 -0.006208540 -0.001297964 0.07391164
Site     0.000000000  0.000422938  0.005301150 0.08001508
Subsite  0.000000000  0.000000000  0.004880276 0.07962582
Trap     0.000000000  0.000000000  0.000000000 0.07511211

> test.within(loci,test=Trap,within=Subsite,nperm=1000)
$p.val
[1] 0.031

> test.between.within(loci,within=Site,rand.unit=Trap,test=Subsite,nperm=1000)
$p.val
[1] 0.656

> test.between(loci,rand.unit=Subsite,test=Site,nperm=1000)
$p.val
[1] 0.567
```

For Timbéri and Dokoutou

```
> data<-read.table("TimberiDokoutouHier.txt",header=TRUE)
> attach(data)
```

```
> loci<-
data.frame(Locus1,Locus2,Locus3,Locus4,Locus5,Locus6,Locus7,Locus8,Locus9)
> levels<-data.frame(Zone,Subsite,Trap)
> varcomp.glob(levels,loci)
$F
              Zone    Subsite          Trap        Ind
Total   0.07493418 0.09378198  0.0754573949 0.15831925
Zone    0.00000000 0.02037454  0.0005655925 0.09013960
Subsite 0.00000000 0.00000000 -0.0202209403 0.07121605
Trap    0.00000000 0.00000000  0.0000000000 0.08962470

> test.within(loci,within=Subsite,test=Trap,nperm=1000)
$p.val
[1] 0.961

> test.between.within(loci,within=Zone,rand.unit=Trap,test=Subsite,nperm=1000)
$p.val
[1] 0.66

> test.between(loci,rand.unit=Subsite,test=Zone,nperm=1000)
0.196
```

Timbéri and Doukoutou without subsites

```
> data<-read.table("TimberiDokoutouHier.txt",header=TRUE)
> attach(data)
> loci<-
data.frame(Locus1,Locus2,Locus3,Locus4,Locus5,Locus6,Locus7,Locus8,Locus9)
> levels<-data.frame(Zone,Trap)
> varcomp.glob(levels,loci)
$loc
               [,1]         [,2]        [,3]      [,4]
Locus1 -0.005416145  0.049552374  0.11500959 0.3913043
Locus2  0.079303509 -0.036826228 -0.04782076 0.9130435
Locus3  0.025164936 -0.048459828  0.15830136 0.6956522
Locus4  0.094947329 -0.026895027  0.09454775 0.5777778
Locus5  0.073639708  0.010224110  0.19631820 0.2826087
Locus6  0.069050675 -0.019157955  0.18475597 0.5434783
Locus7  0.102895393 -0.026648790 -0.02472293 0.8043478
Locus8  0.100433760  0.004003843 -0.02738751 0.8222222
Locus9  0.073513023  0.008805412 -0.06188017 0.9333333
```

Mis en forme : Anglais (États-Unis)

Mis en forme : Français (France)

Mis en forme : Anglais (États-Unis)

Mis en forme : Français (France)

```
$overall
       Zone        Trap         Ind       Error
 0.61353219 -0.08540209  0.58712151  5.96376812

$F
            Zone        Trap        Ind
Total 0.08666909  0.07460498 0.15754323
Zone  0.00000000 -0.01320892 0.07759963
Trap  0.00000000  0.00000000 0.08962470

> test.between(loci,rand.unit=Trap,test=Zone,nperm=1000)
$p.val
[1] 0.004
```

This shows that without Subsites, Zone becomes significant

Dokoutou and Timbéri without Traps

```
> data<-read.table("TimberiDokoutouHier.txt",header=TRUE)
> attach(data)

> levels<-data.frame(Zone,Subsite)
> varcomp.glob(levels,loci)
$F
              Zone    Subsite        Ind
Total   0.07704913 0.08702932 0.15810591
Zone    0.00000000 0.01081335 0.08782351
Subsite 0.00000000 0.00000000 0.07785200

> test.within(loci,test=Subsite,within=Zone,nperm=1000)
$p.val
[1] 0.648

> test.between(loci,rand.unit=Subsite,test=Zone,nperm=1000)
$p.val
[1] 0.186
```

This show that, without traps, Subsite stays non-significant.

Mis en forme : Français (France)

Appendix 3: Detailed analyses of quality testing of genetic markers and sampling

In dioecious species as tsetse flies, heterozygote deficits can occur as a result of amplification problems (null alleles, short allele dominance, stuttering or allelic dropouts), under-dominant selection, assortative mating, systematic breeding between relatives (sib mating) and Wahlund effect.

Null alleles occur when a particular kind of allele cannot be amplified and then appears homozygous for the other allele with which it is heterozygous, or as a missing data when homozygous itself. In case of null alleles, we expect that StdrdErrFIS≥2×StdrdErrFST, a positive correlation between F_{IS} and F_{ST} across loci, and a positive correlation between the number of missing genotypes (N_{blanks}) and F_{IS} across loci (De Meeûs, 2018). We tested these correlations with rcmdr (one-sided Spearman's rank correlation tests). We also undertook the regression F_{IS}~N_{blanks}, where the determination coefficient provided a proxy of the percentage of variance of F_{IS} explained by null alleles, and where the intercept provides a proxy of the "true" F_{IS} in absence of null alleles. Null allele frequencies were estimated with Brookfield's second method (Brookfield, 1996) with MicroChecker (Van Oosterhout et al., 2004). We used these to estimate the total expected number of missing genotypes per locus ($N_{blanks\text{-}expected}$) and when useful, compared it to N_{blanks} with a one-sided (less) exact binomial test under R (command binom.test).

Short allele dominance (SAD) occurs when competition for the Taq polymerase favors the shortest allele in a heterozygote individual (De Meeûs et al., 2004). It was tested with a one sided (negative correlation) Spearman's rank correlation between F_{IT} and allele size (Manangwa et al., 2019). In case of doubt, we validated the result with a linear regression F_{IS}~Allele size weighted by $p_i(1\text{-}p_i)$ (De Meeûs et al., 2004), where p_i is the frequency of allele i. These tests were undertaken with rcmdr.

Stuttering is the result of inaccurate PCR amplification through Taq slippage of a specific DNA strand. This generates several PCR products that differ from each other by one repeat and can cause difficulties when discriminating homozygotes and individuals that are heterozygous for alleles with a single repeat difference. The presence of stuttering was detected with the graphic output of MicroChecker. As recommended (De Meeûs et al., 2021), we considered that the observed deficit of heterozygous individuals for one repeat difference was a likely consequence of stuttering (we ignored the comments panel that happened to contradict the graphic in some instances) and set the randomization at the maximum value (10000).We tried to correct loci with stuttering as in (De Meeûs et al., 2021): Alleles that are close in size were pooled into one synthetic allele, providing that

1528 one of these alleles has a frequency $p \geq 0.05$, in order to avoid giving too much weight to a
1529 collection of rare alleles. If all alleles are one repeat difference, we tried pooling alleles two
1530 by two. If close alleles are all rare, we did not pool those. These corrections were kept only
1531 for the loci for which F_{IS} of corrected data displayed a decrease as compared to the
1532 uncorrected data.
1533 Underdominance is a process that affects loci where the heterozygous individuals
1534 are less fit that all homozygous genotypes. This phenomenon must be very rare because it
1535 induces a rapid elimination of the rarest alleles, since rare alleles are mostly found
1536 heterozygous. The only documented example is the Rhesus system Rh-/Rh-, where
1537 heterozygous fetuses carried by mothers that are homozygous for Rh- are strongly
1538 disfavored (see for example the book from Hedrick page 180 (Hedrick, 2005a)). The rarity
1539 of such systems, is explained by the fact that rare alleles, which are mostly found in
1540 heterozygous individuals, tend to be rapidly eliminated from populations. Underdominance
1541 is thus highly unlikely to be found associated with a microsatellite marker.
1542 Assortative pairing occurs when individuals mate according to their genotype:
1543 carrier of a given allele prefer to mate with those that carry the same allele. This kind of
1544 systems are not expected to be frequently met in nature as it strongly disfavors the rarest
1545 alleles. There are however some examples with complex determinisms as assortative
1546 mating for size or assortative mating for parasite load (Pearson, 1903; Thomas et al.,
1547 1995). Again, microsatellite markers should not be concerned.
1548 Systematic breeding between relatives occurs when individuals mate preferentially
1549 between relatives as sib mating, due to constraints of life cycles like in some arthropods
1550 like *Nasonia* parasitoid wasps (Werren, 1980) or *Varroa* mites (Traynor et al., 2020).
1551 Wahlund effect (Wahlund, 1928; De Meeûs, 2018) corresponds to a population
1552 genetics syndrome coming from the admixture of individuals from different subpopulations
1553 that do not share the same allele frequencies into the same sample. It produces
1554 heterozygote deficits as compared to Hardy-Weinberg expected genotypic proportions,
1555 and also affects linkage disequilibrium between loci, positively or negatively so, depending
1556 on the initial genetic structure of the different subsamples (Prugnolle & De Meeûs, 2010).
1557
1558 *In Mandoul*
1559 Taking subsites as subpopulation units, only one LD test was significant (p-
1560 value=0.0446), which did not stay significant after BY correction (p-value=1). The global
1561 F_{IS}=0.128 in 95%CI=[0.039, 0.243], was significantly different from 0 (p-value<0.0002).
1562 Population structure was weak, with a small and marginally not significant F_{ST}=0.005 in

95%CI=[-0.007, 0.016] (p-value=0.0722). Interestingly, F_{IT}=0.132 in 95%CI=[0.047, 0.244] was not significantly different from the F_{IS} (p-value=0.2129). It is thus possible that the whole focus behaves as a single population.

Using criteria defined in previous works (De Meeûs, 2018; Manangwa et al., 2019; De Meeûs et al., 2021), null alleles explained well observed heterozygote deficits. Indeed, StdrdErrFIS was 10 times StdrdErrFST, and the correlation between missing data and F_{IS} was significant (ρ=0.661, p-value=0.0263) with a regression's R^2=0.55. With F_{IT}, the relationship improved (ρ=0.6738, p-value=0.0233, R^2=0.5795). Using F_{IS} or F_{IT} regressions, the intercept was used to estimate the residual values in absence of null alleles, which were F_{IS_res}=-0.0547 and F_{IT_res}=-0.0474. No signature of SAD (smaller p-value=0.175), or of stuttering could be detected. Null alleles average frequency was around p_{nulls}=0.177 with Brookfield's second method (MicroChecker).

There was no evidence of any Wahlund effect.

In Maro

Only one locus pair displayed a marginally significant LD (p-value=0.0444), which did not stay significant after BY correction (p-value=1).

There was a highly significant heterozygote deficit within traps in that focus: F_{IS}=0.091 in 95%CI=[0.026, 0.164]. Interestingly, the F_{IT} was smaller than F_{IS}: F_{IT}=0.088 in 95%CI=[0.020, 0.162], but not significantly so (p-value=0.1548, two-sided Wilcoxon signed rang test for paired data). This is due to a global negative F_{ST}=-0.005 in 95%CI=[-0.01, 0.001] (p-value=0.3363). We thus considered the whole focus as a single population. Doing so, the within focus F_{IS}=0.088 in 95%CI=[0.021, 0.163], which is smaller than the within traps F_{IS}, but again, not significantly so (p-value=0.1379, two-sided test). There was thus potentially a free migration within this focus, and in particular between the most distant traps that captured tsetse flies that were 33 km distant from each other's.

Within traps, StdrdErrFIS was 12 times STdrdErrFST, and there was a positive correlation between F_{IS} and F_{ST} (ρ=0.2176, p-value=0.2869), which suggests the existence of null alleles. Within the whole focus, the observed F_{IS} was poorly explained by missing data (ρ=0.11, p-value=0.389). No significant SAD signature could be found at any locus (all p-values>0.1478). According to Brookfield's second method, null alleles frequencies explain well the observed F_{IS} and missing data (all p-values>0.5). Additionally, there was a highly significant signature of stuttering (p-value<0.01) for locus Gff18. Stuttering detection is not very powerful and null alleles do not explain very well the observed F_{IS}. We thus tried to correct stuttering for all loci that displayed a deficit in heterozygosity for alleles with

one repeat difference: Gff3, Gff4, Gff12, Gff16, Gff18 and Gff27, following the rules described in (De Meeûs et al., 2021). For locus Gff3, we pooled allele 196 to 202 into one allele and the same for 214-218; for locus Gff4, we pooled alleles 140-152 and 156-172; for locus Gff12, we pooled 137 with 139 and 143-155; for locus 16, 156-166; for locus 18, 212 with 214 and 220-228; and for locus Gff27, 167 with 169 and 187-207. The consequences of this new coding and possible cure of stuttering effects were first explored on F_{IS} within traps. The correction improved the results for locus Gff3 (-0.031 before, -0.119 after), for Gff12 (0.108 before, 0.024 after), for Gff16 (0.267 before, 0.067 after), for Gff18 (0.269 before, -0.161 after), and for Gff27 (0.173 before, -0.051 after). Stuttering correction had no effect on Gff4 (0.025 before, 0.044 after). We thus kept these stuttering recoding for all loci but Gff4 for further analyses.

There was no evidence of any Wahlund effect.

In Dokoutou and Timbéri

Given the results obtained with the hierarchical analysis, we took directly the whole zones as subpopulation units, except when specified otherwise.

Within the two zones, only one pair of loci appeared in significant linkage (p-value=0.0307), which did not stay significant after BY correction (p-value=1). There was a substantial and highly significant heterozygote deficit, F_{IS}=0.08 in 95%CI=[-0.011, 0.191] (p-value=0.0028). It was in fact smaller, but not significantly so, than the F_{IS}=0.09 in 95%CI=[0.001, 0.196] measured within traps (p-value=0.4258, two-sided test). The site was thus probably the correct subpopulation scale. The standard error of F_{IS} was four times the one of F_{ST}, which suggested the presence of null alleles or other amplification problems. The correlation between F_{IS} and F_{ST} was weak and not significant (ρ=0.1255, p-value=0.3738). The correlation between F_{IS} and the number of missing genotypes was negative (ρ=-0.3651, p-value=0.8331). However, with three blank genotypes there was little opportunity to find anything. No significant signature of SAD could be found (smallest p-value=0.1332). According to Brookfield's second method, missing data were enough to explain the observed heterozygote deficit with null alleles (smallest p-value=0.4242). But again, subsample sizes may not have been big enough. Stuttering was significant for Gff16 and Gff18 in Dokoutou. Given the low power of the detection procedures, we tried to correct for stuttering for all loci with heterozygote deficits: Gff3 (F_{IS}=0.281), Gff8 (F_{IS}=0.148), Gff12 (F_{IS}=0.113), Gff16 (F_{IS}=0.419) and Gff18 (F_{IS}=0.238). For Gff3, we pooled alleles 202 and 204 with 200; for Gff8, 160 with 158, 176 to182 with 174, and 192 with 190; for Gff12, 145 with 143, and 151 and 153 with 149; for Gff16, 158 with 156, and

162-166 with 160; and for Gff18, 224 with 222, 234-238 with 232, and 244 with 242. The results was very good for Gff8 (F_{IS}=0.018), Gff12 (F_{IS}=0.001) and Gff18 (F_{IS}=0.035), but very bad for Gff3 (F_{IS}=0.331) and Gff16 (F_{IS}=643). We thus further kept stuttering correction for Gff8, Gff12 and Gff18 only.

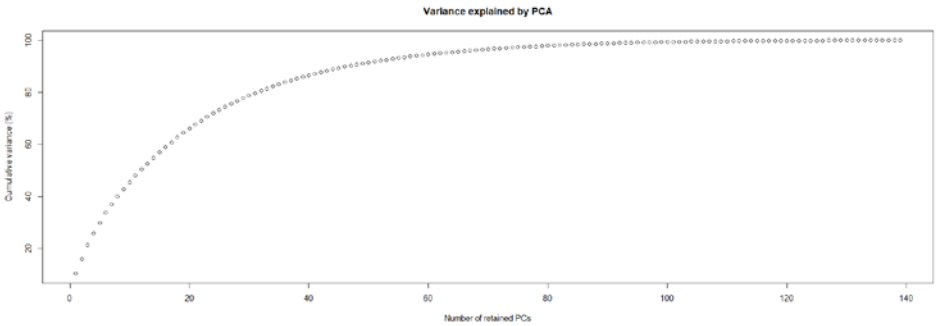
Four locus pairs appeared in significant LD (smallest *p*-value=0.0282), none of which stayed significant after BY correction (all *p*-values=1). The heterozygote deficit (F_{IS}=0.031) was not significant any more (*p*-value=0.2906). The standard error of F_{IS} was still four times the one of F_{ST}, suggesting some kind of amplification problems at some loci, which are not very well explained by null alleles (correlations between F_{IS} and F_{ST} or number of missing genotypes were both negative). Nevertheless, Gff3 and Gff16, that did not display any missing genotype, could be explained by null alleles according to Brookfield's second method, with frequencies 0.09 and 0.14 (*p*-value=0.6868 and *p*-value=0.4242), for Gff3 and Gff16 respectively.

There was no evidence of any Wahlund effect.

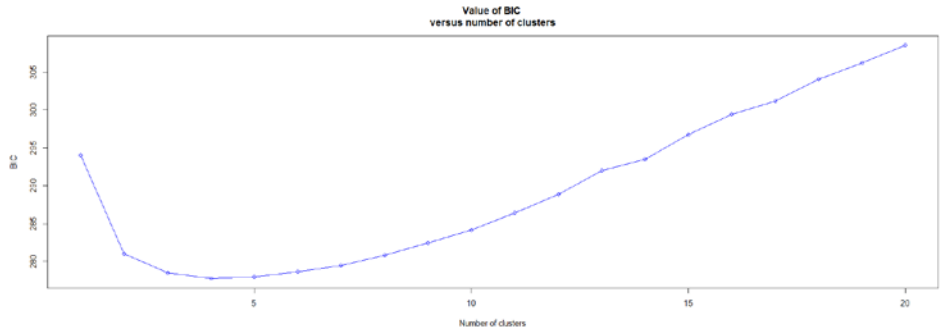
Appendix 4: script, outputs and discussion for the DAPC analysis of *Glossina fuscipes fuscipes* from southern Chad, with the R package adegenet

Scripts and outputs

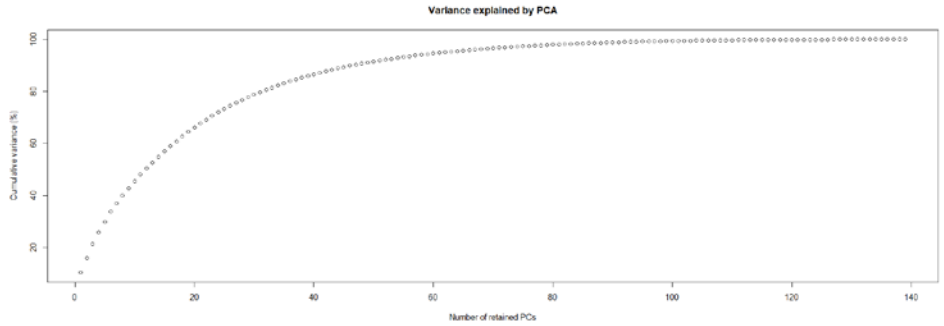
```
> GffChadSpatial<-read.table("GffChadSpatialTrapsDAPC.txt", header=TRUE,
sep="\t", na.strings="NA", dec=".", strip.white=TRUE)
> GffChadSpatialADE<-df2genind(GffChadSpatial, sep = NULL, ncode = 3, ind.names
= NULL, loc.names = NULL, pop = NULL, NA.char = "NA", ploidy = 2, type =
"codom", strata = NULL, hierarchy = NULL)
> x<-GffChadSpatialADE
> grp<-find.clusters(x,max.n.clust=20)
```



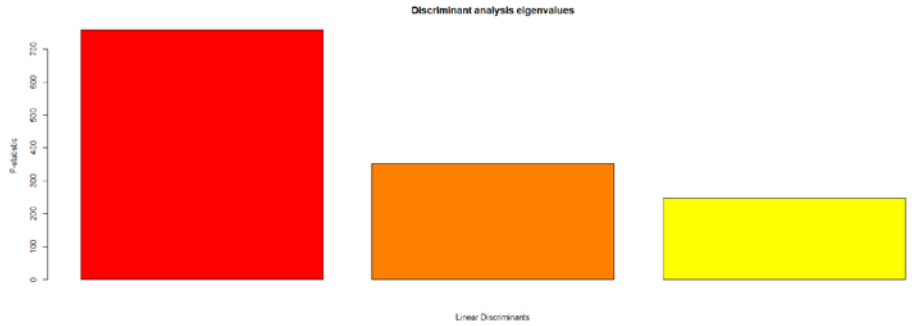

```
Choose the number PCs to retain (>= 1): 100
```



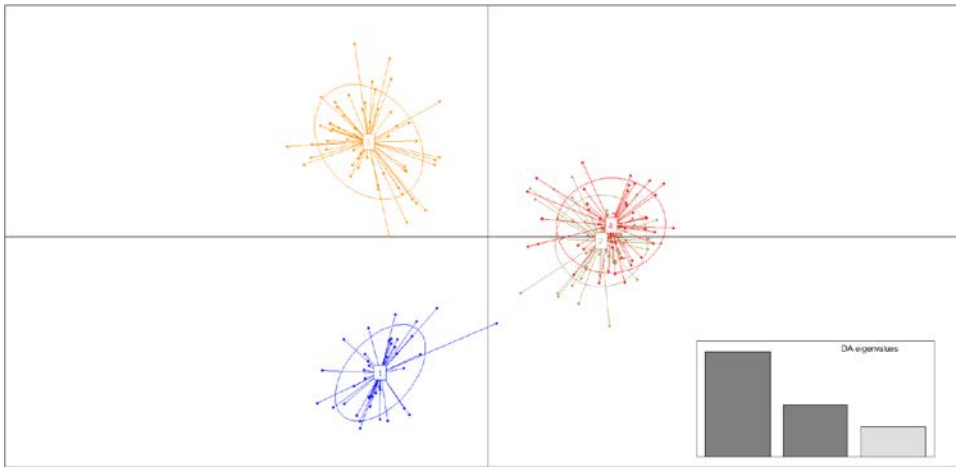

```
Choose the number of clusters (>=2): 4
> dapc1 <- dapc(x, grp$grp,n.pca= NULL, n.da= NULL, var.contrib = TRUE, scale =
FALSE)
```




```
Choose the number PCs to retain (>=1): 100
```




```
Choose the number discriminant functions to retain (>=1): 3
scatter(dapc1)
```




```
> summary(dapc1)
$n.dim
[1] 3

$n.pop
[1] 4

$assign.prop
[1] 1

$assign.per.pop
1 2 3 4
1 1 1 1

$prior.grp.size

 1  2  3  4
44 57 60 44

$post.grp.size

 1  2  3  4
44 57 60 44
> tabGffChadSpatial<-data.frame(Cluster=c(grp$grp),Proportion_assign_cluster
=dapc1$posterior,geno=GffChadSpatial)
> write.table(tabGffChadSpatial,"tabGffChadSpatialTDAPCResK4.txt",col=NA,
sep="\t", dec=".")
> write.table(dapc1$ind.coord, "CoordDAPC.txt", sep="\t")
```

```
> write.table(dapc1$means, "GroupMeansDAPC.txt", sep="\t")
> write.table(dapc1$grp.coord, "GroupCoordDAPC.txt", sep="\t")
```

Results and discussion

The optimal partition consisted of four clusters (as the number of samples), with a strong average assignment (~1), but containing admixtures of individuals from different zones, even if some clusters contained more individuals from particular zones than others (Figure A1).

Combined effects of occasional exchange, isolation by distance, temporal effects and amplification issues probably explain why the DAPC analysis provided hardly interpretable results. This challenges the relevance of this approach in some instances, but this would require further new theoretical approaches.

Figure A1: Projection on the two first axes (top) and axes 1 and 3 (bottom) of the DAPC analyses of individuals of *Glossina fuscipes fuscipes* from Southern Chad. The belonging to a particular focus/site are represented by different colors. Averages of the four clusters are symbolized by big circles of different colors. Mandoul flies belong to cohort 1, Maro to cohort 22 and Dokoutou and Timbéri to cohort 32.