Version of Record: https://www.sciencedirect.com/science/article/pii/S1471492219301114
Manuscript_0c4ef7a2de95c6051a49242b0908034f

Negative density dependent dispersal in tsetse flies: a risk for control campaigns?

Thierry De Meeûs[1]*, Sophie Ravel[1], Philippe Solano[1] and Jérémy Bouyer[1,2,3]

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

[2] Astre, Cirad, Inra, Montpellier, France.

[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

* Corresponding author: thierry.demeeus@ird.fr

Keywords: *Glossina*, Sleeping sickness, Nagana, Control campaigns, Effective population size, Migration rate.

Abstract

Tsetse flies are vectors of parasites that cause diseases responsible for significant economic losses and health issues in sub-Saharan Africa, including sleeping sickness in humans and nagana in domestic animals. Efficient vector control campaigns require good knowledge of the demographic parameters of the targeted populations. In the last decade, population genetics emerged as a convenient way to measure population densities and dispersal in tsetse flies. Here by revealing a strong negative density-dependent dispersal in two dimensions, we suggest that control campaigns might unleash dispersal from untreated areas. If confirmed by direct measurement of dispersal before and after control campaigns, area wide and/or sequential treatments of neighboring sites will be necessary to prevent this issue.

© 2019. This manuscript version is made available under the Elsevier user license
https://www.elsevier.com/open-access/userlicense/1.0/

Tsetse fly control and population genetics tools

Flies of the genus *Glossina* are the vectors of trypanosomes, parasites responsible for diseases that lead to a significant disease burden and economic losses in sub-Saharan Africa. In humans, tsetse flies transmit trypanosomiasis, also known as sleeping sickness" which is usually mortal if the patient is not treated [1]. In animals, particularly cattle, it causes nagana, a disease known to cause considerable economic losses in sub-Saharan Africa [2, 3]. Vector control has become a recognized key component in the management of these plagues [3-5]. Nevertheless, to optimize control campaigns, good knowledge of the biology of the target populations is needed, in particular their density and dispersal ability [6]. Lack of such knowledge has indeed been implicated in the failure of some elimination campaigns due to the raid reinvasion of treated populations by local flies or by flies from neighboring sites [7]. **Population genetics** (see Glossary) offers useful tools [8] in particular for tsetse flies [3].

Considerable work has been undertaken in recent years to study the population biology of different species of tsetse flies in a range of different countries with varying success in terms of population density and dispersal estimates (e.g. see [4, 7, 9-12]).

The best way to accurately estimate dispersal of a given population is isolation by distance. Isolation by distance is a common feature of population structure in which the genetic relatedness (or genetic distance) between individuals or subpopulations (see Box 1) is a decreasing function of the geographic distance separating them (the shorter the distance between them, the more similar they are). When such genetic distances are known, an isolation by distance model can be built and its significance tested (Box 2). If it is significant and if **effective population density** D_e can be estimated, then dispersal δ can be extracted from the model (Box 2).

In the last decade, several data sets have provided the opportunity for such inferences for different tsetse fly species in different countries in Africa: *Glossina palpalis*

52 *gambiensis* along the northern reaches [13] and the southern reaches of the Mouhoun
53 River [14] in Burkina-Faso (in one dimension=1D, see Box 2); *G. palpalis palpalis* in
54 Cameroon (two dimensions=2D, see Box 2) [11, 15]; *G. palpalis gambiensis* and *G.*
55 *tachinoides* in Southern Burkina Faso (2D) across river basins [9]; *G. tachinoides* in Ghana
56 (2D) [7]; *G. fuscipes fuscipes* from Uganda [12, 16], in Tanzania and Kenya [17] (2D); *G.*
57 *pallidipes* from Kenya Nguruman escarpment and Ruma [18] and from the Serengeti Park
58 Reserve in Tanzania (2D) [10]. This offered the opportunity to check if any relationship
59 existed between these different inferences of population density and dispersal. The
60 objective of the present paper is to discuss the evidence that tsetse dispersal may be
61 density dependent.
62
63 **Regression between dispersal and effective population density**
64 The detailed protocol for using or reanalyzing available data and computing the
65 necessary parameters are given in the supplementary file S1. Positive density
66 dependence dispersal occurs when dispersal increases with density, as expected if
67 individuals leave to escape overcrowding. Negative density dependent dispersal happens
68 when crowded populations can no longer accept any immigrants. To investigate evidence
69 for density-dependent dispersal in tsetse, we computed effective population densities as
70 the ratio of **effective population size** (N_e) to the average surface area occupied by a
71 population (Box 2). For the sake of homogeneity across studies, N_e was estimated using
72 **linkage disequilibrium** methods [19-22] and the surface area occupied by a
73 subpopulation corresponds to the surface of subsamples when available or the disc
74 defined by the minimum distance between the closest subsamples taken as the diameter
75 of a population (Box 2). Like for dispersal, we extracted **slope** estimates and 95%
76 **bootstrap** confidence intervals over loci (95%CI) (when available) to compute dispersal
77 (δ) as explained in Box 2. Because the relationship is exponential, we log transformed

both geographic and genetic distances to perform the **regression**. We kept only data with 2D population structure regressions [7, 9-11]. For some studies [12, 15-18], we needed to re-compute several statistics, as explained in detail in supplementary file S1. Only two studies where 1D dispersal inferences were available along the Mouhoun River in Burkina Faso [13, 14] were not included. According to Watts et al. [23], high mutation rates have a stronger effect on the accuracy of the method in linear habitats than in two-dimensional habitats and this may affect analyses of **riparian** systems. The fact that the metric in 1D is in m and in 2D in m² also makes comparisons difficult. Additionally, we wanted to check that the relationship was not due to a systematic bias in D_e and δ estimates by using independent estimates of D_e and census densities (see below). N_e can strongly depend on the reproductive system, on fluctuations in population sizes and/or generation overlap. This last point should be minimized if samples taken at intervals of not more than two months are used, as we recommend (Supplementary File S1). We also expect that N_e is strongly correlated with the census size of the population, otherwise all population genetics studies of tsetse flies would need to be called into question. Finally, a census of flies captured during the studies, when available, was also analysed and, likewise, was seen to be correlated with the real census of the corresponding tsetse populations [24].

The results of the regressions are presented in Figure 1, in 1a with effective population densities (D_e), and in 1b with census population densities (D_c). An exponential, tight and negative relationship can be seen between population density and dispersal. The regression explains most of the variance ($R^2 \approx 0.85$ for D_e; $R^2 \approx 0.86$ for D_c). Spearman's rank correlation testing with Rcmdr package [25, 26] for R [27] gave significant *p*-values (0.0056 and 0.0167 for D_e and D_c respectively).

Possible causes for negative density dependent dispersal

Negative density dependence appears to be much less frequent than positive density dependence (the intuitively expected sign of correlation), in particular in insects [28, 29]. "The occurrence of negative density dependent dispersal is in agreement with the ‘‘social fence’’ and related hypotheses which have been proposed in particular for small mammal populations, where movements are increasingly inhibited by aggression at high densities" [30]. Negative density dependence can result from **Allee effects** or because other factors, such as predation, interact with density to negatively influence dispersal [28]. In the case of insects, if there is a trade-off between wing development and reproductive capacity, dispersal may also be promoted by harsh local conditions, which can be correlated with low levels of population density, especially when dispersal is costly [31].

There are nevertheless several reported cases of negative density dependent dispersal. Experimental studies on insects have shown that juvenile hormone titer, which is influenced by the diet at earlier stages, can influence wing size and hence dispersal capacities [31]. Negative density dependence has also been reported in the northern pine processionary moth *Thaumetopoea pinivora* [29].

Dispersal is strongly density dependent in tsetse flies

Negative density dependent dispersal in tsetse populations has been known for some time (Box 3) but its intensity has not been measured to date. The exponential negative density dependence of dispersal observed in the present study raises several questions. The phenomenon implies that competition for space is very harsh and probably occurs mainly during feeding on the host (Box 3). When densities are high, all the sites are crowded and both hosts and local tsetse are accustomed to each other (Box 3). Naïve immigrants from remote sites are characterized by low feeding success, thus increasing their mortality. When population densities are low, densities may vary in both space and time, which renders immigration much easier and allows long-range immigration. In an

empty spot with no other tsetse competitors and naïve hosts, even exhausted immigrants from remote sites can safely settle. This pattern could also reflect favorable host densities. When densities of favorable hosts are low, tsetse flies need to keep moving to find a suitable blood meal.

Since control campaigns aim at considerably reducing tsetse densities, they may unleash dispersal and relatively rapid reinvasion of treated zones from neighboring or even remote sites, as may have occurred in Ghana [7] and certainly occurred in Ivory Coast [32]. Reinvasion of flies in areas depleted as a result of vector control has been well documented for quite some time and emphasizes the role played by fly movement in the development of control strategies [33]. Some evidence suggests strong density-dependence associated with this phenomenon: the bigger the decrease in the population, the higher the potential for reinvasion in a continuous tsetse belt [33]. These authors also report that Hargrove's models use density-dependent mortality, but not dispersal [34]. This limitation has already been reported and actually concerns all existing models [35]. Density-dependent dispersal was only recently incorporated in a tsetse population dynamics model, based on a sigmoidal density-dependent dispersal rate adapted for individuals competing for access to resources [36]. Such immigrations are dangerous for two reasons: i) because they can severely jeopardize the sustained success of control campaigns; ii) these new immigrants can bring with them pathogenic agents or more virulent strains that were not present before the campaign and were unable to invade the zone because of competition for space; and iii) these immigrants may replace the local population and adapt to local pathogens differently and exhibit new vectorial capacities compared to the former population. What is more, negative density-dependent dispersal may partially explain some observed failures. As underlined by Rogers and Randolph [32]: "Control programmes must recognize that the efficacy of population suppression may well be reduced at the low levels of density that need to be maintained in order to reduce the

rates of disease transmission, because at these low levels, tsetse may show entirely unexpected demographic vigor, due to the absence of normal density dependent constraints". If suitable host density is the only force driving the observed pattern, then control campaigns would not be expected to have any impact on dispersal.

Concluding Remarks

More data on isolation by distance slopes and effective population density estimates are needed to confirm the exact relationship found so far. One-dimensional isolation (along river courses) is particularly rare up to now. Most species we studied belong to the *palpalis* subgenus and one species (*G. pallidipes*) to the *morsitans* subgenus. The trends highlighted here probably apply to all tsetse flies but more studies are needed to generalize the trends we found to the whole genus. Studies should also focus on areas where tsetse control has been implemented. Considering that our results strongly support negative density-dependent dispersal, it will be crucial to evaluate the strength of the phenomenon using direct methods like mark-release-recapture in the field before and after control, for instance, combined with population genetics analyses. Alternatively, the prediction that neighboring flies will recolonize the treated area after the campaign ends can easily be assessed with sampling and genotyping before and after a control campaign in the treated area and its surroundings. If confirmed, and if one wants to avoid reinvasions and their associated issues (see Outstanding Questions), it will be even more important to find a way to isolate zones after treatment [4] and to sequentially treat neighboring areas harboring tsetse flies. Finally, several published results could not be included in the present review because of data unavailability. Data availability is now mandatory in several journals but we think this should be generalized.

Acknowledgements

We thank Winnie Okeyo, Norah Saarman, Oliver Manangwa, Trésor Tito Tanekou Melachio, Adalgisa Caccone and Chaz Hyseni for providing assistance with their data sets, which meant we could compute parameters that were missing in their original papers and considerably enriched our perception of density dependence in tsetse flies in the present review. We also thank Dr Raphaël Leblois for his advice concerning theoretical issues in isolation by distance.

188 **References**

189 1. Jamonneau, V. et al. (2012) Untreated human infections by *Trypanosoma brucei*
190 *gambiense* are not 100% Fatal. PLoS Negl Trop Dis 6 (6), e1691.
191 2. Van den Bossche, P. et al. (2010) A changing environment and the epidemiology of
192 tsetse-transmitted livestock trypanosomiasis. Trends Parasitol 26 (5), 236-243.
193 3. Solano, P. et al. (2010) How can tsetse population genetics contribute to African
194 trypanosomiasis control? Trends Parasitol 26 (5), 255-263.
195 4. Bouyer, J. et al. (2015) Mapping landscape friction to locate isolated tsetse populations
196 candidate for elimination. Proc Natl Acad Sci U S A 112 (47), 14575–14580.
197 5. Diall, O. et al. (2017) Developing a progressive control pathway for African animal
198 trypanosomosis. Trends Parasitol 33 (7), 499-509.
199 6. Vreysen, M.J.B. et al. (2013) Tsetse flies: Their biology and control using area-wide
200 integrated pest management approaches. J Invertebr Pathol 112, S15-S25.
201 7. Adam, Y. et al. (2014) Genetic comparison of *Glossina tachinoides* populations in three
202 river basins of the upper west region of Ghana and implications for tsetse control. Infect
203 Genet Evol 28, 588–595.
204 8. De Meeûs, T. et al. (2007) Population genetics and molecular epidemiology or how to
205 "débusquer la bête". Infect Genet Evol 7 (2), 308-332.
206 9. Kone, N. et al. (2011) Contrasting population structures of two vectors of African
207 trypanosomoses in Burkina Faso: consequences for control. PLoS Negl Trop Dis 5 (6),
208 e1217.
209 10. Manangwa, O. et al. (2019) Detecting Wahlund effects together with amplification
210 problems: cryptic species, null alleles and short allele dominance in Glossina pallidipes
211 populations from Tanzania. Mol Ecol Res 19, 757–772.
212 11. Mélachio, T., Tito, Tanekou et al. (2011) Population genetics of *Glossina palpalis*
213 *palpalis* from central African sleeping sickness foci. Parasites and Vectors 4, 140.

214 12. Opiro, R. et al. (2017) Genetic diversity and population structure of the tsetse fly
215 *Glossina fuscipes fuscipes* (Diptera: Glossinidae) in Northern Uganda: Implications for
216 vector control. PLoS Negl Trop Dis 11 (4), e0005485.
217 13. Bouyer, J. et al. (2009) Population sizes and dispersal pattern of tsetse flies: rolling on
218 the river? Mol Ecol 18, 2787-2797.
219 14. Bouyer, J. et al. (2010) Population structure of *Glossina palpalis gambiensis* (Diptera:
220 Glossinidae) between river basins in Burkina-Faso: consequences for area-wide integrated
221 pest management. Infect Genet Evol 10 (321-328).
222 15. Mélachio, T.T. et al. (2015) Effect of sampling methods, effective population size and
223 migration rate estimation in *Glossina palpalis palpalis* from Cameroon. Infection, Genetics
224 and Evolution 33, 150-157.
225 16. Hyseni, C. et al. (2012) The population structure of *Glossina fuscipes fuscipes* in the
226 Lake Victoria basin in Uganda: implications for vector control. Parasit Vect 5, 1-14.
227 17. Manangwa, O. et al. (2017) Genetic diversity of *Glossina fuscipes fuscipes* along the
228 shores of Lake Victoria in Tanzania and Kenya: implications for management. Parasit
229 Vectors 10 (1), 268.
230 18. Okeyo, W.A. et al. (2017) Temporal genetic differentiation in Glossina pallidipes tsetse
231 fly populations in Kenya. Parasit Vect 10 (1), 471.
232 19. Bartley, D. et al. (1992) Use of linkage disequilibrium data to estimate effective size of
233 hatchery and natural fish populations. Conserv. Biol. 6 (3), 365-375.
234 20. Do, C. et al. (2014) NeEstimator v2: re-implementation of software for the estimation of
235 contemporary effective population size (N_e) from genetic data. Mol Ecol Res 14 (1), 209-
236 214.
237 21. Peel, D. et al. (2013) Accounting for missing data in the estimation of contemporary
238 genetic effective population size (N_e). Mol Ecol Res 13, 243-253.

239 22. Waples, R.S. and Do, C. (2010) Linkage disequilibrium estimates of contemporary N_e
240 using highly variable genetic markers: a largely untapped resource for applied
241 conservation and evolution. Evol Appl 3, 244-262.
242 23. Watts, P.C. et al. (2007) Compatible genetic and ecological estimates of dispersal
243 rates in insect (*Coenagrion mercuriale*: Odonata: Zygoptera) populations: analysis of
244 'neighbourhood size' using a more precise estimator. Mol. Ecol. 16 (4), 737-51.
245 24. Barclay, H.J. and Hargrove, J.W. (2005) Probability models to facilitate a declaration of
246 pest-free status, with special reference to tsetse (Diptera: Glossinidae). Bull Entomol Res
247 95 (1), 1-11.
248 25. Fox, J. (2005) The R commander: a basic statistics graphical user interface to R. J Stat
249 Software 14 (9), 1–42.
250 26. Fox, J. (2007) Extending the R commander by "plug in" packages. R News 7 (3), 46–
251 52.
252 27. R-Core-Team, R: A Language and Environment for Statistical Computing, R
253 Foundation for Statistical Computing, Vienna, Austria, http://www.R-project.org, 2018.
254 28. Baines, C.B. et al. (2014) The interactive effects of competition and predation risk on
255 dispersal in an insect. Biol Let 10 (6), 20140287.
256 29. Ronnås, C. et al. (2011) Limited emigration from an outbreak of a forest pest insect.
257 Mol Ecol 20 (22), 4606-4617.
258 30. Matthysen, E. (2005) Density-dependent dispersal in birds and mammals. Ecography
259 28 (3), 403-416.
260 31. Kamioka, T. and Iwasa, Y. (2017) Evolution of density-dependent wing polymorphism
261 in insects. Evol. Ecol. Res. 18 (3), 335-348.
262 32. Rogers, D.J. and Randolph, S.E. (1984) A review of density-dependent processes in
263 tsetse populations. Insect Science and Its Application 5 (5), 397-402.

264 33. Gouteux, J.P. et al. (2001) A density-dependent model with reinvasion far estimating
265 tsetse fly populations (Diptera : Glossinidae) through trapping. Bull Entomol Res 91 (3),
266 177-183.
267 34. Hargrove, J.W. (2003) Tsetse eradication: sufficiency, necessity and desirability, DFID
268 Animal Health Programme, eds, Centre for Tropical Veterinary Medicine, University of
269 Edinburgh.
270 35. Peck, S.L. and Bouyer, J. (2012) Mathematical modeling, spatial complexity, and
271 critical decisions in tsetse control. J Econ Entomol 105 (5), 1477-1486.
272 36. Cecilia, H. et al. (2018) Environmental heterogeneity drives tsetse fly population
273 dynamics. bioRxiv, 493650.
274 37. Séré, M. et al. (2017) Comparison of different genetic distances to test isolation by
275 distance between populations. Heredity 119, 55-63.
276 38. Wright, S. (1965) The interpretation of population structure by F-statistics with special
277 regard to system of mating. Evolution 19, 395-420.
278 39. Weir, B.S. and Cockerham, C.C. (1984) Estimating F-statistics for the analysis of
279 population structure. Evolution 38, 1358-1370.
280 40. Rousset, F. (1997) Genetic differentiation and estimation of gene flow from *F*-statistics
281 under isolation by distance. Genetics 145 (4), 1219-1228.
282 41. Rousset, F. (2000) Genetic differentiation between individuals. J Evol Biol 13 (1), 58-
283 62.
284 42. Cavalli-Sforza, L.L. and Edwards, A.W.F. (1967) Phylogenetic analysis: model and
285 estimation procedures. Am J Hum Genet 19, 233-257.
286 43. Takezaki, N. and Nei, M. (1996) Genetic distances and reconstruction of phylogenetic
287 trees from microsatellite DNA. Genetics 144 (1), 389-399.
288 44. Balloux, F. (2004) Heterozygote excess in small populations and the heterozygote-
289 excess effective population size. Evolution 58 (9), 1891-900.

290 45. Vitalis, R. and Couvet, D. (2001) ESTIM 1.0: a computer program to infer population
291 parameters from one- and two-locus gene identity probabilities. Mol. Ecol. Notes 1 (4),
292 354-356.
293 46. Wang, J.L. and Whitlock, M.C. (2003) Estimating effective population size and
294 migration rates from genetic samples over space and time. Genetics 163 (1), 429-446.
295 47. Mantel, N. (1967) The detection of disease clustering and a generalized regression
296 approach. Cancer Res 27, 209-220.
297 48. WHO, Control and surveillance of human African trypanosomiasis: report of a WHO
298 expert committee, World Health Organization & WHO Expert Committee on the Control
299 and Surveillance of Human African Trypanosomiasis, World Health Organization.
300 http://www.who.int/iris/handle/10665/95732 Geneva, Switzerland, 2013.
301 49. Bouyer, J. et al. (2007) Learning influences host choice in tsetse. Biol Let 3 (2), 113-
302 116.
303 50. Vale, G.A. (1974) Direct observations on the responses of tsetse flies (Diptera,
304 Glossinidae) to hosts. Bull Entomol Res 64, 589-594.
305 51. Randolph, S.E. et al. (1984) Local variation in the population dynamics of *Glossina*
306 *palpalis palpalis* (Robineau-Desvoidy) (Diptera: Glossinidae). II. The effect of insecticidal
307 spray programmes. Bull Entomol Res 74, 425-438.
308 52. Rogers, D.J. et al. (1984) Local variation in the population dynamics of *Glossina*
309 *palpalis palpalis* (Robineau-Desvoidy) (Diptera: Glossinidae). I. Natural population
310 regulation. Bull Entomol Res 74, 403-423.
311
312
313

314 **Glossary**

315 **Allee effect**: A phenomenon where the individual fitness (survival and/or reproduction) is
316 positively correlated with population size or density.

317 **Alleles**: The different states of a locus or a gene (e.g. the three alleles of the ABO system
318 for blood groups).

319 **Bootstrap**: A randomization procedure where an item series (e.g. loci) are resampled with
320 replacement (the same item can be resampled several times) until the number of
321 items present in the raw data is reached. At each resampling, a statistic is
322 measured (e.g. genetic distance, see Box 1). The procedure is repeated a great
323 number of times (e.g. 5000), which produces a distribution of possible values for the
324 statistic. Excluding the 2.5% smallest and the 2.5% biggest values of the bootstrap
325 distribution provides the 95% bootstrap confidence interval of the statistic.

326 **Correlation coefficient**: A measure of the covariation between two variables.

327 **Effective population density**: The ratio of effective population size to the surface area
328 occupied by a given population.

329 **Effective population size**: Quantifies the rate at which a population loses its genetic
330 diversity. Indeed, the reciprocal of the effective size (1/Ne) gives the long-term
331 probability that two randomly sampled genes in the population are replicates (or
332 descend) from a single gene in the parental generation. It can be roughly defined as
333 the number of adults in a population that will leave a genetic signature to the next
334 generation. It is generally smaller than the census size N_c, except when
335 coalescence is delayed due to a particular system of mating (negative assortative
336 mating) or in very small dioecious populations. In any case, N_e and N_c must be
337 strongly positively correlated in most, if not all, situations.

338 **Immigration rate**: The proportion of individuals in a subpopulation that come from other
339 subpopulations of the total population.

Inbreeding: A concept describing how alleles, individuals or subpopulations can be related. Formally, it corresponds to the probability to randomly draw two identical alleles that are identical by descent, i.e. that come from a common ancestor.

Linkage disequilibrium: A measure of the statistical correlation between alleles at two or more loci. If in equilibrium, then the occurrence of alleles at two loci is simply equal to the product of corresponding allele frequencies in the population for the two loci. The main forces influencing linkage disequilibrium are reproductive systems, selection and genetic drift. It can be used to measure effective population size (small populations generate and maintain higher linkage disequilibria than bigger ones).

Locus: A specific segment of the genome, not necessarily a coding sequence (gene).

Neighborhood: The number of individuals connected through migration in an isolation by distance framework.

Parametric test: Statistical tests using population parameters (average and variance). If the constraints to apply such tests (e.g. normality of data, homogeneity of variances) are not met, non-parametric tests must be applied instead (e.g. rank tests or Mantel tests).

Population genetics: The study of the distribution of genetic variation in space and time and its evolution with random genetic drift, selection, mutation, migration etc.

Regression: A mathematical model explaining the relationship between a response variable and one (or several) explanatory variable(s).

Riparian: A riparian system describes the interface between land and a river or stream.

Slope: The parameter of a regression that describes how many units ordinates go up or down for each unit increase in the abscissa.

365 **Box 1: Measuring genetic distances**

366 Several genetic distances exist that are used in isolation by distance procedures
367 (see [37] for a more complete overview). The most popular, Wright's F_{ST} [38] is not a real
368 genetic distance but a measure of the effect of subdivision on **inbreeding**. The parametric
369 definition of this parameter is (e.g. [8]):

370 $$F_{\mathrm{ST}} = \frac{Q_{\mathrm{S}} - Q_{\mathrm{T}}}{1 - Q_{\mathrm{T}}}$$

371 where Q_S is the probability of identity between two **alleles** from two individuals of the same
372 subpopulation and Q_T is the probability of identity between two alleles from two
373 subpopulations of the total population. Its value varies between 0 (no subdivision) and 1
374 (all subsamples fixed for one or the other allele present, i.e. absolute subdivision). It is
375 mainly influenced by subpopulation sizes and **immigration rate** (or dispersal). This
376 parameter is estimated with Weir and Cockerham's unbiased estimator θ [39]. For isolation
377 by distance situations, it has been shown that the use of $F_{ST_R}=\theta/(1-\theta)$ is more useful
378 because it is linearly related to geographic distances without losing its relation to other
379 demographic parameters [40]. An equivalent measure between individuals a_r, and it's
380 unbiased estimator $\hat{a}$, was also designed by Rousset [41] as:

381 $$a_r = \frac{Q_w - Q_r}{1 - Q_w}$$

382 where Q_w is the probability of identity of two genes within an individual and Q_r is the
383 probability of identity of genes at (geographical) distance r. Another statistic $\hat{e}$ can be used
384 in case of very important **neighborhood** (see [23] for more details).

385 In other instances it may be more appropriate to use another genetic distance, e.g.
386 Cavalli-Sforza and Edward's chord distance [42]. D_{CSE} is more appropriate for tree
387 topology design [43] and more powerful in some cases of isolation by distance testing [37]:

388 $$D_{\mathrm{CSE}} = \frac{2}{r\pi} \sum_{j=1}^{r} \sqrt{2\left[\sum_{i=1}^{mj} \sqrt{x_{ij} y_{ij}}\right]}$$

where r is the number of loci, j the **locus** name (from 1 to r), i the allele name (from 1 to mj), mj the number of alleles at locus j, x_{ij} and y_{ij} are the frequencies of allele i at locus j for subpopulations x and y, respectively.

Box 2: Measuring and testing isolation by distance and parameter inferences

Isolation by distance is measured through a regression of geographic distances (explanatory variable) D_{Geo} between individuals or subsamples and corresponding genetic distances (response variable) D_{Genet}. Inferences follow particular models of regression, depending on whether the population structure occurs in one dimension (1D) or in two dimensions (2D) [23, 40, 41]. 1D structures correspond to shores, ecotones or river courses, as it is the case for *Glossina palpalis gambiensis* along forest galleries in savannas. 2D structures are more common, or at least more often reported. Three dimension models (dense forests or water columns for aquatic organisms) remain poorly explored. In 1D, the model is $D_{Genet}=a+b\times D_{Geo}$, where a is the intercept, b the slope of the regression and D_{Genet} stands for F_{ST_R} (for between subpopulations distances), $\hat{a}$ or $\hat{e}$ (between individuals). In 2D, the model is $D_{Genet}=a+b\times\ln(D_{Geo})$, where $\ln(D_{Geo})$ is the natural logarithm of D_{Geo}. The slope b is linked to the effective population density D_e and the average squared axial parent-offspring distance $\overline{\sigma^2}$ with a neighborhood estimated as $Nb=4D_e\times\overline{\sigma^2}=1/b$ in 1D and $Nb=4\pi D_e\times\overline{\sigma^2}=1/b$ in 2D [40].

The average surface (S) occupied by a subpopulation can be computed as the surface area occupied by the different traps used in a given survey site. If only one trap is available per site, the distance between the two closest sites (D_{min}) can be taken as the raw proxy of the distance between the centers of two neighboring subpopulations and hence as their diameter: $S=\pi\times(D_{min}/2)^2$. If the average effective population size N_e is computed with appropriate algorithms (e.g. see [20, 44-46]), then $D_e=N_e/S$, and a rough proxy of parent-offspring average distance (dispersal) can be computed as [37]:

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

In 2D, immigrants from neighboring subpopulations at each generation can be estimated as $N_e m=1/2\pi b$ [40].

For isolation by distance between individuals, $\hat{a}$ should be used instead of $\hat{e}$ when Nb<10000 in 1D or when Nb<50 in 2D [23].

Significance testing cannot be undertaken with a **parametric test** since distance measures are autocorrelated (paired comparisons). The significance of the slope can be tested by a bootstrap over loci based 95% confidence interval (95%CI). If 0 is not included in 95%CI, then the slope is significantly above 0. Otherwise, **a correlation coefficient** (e.g. Pearson) is computed between the two distance matrices and cells of one of those permuted a great number of times (Mantel test [47]). The *p*-value of the test is the proportion of time the randomized correlation was as big as or bigger than the observed one. If the 95%CI is not above 0, the Mantel test may be more powerful if the genetic distance used is D_{CSE}, at least for highly variable markers like microsatellite loci [37].

432 **Box 3: Tsetse fly atypical reproduction and density dependence**

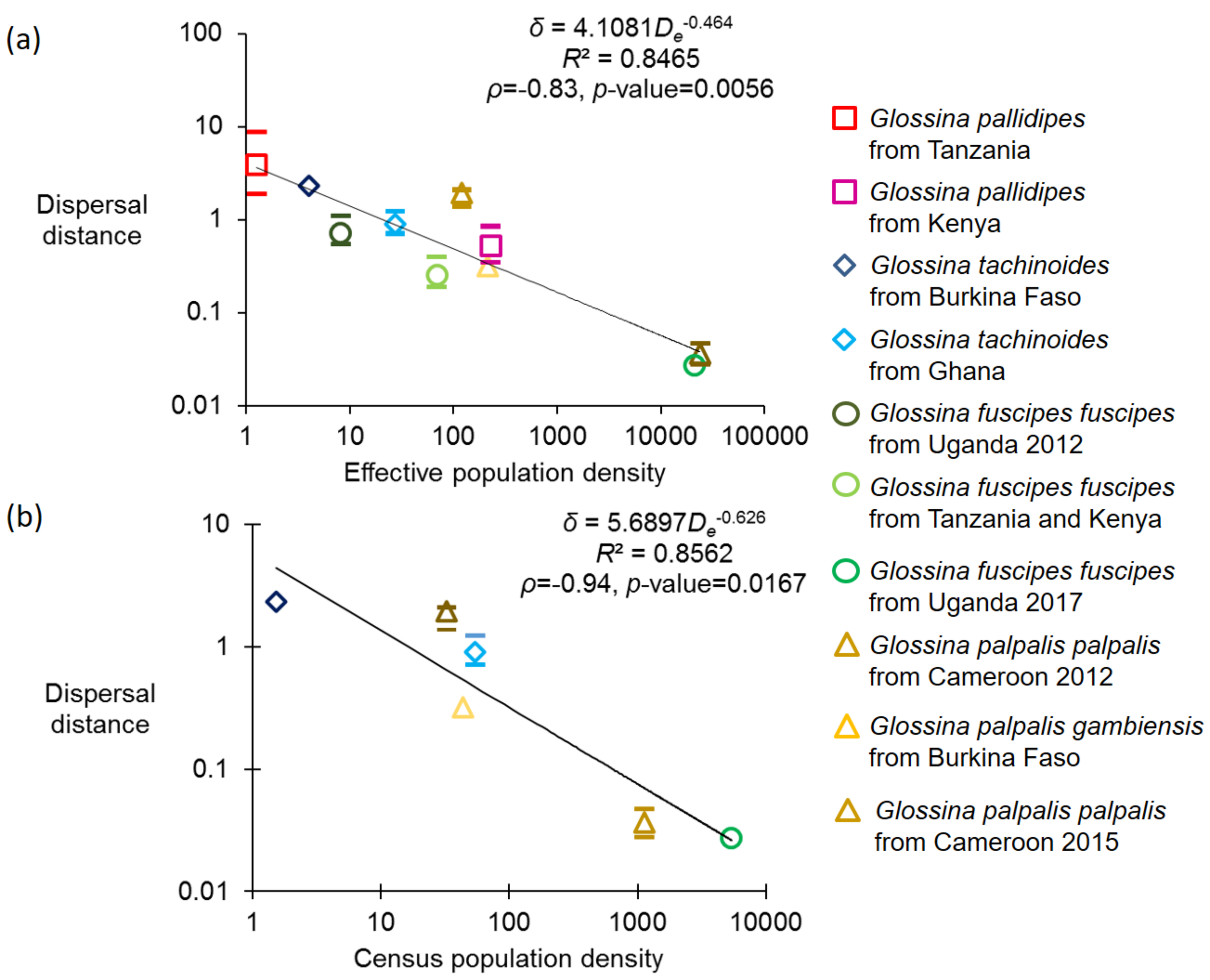
433 Female tsetse flies do not lay eggs but larviposit a single mature larva (3rd instar,
434 L3) in humid soil one at a time. The larva develops feeding from the uterine glands of the
435 mother (adenotrophic viviparity). After larviposition, the larva quickly burrows into the soil
436 surface for pupation. It was shown in *G. morsitans* that a larviposition pheromone is
437 deposited to attract other females to the same site, leading to a strong aggregation of
438 pupae [48]. The adult emerges 20–80 days later. Thus, in nature, each female produces
439 no more than 3–5 offspring during its total life. Lifespan is around 3 months for females, 2
440 months for males. As a result, the intrinsic rate of tsetse population growth is theoretically
441 low. Both females and males feed on vertebrate blood and are therefore both vectors.
442 Learning capacities of tsetse flies may increase their hunting efficiency with age and
443 encourage those returning to their first host [49].

444 Evidence for negative density dependence in tsetse flies has been reported in
445 several studies. Interactions between flies and the irritation of the host animal are
446 responsible for a decreased proportion of fed *G. morsitans morsitans* as the numbers of
447 tsetse flies arriving increase, and frequently disturbed spots might encourage tsetse flies to
448 leave without feeding, simultaneously increasing dispersal and associated mortality and
449 decreasing local density [32, 50]. Host irritation and tsetse learning are important
450 parameters driving the survival and dispersal of tsetse flies.

451 In dense tsetse populations, the populations are self-sustained and are
452 considerably reduced after insecticidal spray, but are recolonized from neighboring sites,
453 which takes several months to complete [51]. Alternatively, low-density sites are naturally
454 sustained by immigrants from neighboring sites and insecticidal treatments do not have as
455 much impact on the total population [32, 51, 52].

456

Figure 1: Strong negative density-dependent dispersal in tsetse fly populations. The graph shows the relationship between effective population densities D_e (individuals per km) and dispersal (δ, in km) (a) or between census population densities (densities of captured flies) (b) across different two-dimensional isolation by distance studies of different tsetse fly species in different African countries. The straight line corresponds to the power regression indicated in the graph with its determination coefficient R^2 and corresponding Spearman's coefficient (ρ) and associated p-value. Dashes indicate bootstrap over loci 95% confidence intervals. The same symbols indicate the same species. Scales were log-transformed for both axes.


(a)
$\delta = 4.1081D_e^{-0.464}$
$R^2 = 0.8465$
ρ=-0.83, p-value=0.0056
Dispersal distance
Effective population density
100
10
1
0.1
0.01
1
10
100
1000
10000
100000
(b)
$\delta = 5.6897D_e^{-0.626}$
$R^2 = 0.8562$
ρ=-0.94, p-value=0.0167
Dispersal distance
Census population density
10
1
0.1
0.01
1
10
100
1000
10000
Glossina pallidipes from Tanzania
Glossina pallidipes from Kenya
Glossina tachinoides from Burkina Faso
Glossina tachinoides from Ghana
Glossina fuscipes fuscipes from Uganda 2012
Glossina fuscipes fuscipes from Tanzania and Kenya
Glossina fuscipes fuscipes from Uganda 2017
Glossina palpalis palpalis from Cameroon 2012
Glossina palpalis gambiensis from Burkina Faso
Glossina palpalis palpalis from Cameroon 2015