



Published in final edited form as:

*Tsetse Trypanos Inf Q.* 2002 ; 25: 141–146.

## Genetic diversity and gene flow in *morsitans* group tsetse flies

**E.S. Krafsur**

Iowa State University, Ames, Iowa 50011-3222 USA. [ekrafsur@iastate.edu]

The question of how to achieve effective levels of tsetse fly control at financially and environmentally acceptable costs is perennial and contentious. Even though tsetse flies are slow to reproduce, populations seem to recover sooner or later after control measures are relaxed. A great capacity and propensity to disperse is said to be characteristic of tsetse flies, and many experts suggest that area-wide control measures and eradication are unobtainable for this reason alone. Others contend that area-wide methods, including the sterile insect technique, can be used successfully to achieve a high degree of control. Can a study of tsetse fly population genetics add anything to the ongoing debate? I believe it can. Here's why.

While the tsetse fly is traditionally shown on maps as being distributed in broad belts, within these belts tsetse fly populations are patchily distributed. Uninfested regions presumably consist of unsuitable or marginally suitable habitat. Even within infested patches, tsetse flies are aggregated into demes among which there may be varying degrees of isolation. Given application of effective control measures, how large an area must be treated to minimize re-invasion? John Hargrove suggests that very large areas are required, greater than 10,000 square kilometers (Hargrove, J.W. 2000. A theoretical study of the invasion of cleared areas by tsetse flies (Diptera: Glossinidae). *Bulletin of Entomological Research*, **90**, 201–209). Suppose, then, that a tsetse patch has been eliminated. What is the risk of invasion from nearby patches? Frontal advance of *morsitans* group flies has been shown to be of the order of 7 km per year and density-dependent responses might increase that value, but the distances between patches are too great to measure experimentally by mark, release, and recapture methods. Moreover, areas between patches are likely to be unsuitable for tsetse reproduction so that frontal advances fail and long range colonization of cleared areas is necessary.

In principle, we could measure gene frequencies of tsetse flies in two or more patches and derive indices of gene flow within and among them. This is the province of population genetics. Evolutionary insights may also come from such studies, as we shall see. Well developed theory teaches that the exchange of approximately one reproducing fly per generation, on average, is sufficient to prevent fixation of genetic differences between populations. Moreover, this 'critical' migration rate is virtually independent of population size! Whether the 'real' number is 0.5 or 2 is not biologically significant – in principle, numerically little gene flow can overcome local genetic differentiation. In theory, then, we have a powerful tool with which to examine the notion of biologically significant exchange of tsetse flies between patches.

Measurement of gene frequencies nowadays is a fairly routine affair. We can examine genetic diversity at loci that code for enzymes – so called isozymes and allozymes – by using starch, paper, or polyacrylamide electrophoresis, coupled with histochemical staining to demonstrate enzyme activity and allelic variation. The chief drawback is that of preserving enzyme activity in field-collected samples. The preferred method, using liquid nitrogen, is not always available, and airlines often refuse to accept shipments of this innocuous substance.

Another source of gene diversity is microsatellite loci. Microsatellites are short repetitive nucleotide sequences that vary in number, for example, [CA]<sub>n</sub>, where the number of repeats *n* varies. DNA can be extracted from rapidly dried or ethanol preserved flies, thereby making

sampling and transport to the laboratory easier. But much energy and skill is required to find microsatellite loci and to design and test the primers necessary to amplify them in the polymerase chain reaction (PCR).

Allozyme, isozyme, and microsatellite loci are present in two copies. Thus they can be used to measure genotypic frequencies. Genotypic frequencies, in turn, allow estimates to be made of departures from random mating within populations. Allele frequencies can be used to test hypotheses about the independence of two or more samples i.e. genetic differentiation.

PCR can be used to measure variation in mitochondrial genes. Mitochondria contain single copy loci and are maternally inherited. Variants do not recombine so the mitochondrial genome is inherited as a unit. Thus mitochondrial loci can be used to measure genetic differentiation of populations and maternal lines of descent because of the clonal pattern of inheritance. Moreover, mitochondrial variation is much more sensitive to population (demographic) events than is nuclear variation because of its inheritance pattern and the fact that its genes are represented by single copies, not double, as is the case for nuclear genes.

The foregoing kinds of genetic variation have been applied to some *morsitans* group tsetse flies. The sampling was carried out by Nigel Griffiths in The Gambia, Kenya, Zambia, Zimbabwe, Mozambique, and Namibia, Reg Allsopp in Botswana, Steve Mihok in Ethiopia, and Marc Vreysen, also in Ethiopia. The laboratory work and analysis was performed in my laboratory at Iowa State University.

## Genetic diversity

So, what do the data show? First, let's examine diversities, i.e. magnitudes of genetic variation. Later, I'll deal with gene flow. Diversities at mitochondrial loci estimate the probability that two randomly chosen tsetse flies have different haplotypes. For nuclear genes, diversity can be expressed in terms of the number of variants (alleles) at each locus and as heterozygosities – the proportion of loci with different (non-matching) alleles. Mitochondrial diversities, averaged over populations, were 41 percent in *G. pallidipes*, 35 percent in *G. m. morsitans*, and 43 percent in *G. m. submorsitans*, but only 22 percent in *G. m. centralis*. There were important contrasts in these diversities among regional populations that I'll return to later.

Allozyme data indicate that *morsitans* group flies are heterozygous at about 6 percent of their loci (heterozygosities at polymorphic loci, however, were about 25 percent). The 6 percent value compares with heterozygosities of about 18 percent in house flies and face flies (*Musca domestica* and *M. autumnalis*, respectively). The same methods show similarly high levels of diversity (heterozygosity) in numerous ladybird beetle and leaf beetle species (Chrysomelidae). Microsatellite diversities (heterozygosities) were much greater than the allozyme diversities, largely because they are probably untranscribed, subject to higher mutation rates, and do not respond to natural selection (i.e. they are 'selectively neutral'). Thus, in *G. pallidipes*, the number of alleles per locus was very much greater at polymorphic microsatellite loci (mean, 20.8 per locus) than at polymorphic allozyme loci (mean, 3 per locus). Microsatellite heterozygosities (diversities), averaged over populations and polymorphic loci, were 71 percent in *G. pallidipes*, 73 percent in *G. m. morsitans*, 81 percent in *G. m. submorsitans*, and 70 percent in *G. m. centralis*.

The magnitude of genetic diversity is important from evolutionary, ecological, and historical points of view. For example, theory shows a direct relationship between historical population sizes and diversity. Thus, the comparatively low diversities in tsetse flies are an indication that historical mean tsetse population sizes have been considerably less than those of many Diptera and Coleoptera, and is consistent with tsetse's low reproduction rates.

Low diversities can suggest historical ‘bottlenecks’ in population size, in which one or more successive generations undergo a great reduction in numbers. Bottlenecks in tsetse populations have been conjectured, as a consequence of rinderpest epizootics in the nineteenth and early twentieth centuries; indeed, such was demonstrated in Zimbabwe and claimed in Uganda (reviews can be found in Ford J., 1971, *The Role of the Trypanosomiasis in African Ecology*. Clarendon Press, Oxford, and Leak S.G.A., 1998, *Tsetse Biology and Ecology: Their Role in the Epidemiology and Control of Trypanosomiasis*. ILRI Nairobi/CABI). *Glossina morsitans centralis* in Botswana and Mamili National Park in Namibia showed a remarkable paucity of mitochondrial variation (only a 3 percent chance that two randomly chosen flies would have different mitochondrial haplotypes), populations having recovered from extensive control schemes in the Okavango region. Mitochondrial diversities in *G. m. submorsitans* were much less in The Gambia (26 percent) than in Ethiopia (84 percent); in Zimbabwe, *G. pallidipes* diversities were only 15 percent, but in Kenya and Ethiopia they averaged 54 percent. Microsatellite variation showed no hint of bottlenecks in *G. morsitans* s.l. but were reduced in Zimbabwean *G. pallidipes*. Allozyme diversities, on the other hand, seemed to be totally unaffected by putative bottlenecks. For example, Zimbabwean *G. pallidipes* showed slightly more heterozygosity than Kenyan populations even though mitochondrial and microsatellite variation was very much less. Thus, it seems that tsetse provide an example of ‘balancing selection’ acting on allozyme heterozygotes, thereby promoting diversity at allozyme loci.

Theory teaches that recovery from bottlenecks requires tens of thousands of generations, far more than the roughly 800 generations since the rinderpest epizootic. Therefore, the tsetse populations that we study today should still exhibit clear evidence of the earlier bottlenecks. The nature of the evidence includes disequilibrium between forces of mutation, migration, and genetic drift. So far, however, we are unable to reject null hypotheses that populations are in mutation-drift equilibrium. Larger sample sizes and the development of more sensitive statistical tests may, in future, allow more definitive investigations.

## Gene flow

Three independent lines of genetic evidence show abundant variation in *morsitans* group tsetse flies and this variation can be used to estimate gene flow within and among populations. How can this be done?

If gene flow is unrestricted, gene frequencies among populations will be statistically homogeneous. But what does it mean if they differ significantly? And why should they differ at all, if they do not respond to natural selection and have a common ancestry? The answer is that differences in gene frequencies arise because the laws of chance operate in the transmission of alleles from one generation to the next; thus, the smaller the breeding population the greater is the variance in gene frequencies. This is termed ‘genetic drift’ and is a major evolutionary force. The major result of drift is that gene frequencies of populations tend to diverge from each other in proportion to their isolation from each other. The isolation may be spatial, temporal, behavioural, premating, postmating, etc. Opposed to drift is immigration, and, as we have seen, numerically little exchange of reproducing migrants is effective in reversing the effects of drift.

We can do better than a simple test for differences by measuring the magnitudes of departures from random mating. The most commonly used index of departures from random mating is  $F$ , the so-called inbreeding coefficient. Now  $F$  may be viewed as a correlation of genes. In theory,  $F$  can take values from  $-1$  (matings only between unlike) to  $1$  (like mates only with like: completely inbred). It makes sense that flies in a particular location are more likely to mate with each other than with flies in another location, so there will be a measure of drift that leads to genetic differentiation. The classical estimate of drift (or genetic differentiation) among

populations is termed  $F_{ST}$ . The null hypothesis is  $F_{ST} = 0$ . Generally speaking,  $F_{ST}$  estimates of  $\geq 0.05$  are considered to indicate a biologically significant degree of differentiation. Genetic differentiation provides a continuous scale of reproductive isolation, in principle varying from zero to unity.

### Gene flow among conspecific tsetse fly populations

Estimates of  $F_{ST}$  based on allozyme, mitochondrial, and microsatellite variation were consistent. In *G. m. submorsitans*,  $F_{ST} \approx 0.17 - 0.35$  (depending on the method of analysis) among seven populations in The Gambia and Ethiopia, but  $F_{ST}$  was only 0.016 among samples within countries. We find that  $F_{ST} \approx 0.18$  among six *G. m. morsitans* populations, five of which originated in Zambia and Zimbabwe. More extensive sampling of *G. m. morsitans* is necessary to get an estimate of gene flow among fly belts. Among seven *G. m. centralis* populations extending from Tanzania to Botswana, it was found that  $F_{ST} \approx 0.19$ . A much greater estimate was obtained at mitochondrial loci, for which  $F_{ST} \approx 0.87$  (recall that mitochondrial loci are much more sensitive to demographic upheavals than are nuclear loci).

*Glossina pallidipes* showed, among 11 populations, a very high degree of differentiation at allozyme loci ( $F_{ST} \approx 0.19-0.24$ , depending on method of analysis), microsatellite ( $F_{ST} \approx 0.29$ ), and mitochondrial loci ( $F_{ST} \approx 0.51$ ). Study revealed that among northern populations, mitochondrial  $F_{ST} \approx 0.52$ , whereas  $F_{ST} \approx 0.28$  among southern populations. Microsatellite loci showed the same trends but allozyme loci did not. These data indicate that genetic drift at allozyme loci is greatly dampened and provides further evidence of balancing selection acting on allozyme loci (Krafsur, E.S., 2002, Population structure of the tsetse fly *Glossina pallidipes* estimated by allozyme, microsatellite, and mitochondrial gene diversities. *Insect Molecular Biology*, **11**, 37–45).

Estimates of  $F_{ST}$  can be converted to hypothetical estimates of the mean number of reproducing organisms exchanged per generation,  $Nm$ . The mathematical relationship is deceptively simple and entails many assumptions, but  $Nm$  can provide useful perspective by indicating the amount of gene flow in simple terms. The means, taken over all populations, were exchanges of 1.1 reproductives per generation among *G. m. morsitans*, 0.04–1.1 among *G. m. centralis* (based on microsatellite vs. mitochondrial loci, respectively), 0.5–1.2 among *G. m. submorsitans*, and 0.6–0.8 among *G. pallidipes*. Most of the foregoing estimates do not differ greatly from 1, that is, the ‘critical’ amount of gene flow below which genetic drift can proceed to fixation of different genotypes in different populations. Most insect species show  $Nm$  values that vary from five or ten to hundreds and thousands.

The high degree of genetic differentiation among *morsitans* group tsetse in general and *G. pallidipes* in particular is surprising when ecological and experimental data are considered. *Glossina pallidipes* is highly mobile – indeed, it is said to be the most vagile tsetse. I shall return to the contrast between ecological and genetical evidences later.

### Gene flow within populations

Genotypic frequencies, in terms of expected and observed heterozygosities, can be used to test hypothesis that matings are random within populations. For example, a deficiency of heterozygotes within a population is evidence that two or more demes of different gene frequencies were sampled; this can happen when samples from different locations are pooled. The chief estimator is  $F_{IS}$  (inbreeding  $F$  for individuals  $I$  in  $S$  (sub)populations). Sampling *morsitans* group tsetse indicated that matings were random within populations. Such data indicate no large scale immigration.

## What may we conclude?

First, with two important exceptions, the foregoing data on gene diversities do not indicate genetically detectable bottlenecks in population sizes. The exceptions concern *G. m. centralis* in Botswana and *G. pallidipes* in Zimbabwe where the historical record suggests that such bottlenecks occurred in *G. morsitans* s.l. and *G. pallidipes*, due to host animal reduction caused by the rinderpest outbreak. The same rinderpest epizootics swept through East and West Africa, but the genetic data suggest that tsetse populations were not so greatly affected as in southern Africa.

Second, tsetse populations are highly structured by genetic drift, with surprisingly little gene flow among them, a seemingly curious result considering the well known propensity for tsetse flies to disperse. Moreover, comparable work on other Diptera and many economically and medically important insects generally show much higher rates of gene flow than recorded in tsetse. The literature suggests frontal advances of tsetse populations averaging 5 to 7 km yearly.

How can the foregoing contradiction be resolved? Are the genetic data faulty or, more likely, is their interpretation simply wrong? Genetic analysis is built on Hardy-Weinberg assumptions which are rarely satisfied when dealing with natural populations and are certainly inapplicable to tsetse. If tsetse populations are recovering from earlier severe bottlenecks and disruptive population fragmentations, they would not be at mutation-drift equilibrium, and conclusions based on an assumption of equilibrium could be in error. Statistical tests for equilibrium, however, provide no evidence that the assumption is false.

Natural selection offers another rationalization for the apparent contradiction between ecological and genetical conclusions. In principle, tsetse far removed from their home territories may be at a significant reproductive disadvantage. Adaptation to local environments may be necessary and most immigrants may die without issue.

We also should consider scale. Ecological data pertain to distances in tens of kilometers. The genetic sampling summarized here involved distances ranging from tens to thousands of kilometers. Means taken over samples that vary so greatly in distance are apt to be misleading because the relationship between  $Nm$  and  $F_{ST}$  is nonlinear. But pairwise population estimates show correlations between genetic and geographical distances and they also generally confirm the low rates of gene flow. This is encouraging because low rates of gene flow support the concept of area-wide control, and predict low rates of re-colonization of habitat lost to tsetse. In principle, the sterile insect technique (SIT) involving sterile fly releases by aircraft may well prove to be more effective in reducing natural populations than are traps, targets, and targeted sprays because larger areas can be treated uniformly and efficaciously. And released, sterile males may turn out to be much more effective in finding small, hard-to-reach tsetse foci than are entomologists, rural sociologists, economists, and stakeholders everywhere.

Let's bear in mind that genetic and ecological research measure different things that are not strictly comparable; indeed, the contrast between ecological and genetic dispersals may be less in practice than in theory. Further research now underway should bring into better focus relationships between geographical distances, genetic distances, and, via physiological adaptation, the possibility of natural selection in explaining the breeding structure of tsetse flies.

As for area-wide control, in my opinion experimental sterile fly releases over large areas should first be made in order to learn something of interactions between released and wild flies in terms of sterile mating rates and target population responses to 'birth control'. Confounding treatments designed to maximize sterile to wild ratios should be avoided in such experiments.

Later, treatment by SIT of entire patches, as defined by satellite imagery and ground reconnaissance, might provide levels of control that would endure for many, many years.

## Acknowledgments

This paper of the Iowa Agriculture and Home Economics Experiment Station, Ames, Iowa, Project 6592, was supported by USPHS Grant AI-5245601, Hatch Act, and State of Iowa funds.