Molecular Ecology (2002) **11**, 2183–2187

SHORT COMMUNICATION

# Genetic structure of *Anopheles gambiae* (Diptera: Culicidae) in São Tomé and Príncipe (West Africa): implications for malaria control

J. PINTO,* M. J. DONNELLY,†‡ C. A. SOUSA,§ V. GIL,¶ C. FERREIRA,¶ N. ELISSA,** V. E. DO ROSÁRIO* and J. D. CHARLWOOD*††

*Centro de Malária e outras Doenças Tropicais, Instituto de Higiene e Medicina Tropical, Universidade Nova de Lisboa. Rua da Junqueira 96 1349–008 Lisbon, Portugal, †Division of Parasite and Vector Biology, Liverpool School of Tropical Medicine. Pembroke Place, Liverpool L3 5QA, UK, ‡Division of Parasitic Diseases, Centers for Disease Control and Prevention. MS; F22, 4770, Buford Highway, Chamblee GA, USA, §Unidade de Entomologia Médica, Instituto de Higiene e Medicina Tropical, Universidade Nova de Lisboa. Rua da Junqueira 96 1349–008 Lisbon, Portugal, ¶Centro Nacional de Endemias, Ministério de Saúde. Caixa Postal 218 São Tomé, Democratic Republic of São Tomé e Príncipe, **Centre International de Recherche Medicale de Franceville. BP 769 Franceville, Gabon*

**Abstract**

**The impact of a vector eradication programme, conducted in the 1980s, on *Anopheles gambiae* populations from the islands of São Tomé and Príncipe, was evaluated by microsatellite DNA analysis. Significant genetic differentiation was observed within and between the two islands and between the islands and a population from Gabon, suggesting a degree of isolation between them. Large estimates of long-term $N_e$ suggested that the control programme did not affect the effective population size of the vector. Heterozygosity tests were also not consistent with a recent bottleneck.**

*Keywords*: *Anopheles gambiae*, islands, malaria control, microsatellites, population structure

*Received 10 February 2002; revision received 14 June 2002; accepted 14 June 2002*

## Introduction

Population genetic studies of malaria vectors are an important means of predicting and assessing the success of control measures. Many conventional control strategies aim to reduce vector abundance. Estimates of effective population size ($N_e$) can be sensitive to strong reductions of population size. If population size varies between generations, $N_e$ will be the harmonic mean of the single-generation effective sizes and thus will approximate the lowest size (Nei 1987; Waples 1991). Such estimates can therefore provide an indication of how successful programmes really are in reducing vector numbers. Inferences on gene flow can also be useful predictors on the likelihood of the spread of insecticide-resistance genes or *Plasmodium* refractory genes introduced into natural populations by means of transgenic technology (Collins *et al*. 2000).

In 1980–82, a nation-wide malaria eradication programme was conducted in the archipelago of São Tomé and Príncipe, based on indoor spraying with DDT and mass drug distribution (chloroquine). During this period, both parasite prevalence and indoor vector densities were dramatically reduced (Ceita 1986). However, the disruption of the programme led to a severe epidemic. Today, malaria has reached or even exceeded pre-intervention levels (Pinto *et al*. 2000).

Using microsatellite data, we evaluated the possible long-term effect of the eradication campaign on the genetic structure of the vector populations in the islands. We also assessed genetic differentiation between populations from the islands and from Gabon, to infer levels of genetic isolation between them.

Correspondence: Virgílio E. do Rosário. Fax: + 351 21 362 24 58; E-mail: cmdt@esoterica.pt

††Present address: Danish Bilharziasis Laboratory. 1 Jaegersborg Alle, Charlottenlund DK 2920, Denmark.

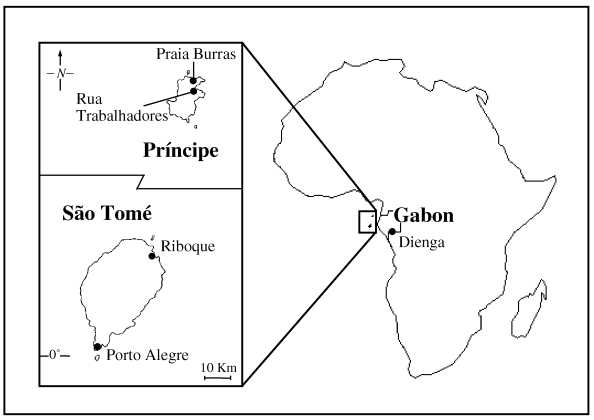


**Fig. 1** Collection sites in São Tomé and Príncipe and in Gabon.

## Materials and methods

The archipelago of São Tomé and Príncipe consists of two main volcanic islands located in the Gulf of Guinea, ≈ 240 km off the coast of Gabon (Fig. 1). The islands are mountainous, particularly in the central-southern parts that are characterized by dense rain forest and are almost uninhabited.

Female *Anopheles gambiae* were collected by human-baited landing catches in two locations on São Tomé and two locations on Príncipe, in March–May 1998 (Fig. 1). In Gabon, mosquitoes were collected in May 1999 in Dienga, a village with similar environmental characteristics to the islands. In all sites, the climate is equatorial with average annual temperature of 25 °C, humidity of 80% and rain fall over 1000 $mm^3$/year.

The 'Forest' cytoform of *Anopheles gambiae s.s.* is the only vector on the islands (Pinto *et al*. 2000). In a previous study, 444 of 445 specimens (99.8%) were identified as being monomorphic standard for inversions in chromosome 2 (Pinto *et al*. 2000). Unfortunately, no cytological information is available for Dienga.

Extraction of DNA from single mosquitoes, species identification by polymerase chain reaction and microsatellite genotyping were performed according to previously described protocols (Scott *et al*. 1993; Donnelly *et al*. 1999). Samples of 60 females from each population, were analysed for 13 microsatellite loci (AGXH7, AGXH8, AGXH25, AG2H147, AG2H175, AG2H197, AG2H1010, AG3H128, AG3H249, AG3H577, AG3H750, 33C1 and 45C1), all being located outside the major chromosomal inversions in *A. gambiae* (Lehmann *et al*. 1996; Zheng *et al*. 1996).

Tests of deviation from Hardy–Weinberg proportions at each locus and of linkage disequilibrium between pairs of loci were performed using GENEPOP v3.2. (Raymond & Rousset 1995).

Population differentiation was assessed by the fixation index $F_{ST}$ (Wright 1978) and the analogous estimator for microsatellite data $R_{ST}$ (Slatkin 1995a). Estimates of $F_{ST}$ were calculated according to Weir & Cockerham (1984) using FSTAT v2.9.1 (Goudet 1995). Estimates of $R_{ST}$ were produced using $R_{ST}$CALC (Goodman 1997), after converting allele length into repeat numbers as in Lehmann *et al*. (1996). Genotypic permutation tests were used to evaluate if $F_{ST}$ and $R_{ST}$ estimates differed significantly from zero.

Estimates of long-term $N_e$ for each population were derived from the expected heterozygosity at each microsatellite locus (Nei 1987) and should reflect events that occurred prior to the study period (Waples 1991; Lehmann *et al*. 1998). Calculations were based on the stepwise mutation model (SMM) and the infinite alleles model (IAM). These are opposite ends of the spectrum of mutation models and thus provide a robust range of estimates of $N_e$. As in previous studies, an average mutation rate of $10^{-4}$ was assumed (Lehmann *et al*. 1998).

Heterozygosity tests compare two estimates of expected heterozygosity, one based on allele frequencies ($H_E$) and the other based on the number of alleles and sample size ($H_{EQ}$) (Cornuet & Luikart 1996). If a population experiences a bottleneck, rare alleles will be lost and $H_{EQ}$ will decrease faster than $H_E$. This apparent excess of heterozygosity ($H_E > H_{EQ}$) is an indicator of a recent bottleneck event, whereas the converse ($H_E < H_{EQ}$) may indicate an expansion event. Wilcoxon signed-ranks tests were used to determine if there was a significant number of loci in which $H_E > H_{EQ}$. Estimates of $H_{EQ}$ were calculated under three mutation models, the SMM, the IAM and an intermediate two-phase model (TPM) with fractions of mutations greater than one repeat of 10, 20 and 30%. Tests were performed using BOTTLENECK 1.2.02 (Cornuet & Luikart 1996).

Whenever multiple tests were performed, the nominal significance level ($\alpha = 0.05$) was adjusted using the sequential Bonferroni procedure (Holm 1979).

## Results

A total of 300 female *Anopheles gambiae* were analysed. The number of alleles per locus ranged from 5 to 20. Two loci (AGXH7 and AGXH8) in Príncipe and one (AG3H577) in São Tomé (Porto Alegre) were monomorphic. Frequencies above 0.95 for the modal allele were found throughout the islands for AG2H1010 and AG3H577. Mean heterozygosity and number of alleles were not significantly different between and within island populations (Wilcoxon signed ranks tests, $P > 0.05$) but these were significantly less diverse than for the Dienga population ($P < 0.005$) (Table 1).

Within populations, significant heterozygote deficits were detected in 13 out of 60 tests at the within-population level (data not shown). These were clustered particularly in loci AG2H147 and AG2H175. None of the tests of linkage

**Table 1** Estimates of genetic diversity, long-term $N_e$ and heterozygosity tests at each population

| | | Riboque ($n$ = 13) | Porto Alegre ($n$ = 12) | Rua Trabalhadores ($n$ = 11) | Praia Burras ($n$ = 11) | Dienga ($n$ = 13) |
|---|---|---|---|---|---|---|
| | *Nall.* | 4.9 | 4.4 | 3.8 | 3.5 | 8.7 |
| | $H_E$ | 0.555 | 0.519 | 0.477 | 0.454 | 0.741 |
| $N_e$ | SMM | 10 393 | 8819 | 6966 | 6747 | 24 977 |
| | (95%CI) | (5504–15 282) | (3611–14 027) | (2600–11 332) | (2438–11 056) | (14 101–35 853) |
| | IAM | 4668 | 4072 | 3409 | 3308 | 8545 |
| | (95%CI) | (3047–6289) | (2265–5879) | (1801–5017) | (1658–4958) | (6571–10 519) |
| $H_E > H_{EQ}$ | SMM | $5^{NS}$ | $6^{NS}$ | $5^{NS}$ | $4^{NS}$ | 2† |
| | TPM (90%) | $7^{NS}$ | $7^{NS}$ | $5^{NS}$ | $4^{NS}$ | $4^{NS}$ |
| | TPM (80%) | $8^{NS}$ | $7^{NS}$ | $5^{NS}$ | $5^{NS}$ | $4^{NS}$ |
| | TPM (70%) | $10^{NS}$ | $7^{NS}$ | $5^{NS}$ | $6^{NS}$ | $9^{NS}$ |
| | IAM | 11* | 9* | $7^{NS}$ | $7^{NS}$ | 12*** |

$n$, number of polymorphic loci; *Nall.*, mean number of alleles per locus; $H_E$, mean unbiased estimate of expected heterozygosity (Nei 1987); $N_e$, mean estimates of long-term $N_e$ across loci. Estimates for X-linked loci were adjusted by a factor of 4/3, assuming a 1 : 1 sex ratio.
$H_E > H_{EQ}$, number of loci in which $H_E > H_{EQ}$.
Significance of one-tailed Wilcoxon signed-ranks tests for heterozygote excess: $^{NS}$nonsignificant; *$P < 0.05$; ***$P < 0.001$.
†Significant ($P < 0.001$) test for heterozygote deficit.
SMM, stepwise mutation model; IAM, infinite alleles model; TPM, two-phase mutation model.

**Table 2** Genetic differentiation estimates for each population pair

| | Riboque | Porto Alegre | Rua Trabalhadores | Praia Burras | Dienga |
|---|---|---|---|---|---|
| Riboque | | **0.025** | **0.176** | **0.208** | **0.080** |
| | | **(0.029)** | **(0.115)** | **(0.127)** | **(0.051)** |
| Porto Alegre | **0.037** | | **0.158** | **0.199** | **0.109** |
| | **(0.043)** | | **(0.109)** | **(0.133)** | **(0.089)** |
| Rua Trabalhadores | **0.167** | **0.188** | | 0.003 | **0.125** |
| | **(0.193)** | **(0.223)** | | (0.000) | **(0.127)** |
| Praia Burras | **0.181** | **0.196** | 0.002 | | **0.128** |
| | **(0.200)** | **(0.226)** | (0.000) | | **(0.127)** |
| Dienga | **0.118** | **0.161** | **0.237** | **0.250** | |
| | **(0.100)** | **(0.152)** | **(0.249)** | **(0.261)** | |

Above diagonal, $R_{ST}$; below diagonal, $F_{ST}$. Values in bold type indicate significance after adjustment by the sequential Bonferroni procedure. Values in parentheses are estimates excluding AG2H147 and AG2H175 that showed significant heterozygote deficits.

disequilibrium, which can indicate population subdivision, were significant. This suggests that samples were drawn from homogeneous, randomly mating populations and that heterozygote deficits are due to locus-specific events, such as null alleles or linkage with genes under selective pressure (Callen *et al*. 1993; Slatkin 1995b).

For both mutation models, $N_E$ estimates in the thousands were obtained for all populations (Table 1). Regardless of mutation model, $N_E$ was significantly higher in the Dienga population than in the island populations (Wilcoxon signed-ranks test, $P < 0.01$).

Further evidence for the absence of recent bottlenecks were given by the heterozygosity tests (Table 1). In the islands, deviations ($0.02 < P < 0.05$) from mutation-drift equilibrium, as a result of heterozygote excess, were found only in São Tomé under the less realistic IAM model. These deviations were not observed when analyses were performed under the TPM and SMM models. The population of Dienga showed, under the SMM, a significant deviation associated with heterozygote deficit. This suggests a recent population expansion, in agreement with previous findings in continental African *A. gambiae* populations (Donnelly *et al*. 2001).

Estimates of $F_{ST}$ and $R_{ST}$ showed similar patterns of population differentiation (Table 2). All pairwise comparisons revealed significant population differentiation, with the exception of the populations in Príncipe (Table 2). Even within São Tomé, mean estimates of $F_{ST}$ (0.037) and $R_{ST}$ (0.025) were significantly different from zero. Overall, $F_{ST}/R_{ST}$ were significant for most loci (data not shown),

indicating that differentiation is genome-wide. Similar patterns of differentiation were obtained when analysis was performed without loci AG2H147 and AG2H175, which exhibited the highest heterozygote deficits, suggesting that deviations from Hardy–Weinberg equilibrium had little effect on the estimates (Table 2).

## Discussion

This study highlights the value of population genetic analyses in assessing the effectiveness of malaria vector control programmes. Our data suggest that, in São Tomé and Príncipe, the spraying campaign of the 1980s did not significantly reduce the overall effective population size of the vector. The large values of $N_e$ observed for the island populations are within the range observed in other parts of Africa (Lehmann *et al*. 1998) and do not support recent bottleneck events. Further support was provided by the heterozygosity tests, in which populations appeared to be at equilibrium. Populations reach mutation-drift equilibrium if their gene frequencies remain stable for a period of $2–4N_e$ generations (Nei & Li 1976). If we conservatively take the lowest 95% CI limit of $N_e$ obtained (1658), assuming equilibrium and 12 generations/year, this population should have remained stable for at least 3316 generations, or around 276 years. Populations could have remained stable during the eradication campaign, possibly due to their ability to exploit outdoor resources (Sousa *et al*. 2001).

Significant genetic differentiation was found between the islands and the mainland, suggesting that gene flow between them is restricted. However, care should be taken regarding these data. A preliminary analysis on the non-panmictic M/S molecular forms of *Anopheles gambiae* (Favia *et al*. 1997), indicated that the M form is found in São Tomé and Príncipe and the S form is found in Dienga. This, coupled with differences in $N_e$ (Whitlock & McCauley 1999), may compromise the estimation of contemporary gene flow. The analysis of confirmed 'Forest' cytoform and M molecular form *A. gambiae*, preferably from the coast of the Guinean Gulf, should provide further clarification on the degree of genetic isolation of the São Tomé and Príncipe populations.

Whereas the two populations of Príncipe seemed to represent the same panmictic unit, significant differentiation was found between the two populations from São Tomé. The distance between the two sites (*c*. 40 km) is lower than the 50-km diameter of the minimum area associated with a deme of *A. gambiae* in western Kenya (Lehmann *et al*. 1997). Analysis of samples from intermediate locations of São Tomé will provide information on whether this differentiation reflects the presence of barriers to gene flow or if it is a consequence of geographical distance between two extremes of distribution. Should malaria control using genetically modified mosquitoes be attempted, the population structuring observed in São Tomé suggests that on this island, as in areas of continental Africa where microgeographic structuring is present, multiple releases, with more complex co-ordination, will be required. More detailed ecological and genetic studies are therefore needed.

## Acknowledgements

The authors are indebted to the people of São Tomé and Príncipe, to the entomology team of CNE for support in the field collections and the Ministry of Health (RD São Tomé and Príncipe) for logistics. Thanks go to T. Lehmann (CDC, Chamblee, USA), H. Townson (LSTM, Liverpool, UK) and V. Petrarca (Ist. Parassitologia, Rome, Italy) for technical training and scientific advice. This work was financed by INCO-DC/European Union (IC18CT960030 and IC18CT970244). J. Pinto and J. D. Charlwood were funded by PRAXIS XXI/FCT, Portugal (BD/15754/98 and BCC/7328/96).

## References

Callen DF, Thompson AD, Shen Y, *et al*. (1993) Incidence and origin of 'null' alleles in the (AC) n microsatellite markers. *American Journal of Human Genetics*, **52**, 922–927.

Ceita JGV (1986) Malaria in São Tomé and Príncipe. In: *Proceedings of the Conference on Malaria in Africa* (ed. Buck AA), pp. 142–155. American Institute of Biological Sciences/USAID, Washington DC.

Collins FH, Kamau L, Ranson HA, Vulule JM (2000) Molecular entomology and prospects for malaria control. *Bulletin of the World Health Organization*, **78**, 1412–1423.

Cornuet J-M, Luikart G (1996) Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics*, **144**, 2001–2014.

Donnelly MJ, Cuamba N, Charlwood JD, Collins FH, Townson H (1999) Population structure in the malaria vector, *Anopheles arabiensis* Patton, in East Africa. *Heredity*, **83**, 408–417.

Donnelly MJ, Licht M, Lehmann T (2001) Evidence for recent population expansion in the evolutionary history of the malaria vectors *Anopheles arabiensis* and *Anopheles gambiae*. *Molecular Biology and Evolution*, **18**, 1353–1364.

Favia G, della Torre A, Bagayoko M, *et al*. (1997) Molecular identification of sympatric chromosomal forms of *Anopheles gambiae* and further evidence of their reproductive isolation. *Insect Molecular Biology*, **6**, 377–383.

Goodman SJ (1997) $R_{ST}$ Calc: a collection of computer programs for calculating estimates of genetic differentiation from microsatellite data and determining their significance. *Molecular Ecology*, **6**, 881–885.

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

Holm S (1979) A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, **6**, 65–70.

Lehmann T, Besansky NJ, Hawley WA, Fahey TG, Kamau L, Collins FH (1997) Microgeographic structure of *Anopheles gambiae* based on mtDNA and microsatellite loci. *Molecular Ecology*, **6**, 243–253.

Lehmann T, Hawley WA, Grebert H, Collins FH (1998) The effective population size of *Anopheles gambiae* in Kenya: implications for population structure. *Molecular Biology and Evolution*, **15**, 264–276.

Lehmann T, Hawley WA, Kamau L, Fontenille D, Simard F, Collins FH (1996) Genetic differentiation of *Anopheles gambiae* populations from East and West Africa: comparison of microsatellite and allozyme loci. *Heredity*, **77**, 192–208.

Nei M, Li WH (1976) The transient distribution of allele frequencies under mutation pressure. *Genetical Research*, **28**, 205–214.

Nei M (1987) *Molecular Evolutionary Genetics*. Columbia University Press, New York.

Pinto J, Sousa CA, Gil V, *et al*. (2000) Malaria in São Tomé and Príncipe: parasite prevalences and vector densities. *Acta Tropica*, **76**, 185–193.

Raymond M, Rousset F (1995) GENEPOP (Version 1.2): population genetics software for exact tests and ecumenicism. *Journal of Heredity*, **86**, 248–249.

Scott JA, Brogdon WG, Collins FH (1993) Identification of single specimens of the *Anopheles gambiae* complex by the polymerase chain reaction. *American Journal of Tropical Medicine and Hygiene*, **49**, 520–529.

Slatkin M (1995a) A measure of population subdivision based on microsatellite allele frequencies. *Genetics*, **139**, 457–462.

Slatkin M (1995b) Hitchhiking and associative overdominance at a microsatellite locus. *Molecular Biology and Evolution*, **12**, 473–480.

Sousa CA, Pinto J, Almeida APG, Ferreira C, do Rosário VE, Charlwood JD (2001) Dogs as favoured hosts of *Anopheles gambiae* sensu stricto (Diptera: Culicidae) of São Tomé, West Africa. *Journal of Medical Entomology*, **38**, 122–125.

Waples RS (1991) Genetic methods for estimating effective population size of cetacean populations. In: *Genetic Ecology of Whales and Dolphins* (ed. Hoelzel AR), pp. 279–300. International Whaling Commission, Cambridge.

Weir BS, Cockerham CC (1984) Estimating *F*-statistics for the analysis of population structure. *Evolution*, **38**, 1358–1370.

Whitlock MC, McCauley DE (1999) Indirect measures of gene flow and migration: $F_{ST} \neq 1/(4Nm + 1)$. *Heredity*, **82**, 117–125.

Wright S (1978) *Evolution and the Genetics of Populations, Variability Among and Within Populations*, Vol. 1, 2nd edn. University of Chicago Press, Chicago.

Zheng L, Benedict MQ, Cornel AJ, Collins FH, Kafatos FC (1996) An integrated genetic map of the African human malaria vector mosquito, *Anopheles gambiae*. *Genetics*, **143**, 941–952.

---

J. Pinto is a PhD student investigating vector population genetics in relation to malaria control and epidemiology. M. J. Donnelly is an entomologist working on evolution of insect vectors. C. A. Sousa is an entomologist working in vector biology and taxonomy. V. Gil is a malariologist and C. Ferreira is a medical entomologist, both at STP. N. Elissa is a medical entomologist in Franceville, Gabon. V. E. Do Rosário is head of CMDT/IHMT. J. D. Charlwood was project co-ordinator in STP.

---