

METHODOLOGY

# Spatial pattern and genetic diversity estimates are linked in stochastic models of population differentiation

José Alexandre Felizola Diniz-Filho<sup>1</sup> and Mariana Pires de Campos Telles<sup>2</sup>

## Abstract

In the present study, we used both simulations and real data set analyses to show that, under stochastic processes of population differentiation, the concepts of spatial heterogeneity and spatial pattern overlap. In these processes, the proportion of variation among and within a population (measured by  $G_{ST}$  and  $1 - G_{ST}$ , respectively) is correlated with the slope and intercept of a Mantel's test relating genetic and geographic distances. Beyond the conceptual interest, the inspection of the relationship between population heterogeneity and spatial pattern can be used to test departures from stochasticity in the study of population differentiation.

## INTRODUCTION

Spatial analysis of genetic divergence among local populations has always played a central role in population genetics and evolutionary biology. Genetic divergence may increase with geographic distance both because environmental variation (and associated selective effects) become more heterogeneous in large distances or because the low migration rates do not constrain divergence by random drift.

The analysis of spatial population structure has been traditionally performed in an implicit way by Wright's (1965) F-statistics and more recently developed but related techniques, such as  $\phi$  and  $G_{ST}$  estimates (Nei, 1973; Weir, 1996). These statistics have been criticized because they do not furnish a detailed description of the spatial patterns of genetic divergence (Barbujani, 1987), furnishing only a general description of spatial heterogeneity among local populations. Explicit spatial methods, such as autocorrelation analyses (Sokal and Oden, 1978a,b; Sokal and Jacquez, 1991; Epperson, 1995a,b) and matrix comparison techniques (Mantel's tests - Smouse *et al.*, 1986; Manly, 1991; Thorpe *et al.*, 1996), have been used to overcome this difficulty and describe in more detail the spatial patterns in genetic data.

The differences between spatial heterogeneity and spatial pattern analyses in this sense (Sokal and Oden, 1978a,b; Sokal, 1986; Diniz-Filho, 1998) reflect, in fact, implicit and explicit spatial approaches to the analysis of population differentiation. More recently, Rousset (1997) showed that it would be possible to calculate pairwise  $F_{ST}$  statistics between populations and that a plot of  $F_{ST}/(1 -$

$F_{ST})$  against geographic distances produced slopes and intercepts that could furnish interpretations of demographic parameters.

In this communication, we show that the concepts of spatial heterogeneity and spatial pattern indeed overlap in stochastic models of population differentiation, such as Wright's (1943) isolation-by-distance or Kimura's stepping stone (Kimura and Weiss, 1964). Because they generate functional (exponential) relationships between genetic divergence and geographic distances, there is a close correspondence between measurements of spatial heterogeneity (e.g., as F-statistics and related estimates) and parameters of explicit spatial models.

## SIMULATION STUDY

We simulated 10 local populations randomly distributed in geographic space and, based on the geographic distance among them, we used the Ornstein-Uhlenbeck (O-U) stochastic process in the PDAP (*Phenotypic Diversity Analysis Program* - Díaz-Uriarte and Garland, 1996) program to simulate the evolution of five allele frequencies, keeping an exponential relationship between genetic divergence and geographic distances (for details, see Felsenstein, 1988; Hansen and Martins, 1996, and Telles, M.P.C. and Diniz-Filho, J.A.S., unpublished results).

We generated 50 data matrices containing 10 local populations and 5 allele frequencies, representing distinct loci, and for each one we estimated the  $G_{ST}$  statistics, according to Alfenas *et al.* (1991). Estimates of  $H_T$ ,  $H_S$ ,  $D_{ST}$  and  $G_{ST}$  for this multiple data set were performed using

<sup>1</sup>Departamento de Biologia Geral, ICB, Universidade Federal de Goiás. Caixa Postal 131, 74001-970 Goiânia, GO, Brasil.

Send correspondence to J.A.F.D.-F. E-mail: diniz@icb1.ufg.br

<sup>2</sup>Mestrado em Agronomia, Genética e Melhoramento de Plantas, Escola de Agronomia, Universidade Federal de Goiás, Goiânia, GO, Brasil.

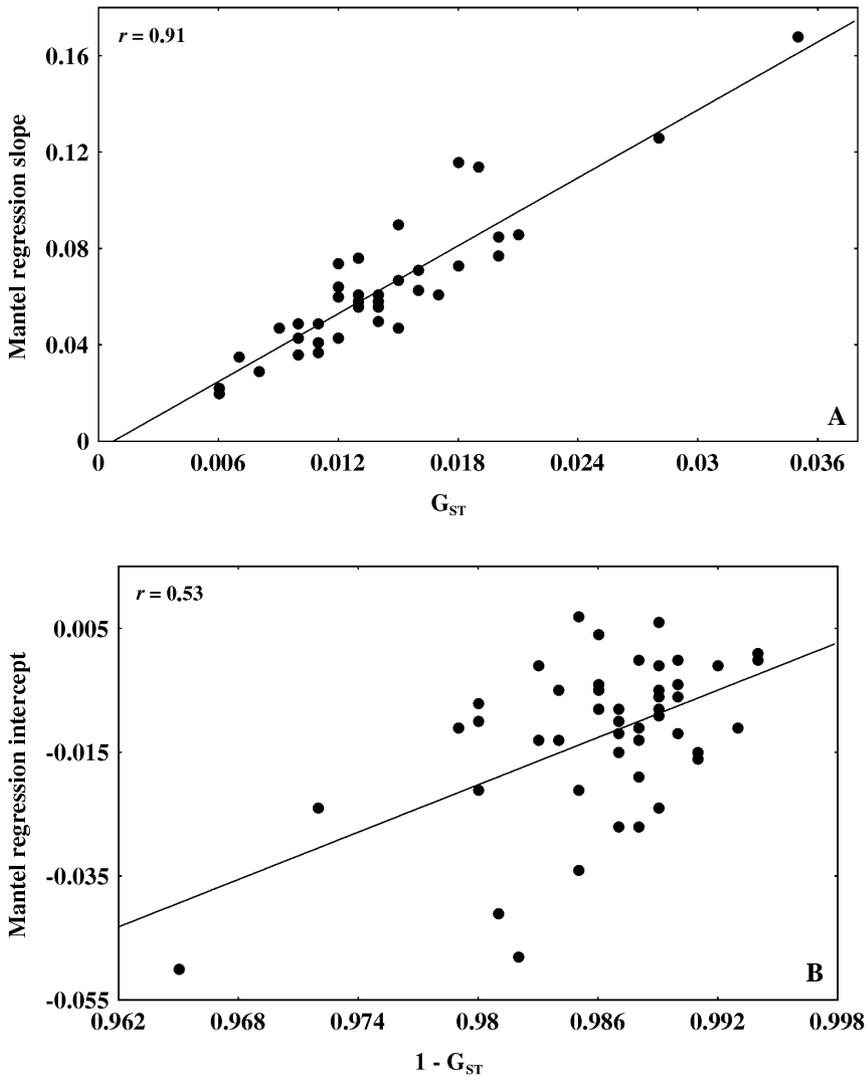
the GSTRUN program. For each data set, we also performed a Mantel's test (Smouse *et al.*, 1986; Manly, 1991) comparing Nei's (1972) pairwise genetic distances between populations with their geographic distances. Since we assume here that genetic divergence is a function of geographic distances, we also estimated the regression parameters of the linear model

$$N_{ij} = a + b \text{Dij} + \epsilon$$

where  $N_{ij}$  is the Nei's (1972) genetic distance between populations  $i$  and  $j$ ,  $\text{Dij}$  is the geographic distance between the same pair of populations and  $\epsilon$  is the residual term. The intercept of this matrix regression ( $a$ ) can be interpreted as the estimated genetic distance when the geographic distance is zero, which should be then related to the proportion of genetic variation within local populations ( $1 - G_{ST}$ ). Its slope ( $b$ ), in turn, must indicate the rate at which genetic divergence increases with geographic dis-

tance. We used both model I and II regression estimates of  $a$  and  $b$  (Sokal and Rohlf, 1995) in the analyses, calculated by the MATREG program. The empirical results for model II regression parameters (assuming that  $X$  is also defined with error) are much clearer, and only these will be shown here. This occurred probably because both genetic and geographic distances are estimated with an error related to the definition of patches of genetic similarity caused by the stochastic variation in the simulations. Both programs (GSTRUN and MATREG) were written in Basic language by one of us (J.A.F.D.-F.) especially for the simulation analyses and are available upon request.

As predicted, slopes and intercepts of Mantel's regressions are significantly correlated with the proportion of variation among ( $G_{ST}$ ) and within ( $1 - G_{ST}$ ) local populations, respectively (Figure 1). These linear patterns are indeed coherent with simple stochastic processes of ge-



**Figure 1** - Relationship between slope (A) and intercept (B) of the matrix regression of Nei's (1972) genetic distances against geographic distances with the estimated proportion of variation among ( $G_{ST}$ ) and within ( $1 - G_{ST}$ ) local populations, for 50 simulated data matrices.

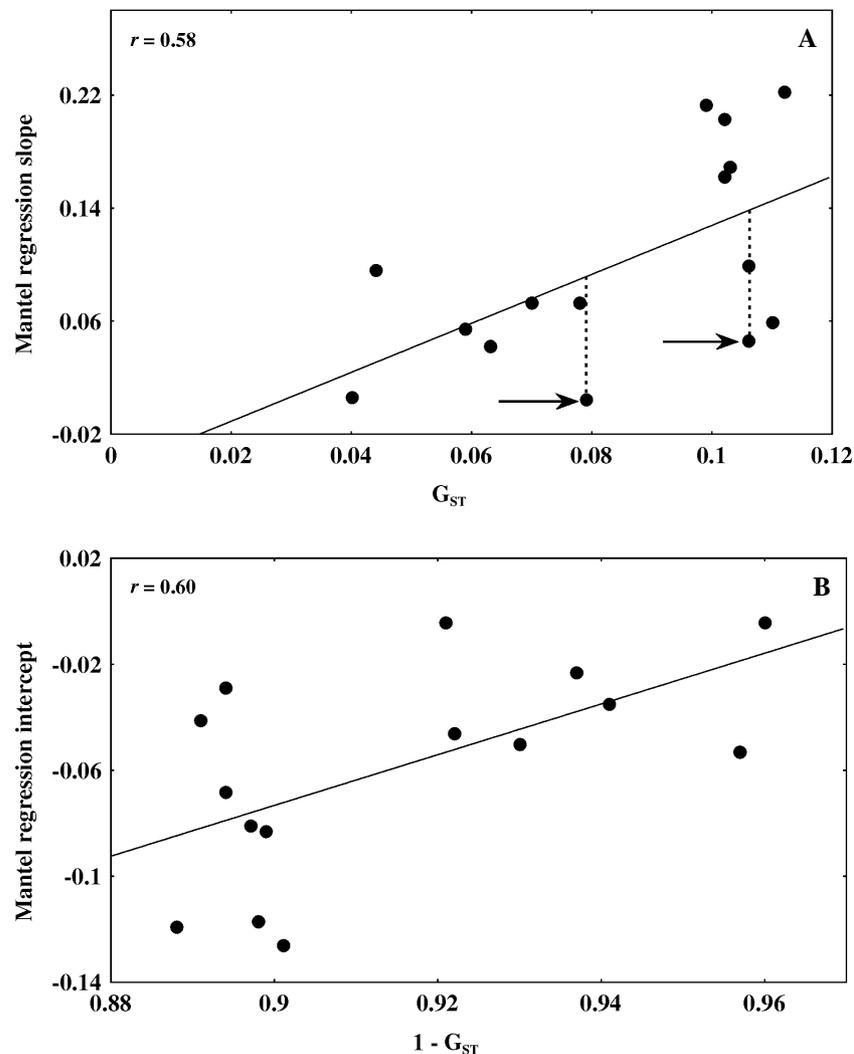
netic divergence. When genetic divergence among populations is high, this indicates that populations distributed in geographic space are more differentiated, and so slopes of Mantel's tests are higher because space possesses a more clear influence in divergence (Figure 1A). Also, because the stochastic divergence contains necessarily a spatial component, the absence of a spatial pattern in the data is an indication that most of the genetic divergence is within the local populations (Figure 1B).

#### REAL DATA SET

We tested the patterns discussed above using the data matrix provided by Sokal *et al.* (1986), with 15 allele frequencies for 15 bi-allelic loci estimated in 50 Yanomama villages. Both  $G_{ST}$  and Nei's (1972) genetic distances were estimated for each locus, and the  $a$  and  $b$  parameters of

Mantel's regression were correlated with  $1 - G_{ST}$  and  $G_{ST}$ , respectively. Results are similar to those obtained with the simulations (Figure 2), and this supports the previous conclusions that these significant relationships appear in stochastic genetic divergence. Sokal *et al.* (1986) discuss the hierarchical genetic structure among Yanomama, which seems to be more related to stochastic processes of divergence in space, caused by historical fission among villages through time.

No outliers are found in the two relationships of Figure 2, indicating, for example, that the spatial pattern of each locus (as measured by the Mantel's regression slope) is within the expected value for its magnitude of genetic heterogeneity. Indeed, the pairwise difference between residuals for the regressions in Figure 2 is not correlated ( $r = 0.03$  and  $0.02$ , as tested by a Mantel's test with 5000 permutations) with the pairwise difference between spa-



**Figure 2** - Relationship between slope (A) and intercept (B) of the matrix regression of Nei's (1972) genetic distances against geographic distances with the estimated proportion of variation among ( $G_{ST}$ ) and within ( $1 - G_{ST}$ ) local populations, for different loci estimated for 50 Yanomama villages (data from Sokal *et al.*, 1986). The arrows indicate the two loci that are more influenced by admixture with a neighbor tribe.

tial patterns in each locus, that were calculated using Manhattan distances between correlograms using Moran's  $I$  in 10 distance classes. Despite this, it is interesting to note that the two loci combining high genetic divergence and the low magnitude of the spatial pattern in Figure 2A (negative residuals -  $Gm^{as}$  and  $Acp^b$ ) are exactly those that depart more clearly from the fission history. These loci seem to be more influenced by admixture of the Ninam cluster of the Yanomama villages with an adjoining tribe, the Makiritare (Sokal *et al.*, 1986, pp. 273-274).

#### CONCLUDING REMARKS

Both simulations and real data set analyses show that the concepts of spatial heterogeneity and spatial pattern are not independent in common stochastic processes of genetic divergence generating hierarchical structure at population level. In principle, since this correspondence occurs in stochastic processes, the lack of relationship or outliers in the relationship would indicate departures of this stochasticity in the genetic divergence by the action of other microevolutionary processes.

For instance, the inspection of a plot of the measurements of spatial heterogeneity for each locus, such as  $G_{ST}$ , against parameters extracted from spatial pattern analyses (such as the slope of the Mantel's test) would be useful to detect which systems departure more from linear relationships. Selective pressures would explain this departure in such a specific system producing, at the same time, both complex relationships between divergence and space (which turns down the slope of Mantel's test) and strong heterogeneity among local populations.

#### ACKNOWLEDGMENTS

Research partially supported by FUNAPE (Master's degree grant to M.P.C.T) and CNPq (PQ-2B researcher fellowship to J.A.F.D.-F.).

#### RESUMO

Nessa comunicação, nós utilizamos análises de dados simulados e reais para demonstrar que, sob processos estocásticos de diferenciação entre populações, os conceitos de heterogeneidade espacial e padrão espacial são equivalentes. Nesses processos, a proporção de variação entre populações locais, estimada com base nas estatísticas  $F_{ST}$ ,  $G_{ST}$  ou  $\phi_p$ , está correlacionada com o coeficiente angular do teste de Mantel relacionando distâncias genéticas de Nei e distâncias geográficas. O intercepto dessa regressão matricial indica o valor da divergência genética quando a distância geográfica é zero, estando assim correlacionado com o valor de  $1 - G_{ST}$ . Além do interesse conceitual, a avaliação da relação entre medidas de heterogeneidade e padrão espacial pode

ser utilizada para testar desvios de processos estocásticos de divergência genética, comparando diferentes loci ou grupos de espécies.

#### REFERENCES

- Alfenas, A.C., Peters, I., Brune, W. and Passador, G.C. (1991). *Eletroforese de Proteínas e Isoenzimas de Fungos e Essências Florestais*. Universidade Federal de Viçosa, Viçosa, MG.
- Barbujani, G. (1987). Autocorrelation of gene frequencies under isolation-by-distance. *Genetics* 177: 772-782.
- Díaz-Uriarte, R. and Garland, T. (1996). Testing hypotheses of correlated evolution using phylogenetically independent contrasts: sensitivity to deviations from Brownian motion. *Syst. Biol.* 45: 27-47.
- Diniz-Filho, J.A.F. (1998). Análise de autocorrelação espacial: inferências microevolutivas e aplicações em genética de populações. In: Duarte, F.A.M. (ed.). *Série Monografias 6*. Sociedade Brasileira de Genética, Ribeirão Preto, pp. 83-145.
- Epperson, B.K. (1995a). Spatial distribution of genotypes under isolation by distance. *Genetics* 140: 1431-1440.
- Epperson, B.K. (1995b). Spatial distribution of two-locus genotypes under isolation by distance. *Genetics* 140: 365-375.
- Felsenstein, J. (1988). Phylogenies and quantitative characters. *Ann. Rev. Ecol. Syst.* 19: 445-471.
- Hansen, T.F. and Martins, E.P. (1996). Translating between microevolutionary process and macroevolutionary patterns: the correlation structure of interspecific data. *Evolution* 50: 1404-1417.
- Kimura, M. and Weiss, G. (1964). The stepping stone model of population structure and the decrease of genetic correlation with distance. *Genetics* 49: 561-576.
- Manly, B.F.J. (1991). *Randomization and Monte Carlo Methods in Biology*. Chapman and Hall, London.
- Nei, M. (1972). Genetic distance between populations. *Am. Nat.* 106: 283-292.
- Nei, M. (1973). Analysis of genetic diversity in subdivided populations. *Proc. Natl. Acad. Sci. USA* 70: 3321-3323.
- Rousset, F. (1997). Genetic differentiation and estimation of gene flow from F-statistics under isolation-by-distance. *Genetics* 145: 1219-1228.
- Smouse, P.E., Long, J.C. and Sokal, R.R. (1986). Multiple regression and correlation extensions of the Mantel test of matrix correspondence. *Syst. Zool.* 35: 627-632.
- Sokal, R.R. (1986). Spatial data analysis and historical processes. In: *Data Analysis and Informatics IV* (Diday *et al.*, eds). Science Publishers, Holland, pp. 29-43.
- Sokal, R.R. and Jacquez, G.M. (1991). Testing inferences about microevolutionary processes by means of spatial autocorrelation analysis. *Evolution* 45: 152-168.
- Sokal, R.R. and Oden, N.L. (1978a). Spatial autocorrelation in biology. 1. Methodology. *Biol. J. Linn. Soc.* 10: 199-228.
- Sokal, R.R. and Oden, N.L. (1978b). Spatial autocorrelation in biology. 2. Some biological implications and four applications of evolutionary and ecological interest. *Biol. J. Linn. Soc.* 10: 229-249.
- Sokal, R.R. and Rohlf, F.J. (1995). *Biometry*. 3rd edn. W.H. & Freeman, New York.
- Sokal, R.R., Smouse, P.E. and Neel, J.V. (1986). The genetic structure of a tribal population, the Yanomama Indians. XV. Patterns inferred by autocorrelation analysis. *Genetics* 114: 259-287.
- Thorpe, R.S., Black, H. and Malhotra, A. (1996). Matrix correspondence tests of the DNA phylogeny of the Tenerife lacertid elucidate both historical causes and morphological adaptation. *Syst. Biol.* 45: 335-343.
- Weir, B.S. (1996). *Genetic Data Analysis II*. Sinauer, Sunderland.
- Wright, S. (1943). Isolation by distance. *Genetics* 28: 114-138.
- Wright, S. (1965). The interpretation of population structure by F-statistics with special regard to mating systems. *Evolution* 19: 395-420.

(Received April 3, 2000)