



Mitochondrial DNA phylogeography of Brazilian populations of *Drosophila buzzatii*

Reinaldo A. de Brito¹, Maura H. Manfrin² and Fabio M. Sene³

¹Dept. Biology, Campus Box 1137. Washington University. St. Louis, USA.

²Dept. Biologia/FFCLRP, Universidade de São Paulo, Ribeirão Preto, SP, Brazil.

³Dept. Genética/FMRP, Universidade de São Paulo, Ribeirão Preto, SP, Brazil.

Abstract

The Argentinean Chaco region has been considered the center of origin of *Drosophila buzzatii* in South America because it contains most of the chromosomal polymorphism detected in natural populations. Two hypotheses have been put forward to explain the distribution of *D. buzzatii* in Brazil, one proposing that it has only recently passively colonized Brazil via human dispersal and the other suggesting that *D. buzzatii* has actively migrated to Brazil some time ago. Data from chromosomal inversions support recent colonization, whereas data from allozymes and mtDNA variation indicate that *D. buzzatii* has been in Brazil longer, favoring an active dispersal hypothesis. In our present work we analyzed data on 56 South American flies, mostly from Brazil, sequenced for the 5' end of the mtDNA COI gene. The combined use of many neutrality tests and phylogeographic methods (e.g. nested clade analysis) indicated high gene flow throughout most of the range of *D. buzzatii*, although significant population structure was still detected. The high nucleotide diversity in the Northeast region of Brazil and the results from the nested clade analysis suggest that *D. buzzatii* has been in Brazil longer than proposed by the passive dispersal hypothesis. Our data indicate that *D. buzzatii* has been distributed throughout Brazil and Argentina since the Quaternary, though more data from different localities and markers need to be gathered to determine how the occupation of South America by *D. buzzatii* has occurred.

Key words: *Drosophila buzzatii*, population structure, mitochondrial DNA, phylogeography, Brazil, climatic cycles.

Received: March 19, 2002; accepted: June 15, 2002.

Introduction

Drosophila buzzatii is a cactophilic member of the repleta group of fruit-flies which is found throughout Argentina, Bolivia, Paraguay and Brazil (Figueiredo and Sene, 1992) that has recently invaded Europe and Australia (Fontdevila *et al.*, 1981; Barker *et al.*, 1985). The most common host plants for *D. buzzatii* in South America are various prickly pear species, principally *Opuntia quimilo*, *Opuntia sulfurea* and *Opuntia vulgaris* in Argentina (Hasson *et al.*, 1992), and *O. vulgaris* and *Opuntia ficus-indica* in Brazil (Pereira *et al.*, 1983), although *D. buzzatii* has also been reared from other genera of cacti such as *Cereus* and *Trichocereus* in Argentina (Hasson *et al.*, 1992) and *Cereus* (Pereira *et al.*, 1983) and *Pilosocereus* (de Brito, 1999) in Brazil. Even though many of these species of cacti are widespread their local frequency and distribution vary substantially. There are two

main arid regions east of the Andes in South America which have a high density and diversity of cacti: the Chaco in Northern Argentina and east Bolivia and the Caatinga in Northeast Brazil, areas which are thought to have been connected during the Pleistocene (Ab'Saber, 1990; Prado and Gibbs, 1993). Cacti are not as frequent in other areas, being mostly found in rocky outcrops or in sandy substrates associated with dry deciduous forest.

Natural populations of *D. buzzatii* have been extensively studied in South America, Spain and Australia (Fontdevila *et al.*, 1981; Barker *et al.*, 1985; Ruiz and Santos, 1989; Halliburton and Barker, 1993; Betran *et al.*, 1998). Most of these studies have indicated little population substructure and high levels of gene flow (Thomas and Barker, 1990; Rossi *et al.*, 1996), the detection of private allozyme alleles in Northeastern Brazil (Barker *et al.*, 1985) being the lone exception to this pattern. Although *D. buzzatii* is generally very common in the Argentinean Chaco, where it occurs mainly in overgrazed areas where cacti are abundant (Ruiz and Fontdevila, 1982), its density in natural populations declines moving northeast into

Brazil (Figueiredo and Sene, 1992; Tidon-Sklorz *et al.*, 1994), and this, along with a high level of chromosomal inversion polymorphism in the Chaco, has led to the suggestion that *D. buzzatii* originated in the Argentinean Chaco (Carson and Wasserman, 1965; Vilela *et al.*, 1980; Fontdevila, 1989).

Two main hypotheses have been proposed for the dispersal of *D. buzzatii* into Brazil. The passive dispersal hypothesis proposes that *D. buzzatii* has only recently colonized Brazil following the use of *O. ficus-indica* by cattle ranchers and that this colonization has most probably been associated with some bottlenecks (Barker *et al.*, 1985) in a similar manner to what has happened in Spain and Australia. An alternative hypothesis, the active dispersal hypothesis, also considers that the dispersal of *D. buzzatii* into Brazil may have been recent but was not determined by human intervention, rather *D. buzzatii* first colonized Southern Brazil and then dispersed north (Figueiredo and Sene, 1992). As presented, this hypothesis offers too many different scenarios to allow rigorous scientific testing; therefore, a more strict definition is needed. Because it would be difficult to distinguish a recent (ecological time) active dispersal from a recent passive dispersal, we will consider that if a recent dispersal has happened, it would more likely be due to human intervention. Conversely, the active dispersal is favored if an ancient expansion is detected, either via range expansion or via isolation by distance.

The few studies done to date to test these hypotheses have been contradictory. Data on chromosomal inversions in *D. buzzatii* indicate the absence of chromosomal polymorphisms in the Brazilian Northeast (Figueiredo and Sene 1992), while populations in Southern Brazil present polymorphism levels lower than levels detected in the Iberian Peninsula, where colonization is known to be recent (Fontdevila, 1989). The lack of chromosomal polymorphism in the Northeast is more compatible with recent passive dispersal but the existence of private allozyme alleles in the Northeast (Barker *et al.*, 1985) is concordant with an active dispersal because it favors a more ancient presence of *D. buzzatii* in Brazil.

The present work aims at a better understanding of the population structure of *D. buzzatii* in Brazil. We were particularly interested to determine how and when *D. buzzatii* reached its current distribution in Brazil and whether the active or passive distribution hypothesis explains the observed distribution. To this end we sequenced part of the mitochondrial Cytochrome Oxidase I (mtCOI) gene of 60 individuals from natural populations as well as isofemale lines. Each of the hypotheses presented for the dispersal of *D. buzzatii* in Brazil have clear predictions for the current population structure of this species in Brazil, which are tested using nested clade analysis (Templeton *et al.*, 1995), analysis of molecular variance (AMOVA) and the combination of many distinct neutrality tests. The com-

bin analysis of these different methods allowed us to formulate the best hypothesis to explain the current distribution of *D. buzzatii* in South America.

Material and Methods

Samples

Flies were collected from 27 localities in South America where cacti were present (Figure 1), the sampling procedure being based on the collection of rotten cacti cladodes and banana-orange baits (Tidon-Sklorz *et al.*, 1994). We also used nine mass cultures, obtained from the Bowling Green *Drosophila* species Center (now stored at Tucson Stock Center, <http://stockcenter.arl.arizona.edu>), from eight localities in Argentina, Bolivia, Lebanon and Australia (Table I). For the neutrality tests we combined the localities into four regions defined with respect to ecological and geographical features such as available species of cacti, associated vegetation and geographical position (Table I).

Because *D. buzzatii* females are difficult to distinguish from other repleta species only males were sequenced in this study. Males trapped were kept in 70% ethanol for posterior DNA extraction and females were used to establish isofemale lines from which males were used for DNA extraction. The rotten cladodes were kept isolated in an aquarium where the newly emerged flies were collected. Because there is a high chance that flies emerging from the same cladode do not represent independent samples (Vilardi *et al.*, 1994), these flies were grouped to establish mass lines that were conservatively treated as representing a single individual.



Figure 1 - The geographical localities sampled in the present study. Locality numbers are from Table I.

Table 1 - Haplotypes sampled at various collection sites, listed by region and locality.

Region	# ^a	Site of collection		N	Haplotypes ^b	Coordinates
Northeast	1	Grão Mogol	MG	1	1	16°34'S, 42°54'W
	2	28 km W.M.Chapéu	BA	2	2; 3	11°34'S, 41°12'W
	3	10 km S.M.Chapéu	BA	1	4	11°38'S, 41°11'W
Southeast	27	Nova Ponte	MG	4	4; 4; 4; 24	19°08'S, 47°42'W
	8	Furnas	MG	1	4	20°37'S, 46°15'W
	9	Sertãozinho	SP	2	4; 6	21°10'S, 48°05'W
	15	Altinópolis	MG	2	13; 13	21°02'S, 47°55'W
	16	Faz. Sta. Carlota	SP	2	4; 14	21°18'S, 47°16'W
	17	Morro do Chapéu	SP	3	4; 15; 16	20°03'S, 47°23'W
	18	Estreito	SP	1	17	20°09'S, 47°16'W
	19	Bocaininha	SP	1	4	21°01'S, 47°17'W
	25	Bento Quirino	SP	1	22	21°28'S, 47°33'W
	6	Serra do Japi	SP	1	4	23°14'S, 46°57'W
	7	Atibaia	SP	1	25	22°56'S, 46°56'W
	20	Ibiuna	SP	1	4	23°44'S, 47°12'W
	21	Sorocaba	SP	2	4; 20	23°30'S, 47°34'W
	22	Brotas	SP	1	21	22°17'S, 47°57'W
	23	Itirapina	SP	3	4; 13; 22	22°16'S, 47°48'W
	24	Pau d' alho-Piracicaba	SP	2	4; 23	22°43'S, 47°41'W
	4	Rio Ligeiro	PR	1	4	23°37'S, 52°31'W
	14	Sengés	PR	5	4; 4; 12; 18; 19	24°07'S, 49°24'W
	26	Salto de Santa Rosa	PR	2	4; 4	23°38'S, 50°28'W
	South	10	Arroio Teixeira	RS	3	4; 10; 11
11		Capão da Canoa	RS	1	4	23°47'S, 50°10'W
12		Tramandaí	RS	1	9	29°55'S, 50°13'W
13		Guaritas	RS	3	4; 7; 8	30°45'S, 53°20'W
Chaco	5	San Juan	Arg	3	4; 4; 5	31°35'S, 68°31'W
	28	Cochabamba - BG ^c	Bol	2	4; 4	17°24'S, 66°09'W
	29	Cordoba - BG	Arg	1	4	31°24'S, 64°11'W
	30	San Ruiz - BG	Arg	1	4	33°18'S, 66°21'W
	33	San Raphael - BG	Arg	1	4	34°36'S, 68°18'W
Lebanon	31	Byblos - BG	Leb	1	4	34°07'N, 35°39'E
	32	Ain Anub - BG	Leb	1	4	33°50'N, 35°50'E
Australia	34	New South Wales - BG	Aus	1	4	33°50'S, 151°10'E
	35	Quensland - BG	Aus	1	4	22°00'S, 145°00'E

^aThe numbers refer to the localities shown in Figure 1.

^bHaplotypes of individuals sampled at that locality.

^cMass cultures from Bowling Green Species Center.

DNA was extracted from fresh, dried, or ethanol-preserved flies using QIAgen DNA extraction kits (QIAgen, Valencia, USA). PCR amplification was done using a 1.0 U of *Taq* polymerase and 1/100th U of *Pfu* polymerase to reduce errors during the process (Cline *et al.*, 1996). The primers used to amplify and sequence the 5'-end of the mitochondrial Cytochrome Oxidase I gene were 1460f 5'- atctatgcctaactcagcc - 3', 1519f 5'- atcataargatattggaact -3', 1924r 5' - taaagttgaagaattccggc

-3' and 2195r 5'- gatttttggtcacctgaagt -3' (the first and last are shorter modifications of the described by Simon *et al.*, 1994), with 35 cycles of 20 s at 94 °C, 30 s at 50 °C, and 1 min at 72 °C with a final 10 min extension step at 72° in a PE 9600 thermocycler. PCR products were purified using either a QIAgen PCR purification kit, geneclean (QIAgen, Valencia, USA) or electrophoresis in acrylamide gel followed by passive elution and ethanol precipitation. The templates were manually sequenced using the CircumVent

Cycle Sequencing kit (New England Biolabs Inc., Beverly, USA) and [S^{35}]dATP according to manufacturer's recommendations. The DNA sequences were input into PAUP* version 4.0b1 (Swofford, 1998) and visually aligned.

Neutrality tests

The mean number of nucleotide differences per site (π) the proportion of segregating sites among DNA sequences (S) and the number of haplotypes per sample (k) were estimated using the Arlequin program version 2.0 (Schneider *et al.*, 1997). Tests of neutrality were applied to the data from the four main geographical regions using the DnaSP program version 3.50 (Rozas and Rozas, 1999), the tests being the Tajima D test of selective neutrality (Tajima, 1989), the F^* and D^* tests (Fu and Li, 1993) and the F_s test (Fu, 1997). These tests do not all have the same power to detect departure from neutrality caused by different evolutionary forces such as hitchhiking, population size expansion, background selection or selective sweep (Fu, 1997), but their combined use allows an inferences to be made on the patterns of selection affecting a specific/selected region of DNA. The significance levels of these tests were estimated by the DnaSP program using a distribution of 10000 simulated populations generated by a neutral coalescent process.

Phylogeographic analysis

DNA sequences were analyzed following the procedures described in de Brito *et al.* (2002). Briefly, the phylogenetic relationships between *D. buzzatii* haplotypes were inferred in PAUP* 4.0b1 using 100 random taxa additions and heuristic searches. The most parsimonious trees were input into the MacClade program version 3.05 (Maddison and Maddison, 1992) where we implemented a methodology developed to deal with intraspecific phylogenies to create an unrooted haplotype network (Templeton *et al.*, 1987, 1988, 1992). A nested statistical design was defined according to Templeton *et al.* (1995) and used to investigate the history of *D. buzzatii* in Brazil. The Geodis program (Templeton, 1994, Posada *et al.*, 2000) was used to estimate the *clade distance* (D_c), which measures how geographically widespread the individuals from a clade are, and the *nested distance* (D_n), which measures the distance of a clade from the geographic center of their nesting category (Templeton, 1994). The combined analyses of these indices allow the distinction between historical factors, such as vicariance and range expansion, from ecological factors such as restricted gene flow which have shaped the distribution of a taxon (Templeton *et al.*, 1995). No *a priori* population hierarchies are assumed. We used Appendix I in Templeton (1998) to interpret the significance of the results obtained.

The results of the nested clade analysis were contrasted with the results of AMOVA (Excoffier *et al.*, 1992) set up in Arlequin version 1.1 (Schneider *et al.*, 1997), a technique which determines the amount of variation due to population substructure given an *a priori* set of population hierarchies using phiST. This index is like other F_{st} estimates but takes the number of mutations between haplotypes into consideration (Excoffier *et al.*, 1992). The *a priori* genetic structure was analyzed by a hierarchical analysis of variance, which partitioned the total variance into components due to intra- and inter- individual differences and interpopulation differences (Schneider *et al.*, 1997). The initial model tested considered that there was no hierarchy in the localities, while the second model tested performed a hierarchical AMOVA whereby localities were grouped into four regions. This subdivision was defined with respect to ecological and geographical features, such as available species of cacti, associated vegetation and geographical proximity to other fruit-fly populations (Table I). The first AMOVA provided an overall estimate of population differentiation similar to F_{st} following Slatkin (1991) across the range of *D. buzzatii*, whereas the second allowed the testing of a specific model of population substructure to explain patterns observed. The phiST parameter enabled the estimation of the average effective gene flow between regions (Nm) assuming the island model and migration-drift equilibrium (Hudson *et al.*, 1992).

Results

We sequenced the 5' end of the mitochondrial Cytochrome Oxidase I (mtCOI) gene (646 bases) for 60 individuals from 35 localities, 31 of them being in South America (Figure 1). The other four localities represent recent range expansion events in Lebanon and Australia (Table I). *D. buzzatii* was found at low densities outside Argentina and Southern Brazil, hence the low number of individuals sequenced per site. All individuals sequenced in Lebanon and Australia had the same haplotype, Haplotype 4, the most common one found in the New World. In South America, the Chaco samples showed the lowest level of polymorphism, followed by samples from Southeast Brazil, whereas South and Northeast Brazil had the highest mean number of nucleotide differences per site (π value), with Northeastern Brazil region having a level of polymorphism almost twice as high as the second most polymorphic area (South Brazil) notwithstanding its smaller sample size (Table II).

Neutrality tests

One caveat of neutrality tests in general is their assumption of an infinite-sites model with no recombination (Bertorelle and Slatkin, 1995). Departure from this assumption (*e.g.*, if there is mutation rate heterogeneity) may

Table II - Neutrality tests and mismatch distribution analyses.

Region	N	S (η)	π	Fu's Fs	Tajima's D	Fu and Li's F*	Fu and Li's D*
Northeast	4	10(10)	0.00826 \pm 0.00190	n.a. ^a	n.a. ^a	n.a. ^a	n.a. ^a
Southeast	36	17(19)	0.00224 \pm 0.00051	-8.555***	-2.156**	-2.542	-2.158
South	8	11(12)	0.00465 \pm 0.00130	-1.609	-1.770*	-2.109**	-1.940*
Chaco	8	1(1)	0.00039 \pm 0.00028	-0.182	-1.055	-1.204	-1.127
Brazil	48	34(37)	0.00371 \pm 0.00066	-21.677***	-2.203**	-3.500**	-3.347**
S America	56	35(38)	0.00330 \pm 0.00059	-27.957***	-2.515**	-4.304**	-4.103**

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

S, number of segregating sites; η , number of estimated mutations; π , nucleotide diversity.

^aNot applicable. Sample size was too small to allow inference.

cause rejection of neutrality regardless of whether other evolutionary forces are involved (Tajima, 1996). The infinite-sites model seems reasonably adequate to explain the polymorphism observed in mtCOI, since of the 36 polymorphic nucleotide positions only two had two substitutions inferred in the haplotype network, which is not significantly distinct from that expected by a Poisson distribution model (t-test; $p = 0.9985$). These two instances of homoplasy were observed in longer than average branches, where homoplasy is more likely than elsewhere in the cladogram.

All neutrality tests showed a significant deviation from the null hypothesis of neutrality when all localities in Brazil or in South America were grouped together. Table II shows these results and if each region is considered separately. The smaller sample size in Northeastern Brazil precludes inference on neutrality parameters. Of the other

three regions, only the Chaco showed no significant departure from neutrality by any of the tests. In South Brazil, Fs was the only test that failed to depart from neutrality, while in Southeast Brazil the Fs test and D test indicated departure from neutrality, while the F* and D* tests did not.

Phylogeographic analysis

The use of statistical parsimony (Templeton, 1998) allowed us to define a single unrooted haplotype network connecting all 25 haplotypes found. As can be seen in Figure 2, little structure is observed in this network, being mostly composed of a star phylogeny of short branches. The haplotype network defines the nesting comparisons relevant to investigating the existence of population substructure and gene flow (Templeton *et al.*, 1995). The nested clade analysis was performed considering only

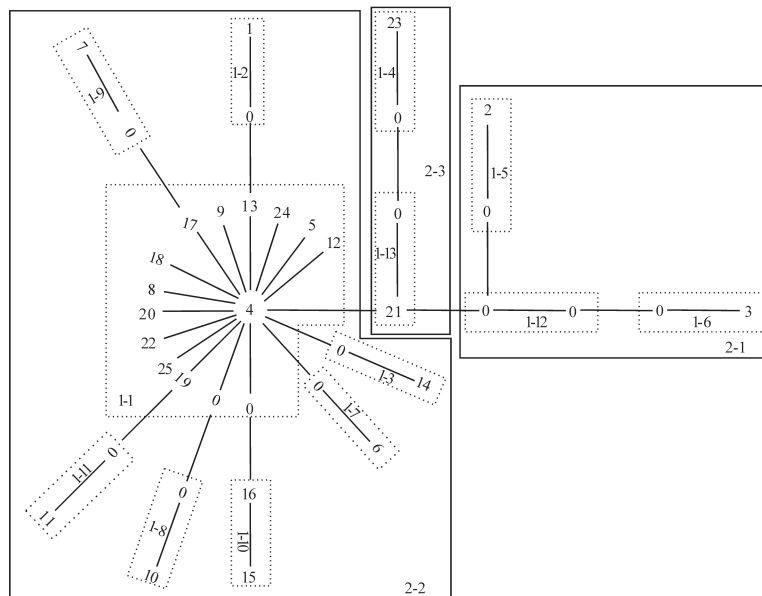


Figure 2 - The estimated 95% haplotype network and associated nested design for the mtCOI haplotypes observed in *Drosophila buzzatii*. Haplotypes are represented as numbers, and zeros indicate haplotype states that are necessary intermediates between sampled haplotypes but that were not observed in the sample. Each line represents a single mutational step that connects two haplotypes. Dashed boxes enclose one-step clades, which are referred to as '1-x', where x is the number that identifies the clade, while solid boxes enclose two-step clades ('2-x').

South American localities and alternatively considering all populations. The results (indicating two main events) were similar for both analyses, therefore only the former is shown in Table III.

According to the restricted gene flow with isolation-by-distance model, and following the coalescent theory, we expect that older haplotypes tend to be more internal in the haplotype network, and more geographically widespread than more recent ones, which would more often be rare and found in tips (Castelloe and Templeton, 1994). As predicted by this model, the internal Haplotype 4 is significantly more widespread than other haplotype tips connected to it. The other three significant results obtained in Clade 1-1 are also concordant with a pattern of restricted gene flow with isolation-by-distance (Table III). A second relevant event indicates an inconclusive outcome for Clade

2-3 in the Northeast, where we are unable to discriminate between fragmentation and isolation-by-distance.

Because a pattern of isolation-by-distance was the outcome of most of the nested analysis, it is particularly relevant to investigate whether the rate of gene flow has been sufficiently high to erase the association between genotype and geographical distribution. One way to look at this is to verify whether the average clade distances level off at the higher-level clades (Templeton *et al.*, 1995). In our data set we did not observe a leveling off of these distances (Figure 3).

We performed a hierarchical AMOVA to investigate population differentiation. Our first estimate considered all localities in an island model (without any specific hierarchy) and produced a F_{st} value of 0.0920, generating a migration rate parameter (Nm) of 4.94. In this partition, 9.2% of the variation can be ascribed to among population differences, which was not significantly different from zero

Table III - Results of the nested clade analysis.

Clade	D_c	D_n	Clade	D_c	D_n	Clade	D_c	D_n
Hap4 ^a	961.6 ^L	915.2 ^L						
Hap5	0	85.6 ^S						
Hap8	0	747.1						
Hap9	0	661.4						
Hap12	0	252.7						
Hap13	43.5 ^S	526.1						
Hap17	0	649.6	1-1	805.6	781.7			
Hap18	0	252.7						
Hap19	0	252.7						
Hap20	0	443.2						
Hap22	34.7	516.6						
Hap24	0	706.2						
Hap25	0	522.3				2-2	777.9	756.3
I-T	955.0 ^L	447.0 ^L						
Hap1	-	-	1-2	0	1193.3			
Hap14	-	-	1-3	0	501.8			
Hap6	-	-	1-7	0	446.6			
Hap10	-	-	1-8	0	658.5			
Hap7	-	-	1-9	0	786.5			
Hap15	-	-						
Hap16	-	-	1-10	0	586.6			
Hap11	-	-	1-11	0	658.5			
			I-T	805.1	91.4			
Hap23	-	-	1-4	0	37.5			
Hap21	-	-	1-13	2.2	18.9	2-1	25.1	319.6
			I-T	2.2	-18.6			
Hap2	-	-	-	-	-			
Hap3	-	-	-	-	-	2-3	0 ^S	1662.7
						I-T	-363.9	-889.9 ^S

^aShading indicates internal haplotypes or clades, while clear indicates tips.

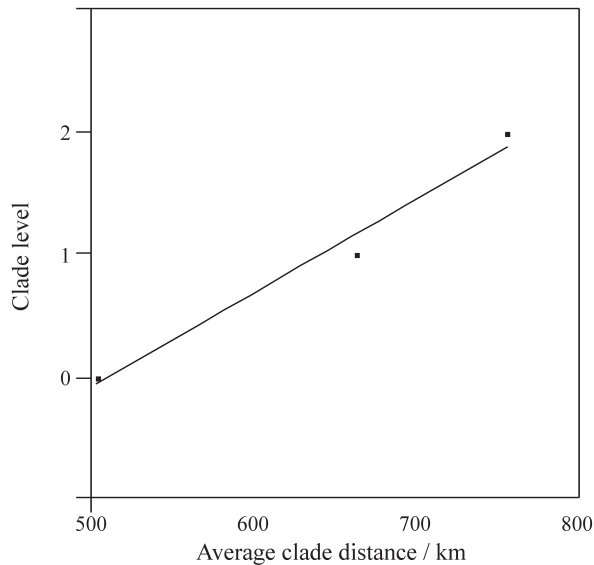


Figure 3 - Average nested clade distance (D_n) in Clade 2-2. Nested clade indicates average D_n for 0-step and 1-step clades nested in Clade 2-2 and D_n of Clade 2-2.

($p = 0.2630$). An AMOVA separating all localities into four main groups showed that 12.58% of the variation present was interpopulational (8.72% for differences between regions and 3.87 due to differences within populations among regions). A significant portion of this variation is due to interregional differences ($F_{CT} = 0.0872$, $p = 0.0244$), indicating that this model explains part of the genetic variation observed.

Discussion

D. buzzatii as a whole presented low π values, although they were within the boundaries of the values observed for mtDNA in other *Drosophila* species (Powell, 1997: 383). In our study, we observed very low levels of polymorphism in Chaco samples, about 15% of the polymorphism detected in Brazilian populations. Many of the individual Argentinean and Bolivian flies sequenced came from stocks maintained for a long time in laboratory culture, which may have caused a reduction in their levels of polymorphism. Nevertheless, the levels of polymorphism we observed in the Argentinean and Bolivian flies were still in the lower range of the levels observed in Argentina reported in the study by Rossi *et al.* (1996). Even if we accept the highest polymorphism values observed in the study of Rossi *et al.* (1996), Brazilian populations would still be, on average, 50% more polymorphic than Argentinean populations.

Neutrality tests

There are many reasons why a DNA region would fail to evolve according to the Fisher-Wright neutral model. If neutrality tests are significantly positive, it is an indication of balancing selection or population substructure. Most

mtDNA sequence studies to date, though, indicate that whenever the neutral model is rejected for a region it is due to a significantly negative test (Rand *et al.*, 1994; Nachman 1998; Nielsen and Weinreich, 1999). This may be caused by recent directional selection (selective sweep), a population bottleneck, recent populational growth, or background selection of slightly deleterious alleles (Tajima, 1989). Although the end result of these events may be similar (*i.e.* an excess of rare polymorphic sites) there are differences in the pattern of nucleotide substitution that allow us to distinguish between some of them (Fu, 1997; Wayne and Simonsen, 1998). If violation of neutrality is due to a population bottleneck, or recent population expansion, it should affect silent and replacement substitutions equally. The small number of replacement substitutions we detected limited our ability to use these tests to infer the forces involved. Because different evolutionary forces create distinct patterns of DNA polymorphisms, the many neutrality tests developed based on different polymorphism estimates do not have the same statistical power to detect events such as population expansion, genetic hitchhiking or background selection that cause this departure (Simonsen *et al.*, 1995; Fu, 1997). In fact, it has been suggested that because Fu's F_s and Tajima's D are more powerful than Fu and Li's D^* and F^* in detecting population growth and genetic hitchhiking (the opposite being true for background selection) the combined use of these tests may indicate the likely mechanism causing departure from neutrality (Fu, 1997).

Before we discuss the combined use of neutrality tests, it is important to point out one caveat with our sampling. Small numbers of individuals collected from many different localities tend to show more polymorphism than larger samples from a single locality (Rand *et al.*, 1994). This will reduce the ability to identify evolutionary forces acting at the population level since some haplotypes may be common at the local level but because of the reduced number of individuals sampled locally present themselves as singletons (Schmid *et al.*, 1999). Despite this, a hierarchical AMOVA performed on this data set indicated significant difference between the different regions, whereas little variation was ascribable to intra-regional differences. This indicates that our grouping procedures were justified since each region is not significantly heterogeneous.

It has been suggested that sample sizes of at least 50 are required in order to achieve reasonable power to detect some evolutionary forces (Simonsen *et al.*, 1995). Therefore, it is possible that even in populations that are experiencing some departure from neutrality, reduced sample size may prevent the detection of these forces, unless the forces are sufficiently strong to overcome the reduced power of the tests chosen (Wayne and Simonsen, 1998). Hence, the investigation of regions with smaller sample sizes may allow us to infer which events have been more important for that specific region, giving us an idea of the factors involved in the evolution of the whole population. Because

we were working with small sample sizes, we expected a reduced number of significant deviations from neutrality. Small sample sizes preclude a reasonable argument for neutrality if neutrality tests fail to be significant. On the other hand, when neutrality is significantly rejected in some regions, the small sample sizes may provide cogent evidence of strong evolutionary forces at work.

Our analysis of all individuals pooled indicated significant departure from neutrality. Many different evolutionary forces may be causing such departure. We looked into each separate region to see whether the same forces are acting in different geographical regions. No test detected significant departure from neutrality in the Chaco. In South Brazil, Fu's F_s test was the only test that failed to indicate significant departure from neutrality, suggesting that background selection may be the most important force affecting populations in the South, which is not surprising considering the high population densities observed in this region (Figueiredo and Sene, 1992). The fate of slightly deleterious mutations is influenced by variance in the effective population size (N_{ev}) and the mutation rate. In small populations, slightly deleterious mutations behave virtually as neutral if $N_{ev}s < 1$, whereas large populations have a higher ability to purge these alleles (Ohta, 1992). Southeast Brazil comprises the largest regional sample size analyzed. Fu's F_s and Tajima D tests indicate significant departure from neutrality, while Fu and Li's D^* and F^* are not significant. Because statistical power depends on sample size, we also investigated the outcome of neutrality tests if Southeast Brazil is subdivided into three sub-regions, with approximately the same number of individuals present in the other regions studied. When we do this, F_s is the only test to reject neutrality in all three sub-regions (data not shown). The F_s and D tests are more powerful in detecting population expansion and hitchhiking (Fu, 1997), which, from a population genetics standpoint, generate basically the same pattern when only one marker is investigated (Slatkin and Hudson, 1991).

Przeworski *et al.*, (1999) determined that the evolution of neutral sites will not be affected by weak purifying selection on selected markers even if there is no recombination between them, but background selection on many strongly deleterious mutations may reduce variation at linked neutral sites. Tests such as Tajima's D and Fu and Li's F^* and D^* have little power to detect single-site purifying selection if linked neutral loci are examined, unless there are very strong selection pressures (Przeworski *et al.*, 1999). Because most of the substitutions we detected were silent, there is little indication that the strongly negative results of the neutrality tests are the result of directional or purifying selection. This suggests that the significant F_s estimates for populations from central Brazil are an indication of population size expansion.

Phylogeographical analysis

Our analysis of haplotypic data suggests homogeneity across Brazilian populations of *D. buzzatii*, the haplotype network indicating little differentiation between individuals separated in some cases by more than 3000 km, as preliminary data has suggested (Manfrin *et al.*, 2000). Despite this reduced heterogeneity, an AMOVA test detected significant population substructure while the nested clade analysis indicated significantly restricted gene flow. These results show that levels of gene flow are not sufficiently high to erase historical information on geographic origin.

Most of the gene flow we detected fits an isolation-by-distance model associated with Clade 1-1. The Nm value estimated from the pooled data was well above 1, which is considered to be strong enough to prevent evolutionary divergence (Wright, 1951). Since this Nm estimate comes from a small/reduced sample of a single molecular mtDNA marker it should be considered a rough estimate to be corroborated by other markers. Furthermore, this Nm estimate is based upon an island model of gene flow and, therefore, any deviation (*e.g.* such as isolation by distance or vicariance) from this model undermines its biological meaning.

High levels of gene flow between Argentinean populations of *D. buzzatii* have been described using restriction digestion patterns of mtDNA in a study which also indicated a complete lack of variation in 88 individuals belonging to 5 colonized populations from the Iberian Peninsula (Rossi *et al.*, 1996). A similar pattern was observed in our study, even though only four individual flies from Lebanon and Australia were sequenced. Concordant with what we detected for our colonized populations, the most common haplotype in South America is the only haplotype present in the Iberian Peninsula (Rossi *et al.*, 1996).

The nested clade analysis detected another significant event in Clade 2-3, which represents the separation between Northeast Brazil and elsewhere. Although range expansion can be ruled out, the data is insufficient to distinguish between fragmentation or isolation-by-distance at this level of the analysis.

The use of nested clade analysis enables us to detect range expansions, but there are some limitations on the ability of this test to detect such expansions. In a study of 13 known range expansions (Templeton, 1998), 12 were correctly inferred by the nested clade analysis. Interestingly, the only range expansion that the nested analysis failed to detect involved the occupation of Iberian Peninsula by *D. buzzatii*. This was because that event occurred only very recently (Fontdevila *et al.*, 1981; Barker *et al.*, 1985) and colonization most likely involved some population bottleneck or selective sweep because all *D. buzzatii* individuals collected in Europe and Australia have only the most common haplotype observed in South America. Though Argentinean and Brazilian populations did not show high levels of polymorphism, the

levels detected suggest that if a range expansion had occurred it would have been detected by nested analysis, unless current levels of gene flow were sufficiently high to erase correlation between geographical location and phylogenetic relationship. The significant results which we obtained at two different levels of the cladogram and the presence of a correlation between clade level and geographical range size (Figure 3) indicate that gene flow is not sufficiently strong enough to erase phylogenetic information on geographical distribution, a conclusion which is also supported by the detection of significant substructure in the AMOVA.

Combined analysis of neutrality and phylogeographic methods

To test the two hypotheses (passive versus active dispersal) put forward to explain occupation of Brazil by *D. buzzatii* we must consider what is expected if Brazilian populations represent a northward range expansion of Argentinean populations.

In a range expansion, often only some of the haplotypes present in the original distribution will expand to the new area, making the haplotypes in the expanding population more widespread than the other haplotypes which remained in the source population (Cann *et al.*, 1987), a pattern confirmed by Templeton (1998). A range expansion entails a sub-sampling of the diversity of the source population and, as a consequence, the expanding populations are not expected to present higher nucleotide diversity than the source population unless they have also experienced a rapid expansion in population size, something which may have happened in Southeast Brazil but is unlikely in the Northeast (de Brito, 1999). Therefore, if *D. buzzatii* had experienced a recent northward expansion from Argentina we would expect genetic polymorphism to taper out as we move away from the source population into Brazil, but our results indicate that this is not what happened. In fact we found that *D. buzzatii* nucleotide diversity in the Northeast Brazilian is the highest of all regions (almost twice as high as in Southern Brazil) and that the most divergent haplotypes occurred in the Northeast, both these observations being in conflict with a recent northward expansion. Furthermore, nested clade analysis indicates that occupation of Northeastern Brazil was not achieved through a range expansion, either recent or ancient.

The different neutrality tests indicate that populations in Southeastern Brazil may have experienced population size expansion or selective sweep, whereas background selection in the South of Brazil indicates large population sizes.

The combination of data from different methods suggests that *D. buzzatii* was distributed through Brazil and Argentina over much of the time period marked by the coalescence of present day mtDNA variation to its most recent common ancestor. Populations in Argentina and in

Northeastern Brazil became separated either by restricted gene flow with isolation-by-distance or via a vicariant event (suggested by the nested clade analysis). More recently, these populations expanded back to Southeast and central Brazil from southern populations (suggested by the nested clade analysis and the neutrality tests). In a previous publication (de Brito, 1999), we used the sudden population expansion model of Rogers and Harpending (1992) to infer that a population size expansion in Southeast Brazil may have happened between about 105 to 154 thousand years ago, suggesting that *D. buzzatii* populations have been present in Brazil much earlier than the Holocene.

The existence of an earlier fragmentation in Brazilian populations of *D. buzzatii* has been suggested by allozyme studies (Barker *et al.*, 1985) but not by chromosomal inversion studies (Wasserman, 1962; Barker *et al.*, 1985; Figueiredo and Sene 1992), in fact, a striking paucity of chromosomal inversion polymorphisms has been observed in Brazilian populations by Figueiredo and Sene (1992). Populations in Southern Brazil show only two of 16 known inversions for the species and no polymorphisms were detected in the Northeast. This is even more remarkable considering that recently colonized areas such as Spain and Australia (where a strong population bottleneck was considered to have occurred) show higher chromosomal polymorphism (Knibb and Barker, 1988; Fontdevila, 1989; Hasson *et al.*, 1991). Chromosomal inversions have been shown to be associated with some ecological gradients not only in *D. buzzatii* (Ruiz *et al.*, 1986; Knibb and Barker, 1988; Hasson *et al.*, 1991) but also in other *Drosophila* species (Ayala *et al.*, 1989). Therefore, their use to detect population substructure may be affected by the effects of natural selection. Should this be the case, the lack of polymorphism in Brazilian populations may be more informative to the study of selective pressure on these inversions than to the study of gene flow and population differentiation.

Our results favor a pre-Holocene presence of *D. buzzatii* in Brazil, not only because populations of *D. buzzatii* in Northeastern Brazil show higher polymorphism levels but also because of the indication of significant restricted gene flow across most of the distribution of *D. buzzatii*. Furthermore, estimates that an expansion in population size in Southeast Brazil occurred at least 100 thousand years ago indicates that the presence of *D. buzzatii* in Brazil predates human colonization.

The two hypotheses (active and passive dispersal) originally proposed to explain the occupation of Brazil by *D. buzzatii* consider Argentina to be the center of origin of the species. This assumption is mostly based on the fact that most of the chromosomal variation found in natural populations of *D. buzzatii* occurs in Argentinean populations. The results presented in this paper indicate that our data does not support the passive dispersal hypothesis. The high polymorphism levels found in *D. buzzatii* populations occurring

in the Northeast of Brazil, as well as the private allozyme alleles found by Barker *et al.* (1985), suggests that *D. buzzatii* was distributed throughout Argentina and Brazil, and that the concept of an Argentinean center of origin for *D. buzzatii* should be viewed with caution.

Acknowledgments

We thank Alan Templeton, Jim Cheverud, Ian Duncan, Allan Larson, Jonathan Losos, Mick Richardson, Guto Kuhn, Sergio Matioli, Andréa Peripato and two anonymous reviewers for help in previous versions of this manuscript. We especially thank Alan Templeton and Jim Cheverud for intellectual input. This work was supported by an NSF DEB-9701809 grant to Alan R. Templeton and R.B.; FAPESP, CAPES and CNPq grants to F.M.S. R.B. was supported by a fellowship from the Division of Biology and Biomedical Sciences - Washington University in St. Louis.

References

- Ab'Saber AN (1990) Paleoclimas quaternários e pré-história da América tropical II. Rev. Bras. Biol. 50:821-831.
- Ayala FJ, Serra L and Prevosti A (1989) A grand experiment in evolution: the *Drosophila subobscura* colonization of the Americas. Genome 31:246-255.
- Barker JSF, Sene FM, East PD and Pereira MAQR (1985) Allozyme and chromosomal polymorphism of *Drosophila buzzatii* in Brazil and Argentina. Genetica 67:161-170.
- Bertorelle G and Slatkin M (1995) The number of segregating sites in expanding human populations, with implications of estimates of demographic parameters. Mol. Biol. Evol. 12:887-892.
- Betran E, Santos M and Ruiz A (1998) Antagonistic pleiotropic effect of second-chromosome inversions on body size and early life-history traits in *Drosophila buzzatii*. Evolution 52:144-154.
- Cann RL, Stoneking M and Wilson AC (1987) Mitochondrial DNA and human evolution. Nature 325:31-36.
- Carson HL and Wasserman M (1965) A widespread chromosomal polymorphism in a widespread species *Drosophila buzzatii*. Amer. Natur. 99:111-115.
- Castelloe J and Templeton AR (1994) Root probabilities for intraspecific gene trees under neutral coalescent theory. Mol. Phylogenet. Evol. 3:102-113.
- Cline J, Braman JC and Hogrefe HH (1996) PCR fidelity of *pfu* DNA polymerase and other thermostable DNA polymerase. Nucl. Acids Res. 24:3546-3551.
- de Brito RA (1999) Contrasting ecological and historical factors shaping the evolution of two cactophilic flies, *Drosophila serido* and *D. buzzatii* in Brazil. Ph.D. Dissertation, Washington University.
- de Brito RA, Manfrin MH, Sene FM (2002) Nested cladistic analysis of Brazilian populations of *Drosophila serido*. Mol. Phylogenet. Evol. 22:131-143.
- Excoffier L, Smouse PE and Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. Genetics 131:479-491.
- Figueiredo VLC and Sene FM (1992) Chromosome variability in Brazilian populations of *Drosophila buzzatii* (Diptera, Drosophilidae) Rev. Bras. Biol. 52:555-561.
- Fontdevila A (1989) Founder effects in colonizing populations: the case of *Drosophila buzzatii*. In: Fontdevila A (ed) Evolutionary Biology of Transient Unstable Populations. Springer-Verlag, Berlin; New York, pp. 74-95.
- Fontdevila A, Ruiz A, Alonzo G and Ocaña J (1981) Evolutionary history of *Drosophila buzzatii*. I. Natural chromosomal polymorphism in colonized populations of the Old World. Evolution 35:148-157.
- Fu Y-X (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. Genetics 147:915-925.
- Fu Y-X and Li W-H (1993) Statistical tests of neutrality of mutations. Genetics 133:693-709.
- Halliburton R and Barker JSF (1993) Lack of mitochondrial DNA variation in Australian *Drosophila buzzatii*. Mol. Biol. Evol. 10:484-487.
- Hasson E, Vilardi JC, Naveira H, Fanara JJ, Rodriguez C, Reig OA and Fontdevila A (1991) The evolutionary history of *Drosophila buzzatii*. XVI. Fitness components analysis in a natural original population from Argentina. J. Evol. Biol. 4:209-225.
- Hasson E, Naveira H and Fontdevila A (1992) The breeding sites of the Argentinian species of the *Drosophila mulleri* complex (subgenus *Drosophila*-repleta group) Rev. Chil. Hist. Nat. 65:319-326.
- Hudson RR, Slatkin M and Maddison WP (1992) Estimation of levels of gene flow from DNA sequence data. Genetics 132:583-589.
- Knibb WR and Barker JSF (1988) Polymorphic inversion and esterase loci complex on chromosome 2 of *Drosophila buzzatii*. Austral. J. Biol. Sci. 41:239-246.
- Maddison W and Maddison D (1992) MacClade Version 3: Analysis of Phylogeny and Character Evolution Sinauer Associates, Sunderland, MA.
- Manfrin MH, Brito RA and Sene FM (2000) Systematics and Evolution of the *Drosophila buzzatii* (Diptera: Drosophilidae) Cluster using mtDNA. Ann. Entomol. Soc. Amer. 94: 333-346.
- Nachman MW (1998) Deleterious mutations in animal mitochondrial DNA. Genetica 102/103:61-69.
- Nielsen R and Weinreich DM (1999) The age of nonsynonymous and synonymous mutations in animal mtDNA and implications for the mildly deleterious theory. Genetics 153:497-506.
- Ohta T (1992) The nearly neutral theory of molecular evolution. Ann. Rev. Ecol. Syst. 23:263-286.
- Pereira MAQR, Vilela CR and Sene FM (1983) Notes on breeding and feeding sites of some species of the *repleta* group of the genus *Drosophila* (Diptera, Drosophilidae) Cienc. Cult. 35:1313-1319.
- Posada D, Crandall KA, Templeton AR (2000) GeoDis: a program for the cladistic nested analysis of the geographical distribution of genetic haplotypes. Mol. Ecol. 9:487-488.
- Powell JR (1997) Progress and Prospects in Evolutionary Biology: The *Drosophila* model, 1st ed. Oxford University Press, New York, Oxford.

- Prado DE and Gibbs PE (1993) Patterns of species distribution in the dry seasonal forests of South America. *Ann. Missouri Bot. Gard.* 80:902-927.
- Przeworski M, Charlesworth B and Wall JD (1999) Genealogies and weak purifying selection. *Mol. Biol. Evol.* 16:246-252.
- Rand DM, Dorfsman M and Kann LM (1994) Neutral and non-neutral evolution of drosophila mitochondrial DNA. *Genetics* 138:741-756.
- Rogers AR and Harpending HC (1992) Population growth makes waves in the distribution of pairwise genetic differences. *Mol. Biol. Evol.* 9:552-569.
- Rossi MS, Barrio E, Latorre A, Quezada-Diaz JE, Hasson E, Moya A and Fontdevila A (1996) The evolutionary history of *Drosophila buzzatii*. XXX. Mitochondrial DNA polymorphism in original and colonizing populations. *Mol. Biol. Evol.* 13:314-323.
- Rozas J and Rozas R (1999) DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. *Bioinformatics* 15:174-175.
- Ruiz A and Fontdevila A (1982) Evolutionary history of *Drosophila buzzatii*. III. Cytogenetic relationships between two sibling species of the *buzzatii* cluster. *Genetics* 101:503-518.
- Ruiz A and Santos M (1989) Mating probability, body size, and inversion polymorphism in a colonizing population of *Drosophila buzzatii*. In: Fontdevila A (ed) *Evolutionary Biology of Transient Unstable Populations*. Springer-Verlag, Berlin; New York, pp. 96-113.
- Ruiz A, Fontdevila A, Santos M, Seoane M and Torroja E (1986) The evolutionary history of *Drosophila buzzatii*. XI. Analysis of selection acting on the inversion polymorphism in a natural population. *Evolution* 40:740-755.
- Schmid KJ, Nigro L, Aquadro CF and Tautz D (1999) Large number of replacement polymorphisms in rapidly evolving genes of *Drosophila*: Implications for genome-wide surveys of DNA polymorphism. *Genetics* 153:1717-1729.
- Schneider S, Kueffer J-M, Roessli D and Excoffier L (1997) Arlequin ver. 1.1: A software for population genetic data analysis. *Genetics and Biometry Laboratory, University of Geneva, Switzerland*.
- Simon C, Frati F, Beckenbach A, Crespi B, Liu H and Flook P (1994) Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved Polymerase Chain Reaction primers. *Ann. Entomol. Soc. Am.* 87:651-701.
- Simonsen KL, Churchill GA and Aquadro CF (1995) Properties of statistical tests of neutrality for DNA polymorphism data. *Genetics* 141:413-429.
- Slatkin M (1991) Inbreeding coefficients and coalescence times. *Genet. Res* 58:167-175.
- Slatkin M and Hudson RR (1991) Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. *Genetics* 129:555-562.
- Swofford DL (1998) PAUP*. *Phylogenetic Analysis Using Parsimony (*and Other Methods)* Version 4 Sinauer Associates, Sunderland, Massachusetts.
- Tajima F (1983) Evolutionary relationship of DNA sequences in finite populations. *Genetics* 105:437-460.
- Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123:585-595.
- Tajima F (1996) The amount of DNA polymorphism maintained in a finite population when the neutral mutation rate varies among sites. *Genetics* 143:1457-1465.
- Templeton AR (1994) "Eve": Hypothesis compatibility versus hypothesis testing. *Am. Anthropol* 96:141-147.
- Templeton AR (1998) Nested clade analyses of phylogeographic data: testing hypotheses about gene flow and population history. *Mol. Ecol.* 7:381-397.
- Templeton AR, Boerwinkle E and Sing CF (1987) A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping. I. Basic theory and analysis of alcohol dehydrogenase activity in *Drosophila*. *Genetics* 117:343-351.
- Templeton AR, Sing CS, Kessling A and Humphries S (1988) A cladistic analysis of phenotype associations with haplotype inferred from restriction endonuclease mapping. II. The analysis of natural populations. *Genetics* 120:1145-1154.
- Templeton AR, Crandall, KA and Sing CF (1992) A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. *Genetics* 132:619-633.
- Templeton AR, Routman E and Phillips CA (1995) Separating population structure from population history: A cladistic analysis of the geographical distribution of mitochondrial DNA haplotypes in the Tiger salamander, *Ambystoma tigrinum*. *Genetics* 140:767-782.
- Thomas RH and Barker JSF (1990) Breeding structure of natural populations of *Drosophila buzzatii*: effects of the distribution of larval substrates. *Heredity* 64:355-365.
- Tidon-Sklorz R, Vilela CR, Sene FM and Pereira MAQR (1994) The genus *Drosophila* (Diptera, Drosophilidae) in the Serra do Cipó, State of Minas Gerais, Brazil. *Rev. Bras. Entomol.* 38:627-637.
- Vilardi JC, Hasson E, Rodriguez C and Fanara JJ (1994) Genetic structure is determined by stochastic factors in a natural population of *Drosophila buzzatii* in Argentina. *Genetica* 92:123-128.
- Vilela CR, Sene FM and Pereira MAQR (1980) On the *Drosophila* fauna of chaco and east slopes of the Andes in Argentina. *Rev. Bras. Biol.* 40:837-841.
- Wayne ML and Simonsen KL (1998) Statistical tests of neutrality in the age of weak selection. *Tree* 13:236-240.
- Wright S (1951) The genetical structure of populations. *Ann. Eugenics* 15:323-352.