Mitochondrial phylogeography of the high Andean frog *Hyla labialis*

Carlos Enrique Guarnizo

Trabajo de grado para optar al título de Magister en Biología

Director: Adolfo Amézquita, PhD.
Co-Director: Eldrege Bermingham, PhD.
Jurado: Martha Vives, MSc.

Universidad de los Andes
Facultad de Ciencias
Departamento de Ciencias Biológicas
Bogotá D.C.
2003

Introduction

The understanding of evolutionary processes at population level is important because those short time genetic variations may partially explain the wide diversity of species on the Earth. In order to understand population genetic structures, it is not enough to scan genetic variability alone. A lot of information is needed to have a closer view of what actually happens. Geographic distribution is an important source of information that, together whit genetic aspects, could give us clues about the past and present population dynamics.

Phylogeography is a recent discipline that mixes geographic and genetic information in order to understand population biology aspects (Avise *et al*. 1987). Only with the appearance of DNA analysis techniques, it was possible to develop this field. Particularly important was the mitochondrial DNA (mtDNA) analysis, because of it's properties of high rate of mutation and no recombination. The first one allows the analysis of populations recently separated, and the second allows the use of common phylogenetic techniques exclusively used at a species level, on populations. The fact that mtDNA is inherited by maternal lines makes it possible to use individuals as taxonomic units without restrictions due to reticulation (Avise, 2000).

Geographic variation of gene sequences is attributable to random or selective (adaptive) processes. According to the first explanation, proposed by Moto Kimura in late 1960's, genetic drift modifies the genetic structure of populations by chance (Lewin, 1997). If so, then differences among populations could be related to geographic linear distance (stepping stone model), or to the presence of geographic barriers for gene flow. This model is commonly known as the neutral model of variation (Lewin, 1997; Hartl & Clark, 1997; Futuyma, 1998). Changes in the genetic structure of a population could be also associated to the fit of some phenotypes to specific environments. During this process natural selection chooses some phenotypes that, over time, could form adaptations (Moody, 1962; Futuyma, 1998). Recently, a controversy appeared between evolutionary biologists defending either neutral variation or natural selection as the main source of change in natural populations. Although the debate remains today, it is widely accepted that, at the species level (macroevolutionary processes), natural selection plays a crucial role generating variation, whereas at the molecular level (microevolutionary processes),

the neutral variation is the leading force. Summing up, the phenotypic and/or genetic changes of organisms over geographic areas could be due, as far as we know today, to random processes or to selective processes.

Study Model

The frog *Hyla labialis* is characterised by its wide altitudinal and latitudinal distribution over the Eastern Andes of Colombia, between 1600 and 3600 masl (Ruiz *et al*. 1996). Populations exhibit phenotypic variation related to the altitudinal gradient in which they occur. This variation could represent adaptations to local conditions (ecogeographic variation; Lüddecke, 1995; Navas, 1996). Variation includes traits like thermal preferences, body size (individuals from paramo populations could be three times as larger as individuals from lower populations), coloration patterns (Amézquita, 1999), growth and developmental rate of the post-metamorphs (Amézquita 1995, Amézquita & Lüddecke, 1999), and some features of the advertisement call (Amézquita 2001).

In a recent study, Amézquita (2001) analyzed the geographic variation in call characteristics and body size of seven populations of *Hyla labialis*. Since vocalizations of anurans are fundamental for reproduction and species recognition, it is expected a high natural selection pressure to maintain the call characteristics with no change and, simultaneously, a high sexual selection pressure to increase call attractiveness (Rand, 2001). The Amézquita's results show variation on dominant frequency due to pleiotropic relations to size end elevation, micro temporal characteristics due to isolation by distance, and thermal sensitivity due to north-south classification (Amézquita, 2001). An interesting result is that in northern populations the call characteristics are very different from the southern ones. That is interesting because the change occurs abruptly and there are no clear geographic barriers (at least in the present), which could make them differentiate in that way.

Aiming at understanding the mitochondrial DNA phylogeographic patterns of the frog *H. labialis*, I performed an analysis of mitochondrial DNA genes on individuals from the same and additional populations included in Amezquita's study (2001). My interest was to know whether the chosen genes had the enough level of variation to make a population level study, and whether this variation was related to *a)* the pattern of altitudinal (adaptive) variation of body size or *b)* the pattern of isolation by distance (neutral) variation of vocalizations. Actually, there is no genetic information on the populational genetic structure of this species.

Aims

-To construct a phylogeographic hypothesis for this species.
-To test several models of geographic variation (stepping stone, geographic barriers, panmixia...) that may explain the biogeographic history of these populations.
-To relate the obtained gene genealogy with previous results about the phenotypic and behavioral variation among these populations.

Materials and Methods

The *H. labialis* tissues were collected during the first semester of 2002. The geographic coordinates for each site were registered using GPS. I had samples of ten individuals from ten carefully chosen populations that represent most of the geographic distribution range of this species (departments of Cundinamarca and Boyacá; fig 1, table 1). Tissues from adults (finger tips) or tadpoles (tail tips) were kept on ethanol 97% and transported (exportation permission No 439) to the NAOS labs of Molecular Biology and Evolution of the Smithsonian Tropical Research Institute (Panamá). Initially, only one or two individuals from each population were sequenced for the mitochondrial genes CO1 (cytochrome oxidase 1) and ND2 (NADH dehydrogenase 2). Both genes have a high rate of evolution and are used mainly for interspecific studies on a wide group of organisms (Hillis *et al*. 1996)

Total genomic DNA was extracted using the standard phenol-chloroform method (Sambrook *et al*. 1989). The CO1 gene was amplified using two external primers (Palumbi *et. al*. 1991): CO1a-H (AGTATAAGCGTCTGGGTAGTC) and CO1f-L

(CCTGCAGGAGGAGGAGAYCC). For the ND2 gene I used two external primers too (Macey *et. al*. 1997): Metf.6 (AAGCTTTCGGGCCCATACC) and CO1r.1 (AGRGTGCCAATGTCTTTGTGRTT). Double stranded amplifications were performed with a reaction mix containing: 1.0 μl of 10X PCR buffer (500 mM KCl, 100mM Tris-HCl, pH 8.5), 1.0 μl of dNTP's 8 μM, 4.5 μl of dH2O (Sigma), 0.4 μl of $MgCl_2$, 0.1 μl of Quiagen DNA polymerase, 0.5 μl of each primer, and 2 μL of the total DNA for a final volume of 10 μL. PCR amplifications included an initial denaturing step at 96°C for 6 min, then, six amplification cycles (45 sec 96°C, 45 sec 50°C, 1:30 sec 72°C), then 29 cycles (45 sec 96°C, 45 sec 56°C, 1:30 sec 72°C) and a final extension of 72°C for 6 min. The PCR fragments were gel purified by Gelase treatment.

The sequencing cycle amplifications were done using the same primers for the CO1 gene and additional internal primers for the ND2 gene: ANSr.2 (GCGTTTAGCTGTTAACTAAA), TRPf.5 (GACCAAAGGCCTTCAAAGCC), and CO1.r.1 (AGRGTGCCAATGTCTTTGTGRTT). Sequences were determined with an automatic sequencer (MJ Research GeneScan) following the manufacturer protocols. Both strands were sequenced in both directions for each individual in order to avoid ambiguities. Sequences were easily aligned by eye using the software Sequencher Version 3.1 (Gene Codes Corporation Inc, 1998). Protein-coding sequences were translated to aminoacids using MacClade (Maddison & Maddison, 1992) for alignment confirmation.

Table 1. Sumary of location and procedures for each studied population of *Hyla labialis*.

Population	Coordinates (lat N, long W)	Altitude (m.a.s.l.)	Genes scanned	# of individuals
El Manzano	05.45.05, 73.10.40	2577	CO1, ND2	2(CO1), 1(ND2)
Cucaita	05.32.45, 73.27.03	2688	CO1, ND2	1(CO1), 1(ND2)
Villa de Leyva	05.38.55, 73.31.56	2170	ND2	1(ND2)
San Carlos	05.35.54, 73.43.02	2590	CO1	2(CO1)
Cucunubá	05.06.07, 73.47.45	2592	CO1, ND2	1(CO1), 2(ND2)
Suesca	05.05.06, 73.46.38	2678	CO1, ND2	2(CO1), 1(ND2)
Cota	04.48.35, 74.06.06	2600	CO1, ND2	1(CO1), 1(ND2)
Las Juntas	04.38.27, 74.13.16	2650	CO1	2(CO1)
Chingaza	04.41.25, 73.48.23	3550	CO1	1(CO1)

Las Brisas	04.26.12, 73.55.10	1970	CO1	1(CO1)

Phylogenetic trees were estimated using PAUP* version 4.0b2 (Swofford, 1999). I calculated Maximum Likelihood (ML) treating gaps as missing data and conducted branch and bound searches using the model of substitution proposed by MODELTEST (Posada & Crandall 1998). The tree searches used a starting tree obtained via stepwise addition and the tree bisection-reconnection (TBR) was used as branch swapping algorithm. I also conducted maximum parsimony (MP) and neighbor joining (NJ) analysis in order to compare the main methods for phylogenetic inference. In order to find a relationship between geographic and genetic distances I made a Mantel test with the software R Package (Legendere & Vaudor, 1991).

Results

I obtained sequences from 13 individuals for the CO1 gene (585 bp) and 7 individuals for the ND2 gene (750 bp). The alignments showed that most of the variation was on the third-codon position, and consisted only of four aminoacides changes. For the CO1 gene 555 characters were constant, 13 characters were variable and parsimony uninformative, and 51 characters were variable and parsimony informative. For the ND2 gene 1041 characters were constant, 40 variable characters were parsimony uninformative and 7 characters were parsimony informative. For the CO1 gene the number of observed transitions was 48 and the number of transversions observed was 6. For the ND2 gene the number of observed transitions was 3 and the number of transversions observed was 1 (fig. 2).

The proportion of each nucleotide for the CO1 gene was A=0.24320, C=0.24563, G=0.18070 and T= 0.33047, and for the ND2 gene was A= 0.29549, C= 0.27039, G= 0.15554 and T= 0.27857. The Maximum Likelihood (ML) analysis was done using the model chosen by MODELTEST: HKY85 + G, that assumes different mutation rates for transitions (Ti) and transversions (Tv), different proportions for each nucleotide and differential mutation rate for different parts of the genome (gamma distribution).

The three methods of phylogenetic inference basically provided the same tree, with small variations on the least bootstrap support branches. The CO1 gene genealogy (fig. 5) presents two main branches, one belongs to the northern populations (Cucaita and

Manzano), and the other the southern ones (Cota, Chingaza, Las Brisas, Las Juntas, Suesca, Cucunubá and San Carlos). Inside the southern clade there are two main sub-groups, one with the populations of San Carlos and Cucunubá, and the other with Cota, Chingaza, Las Brisas, Las Juntas and Suesca. The branch lengths and bootstrap values that separate the north-south clades are much longer than the ones that separate the sub-branches of the south clade. For the ND2 gene (fig. 6), the main clades again group the northern populations (Cucaita and Manzano), and the Sourthern populations (Villa de Leyva, Cucunuba, Suesca and Cota). Although Villa de Leyva is very close to one of the northern populations, it was included in the southern populations.

The Mantel correlation test for the CO1 gene showed a high degree of correlation between geographic and genetic distances on southern populations (r=0.607, 999 permutations, P=0.001), although for the ND2 (fig 3) there's no significant correlation (r=0.177, 999 permutations, P=0.22). The plot that relates the geographic and genetic distances for the CO1 gene (fig. 3) shows a positive correlation (Rsq= 026, ANOVA: df=1, F=16.871, P=0.000;) between geographic and genetic distances for pairs of populations within the same clade (intracluster populations). On the other hand, the tendency shown by the north-south populations (intercluster) is that geographic distance does not affect the genetic distances (Rsq= 0.0064; ANOVA: df=1, F=0.180, P=0.675). The scatter point distribution representing the intra and intercluster populations were very different, being higher the genetic differences of the intercluster ones. The ND2 gene (fig. 3) shows no relation between geographic and genetic distances, and there is no clear separation between intra and intercluster points (Rsq= 0.033, ANOVA: df=1, F=0.658, P=0.427).

Controlling the effect of geographic distances, there is no effect of altitude on genetic distances for the CO1 gene (ANCOVA: df=1, F=2.890, P=0.093; Fig. 4), neither for the ND2 gene (ANCOVA: df=1, F=0.326, P=0.575)

Discussion

Amphibians generally have high degrees of interpopulation genotype structure (Avise, 1994), in comparison with other taxonomic groups as reptiles, birds and mammals. This may be explained because amphibian dispersal abilities are constrained by high humidity places; by contrast, other vertebrates may disperse through longer displacements without

being under physiological risk. Thus, the probability of gene flow with neighboring populations increases, homogenizing the interpopulational genetic structure (Hartl & Clark 1997).

The more variable gene in this study was CO1, and most substitutions were neutral (third codon positions), the transitions were eight times as frequent as transversions. However there were also substitutions on four aminoacids in the CO1 gene, three changes occurring in the northern populations, Cucaita and Manzano, and the other one in the paramo population of Chingaza. The aminoacid substitutions on northern populations support some degree of differentiation, probably due to reduced gene flow during a long time.

Part of the genetic structure of CO1 gene, which positively relates the geographic and genetic distances, supports the isolation-by-distance (stepping stone model; Hutchison & Templeton, 1999). This probably means that closer populations were more recently separated or they more probably experience the homogenizing effect of gene flow. This evidence supports the neutral mode of variation, which is more related with the pattern of vocal differentiation, than the pattern of body size differentiation. Moreover, intercluster genetic differences are higher than intracluster ones, and, this strong variation is not explained by geographic distance alone. There may have existed a sort of geographic barrier for gene flow between southern and northern populations. Today, there is no clear evidence of geographic barriers over the whole range of distribution of this species (Amézquita, com. Pers.).

The ND2 gene fails to show a separation between intra and intercluster populations. The whole ND2 gene wasn't sequenced because some problems with PCR. One of the regions that wasn't sequenced was the more variable region placed on the middle of the gene (Crawford com. Pers). Thus, only a small part of ND2 was analyzed, with only 7 parsimony informative sites available. It is necessary to sequence the whole gene on more populations in order to obtain additional useful information.

The phylogenetic conclusion drawn from the sequence analysis of the CO1 gene is strong, as indicated by the fact that multiple methods of tree reconstruction procedures (ML, MP, NJ) produced almost identical topologies (Hillis *et.al*. 1996). In general, the tree topology

agrees well with the geographic distribution of populations (fig. 8). The most pronounced (100% bootstrap value) phylogeoghraphic pattern shown by the CO1 gene is the separation of the northern populations of Cucaita and Manzano from the other populations. Although the linear distance between Cucaita and Manzano is longer than between Cucaita and San Carlos, the genetic distance between San Carlos and Cucaita is much larger.

The second well-supported clades (99% bootstrap value) group on one hand San Carlos and Cucunubá, and on the other hand Cota, Chingaza, Las Brisas, Las Juntas and Suesca. Given the high values of branch lengths between northern and southern populations, both clades may represent two different species. This idea may be further supported by the pattern of vocal variation.

There is no relationship between the difference in altitude between pairs of populations and the corresponding genetic distances. However, most dots correspond to small altitudinal differences between pairs of populations. Thus, although the CO1 molecule is associated with respiration, and *Hyla labialis* have a wide altitudinal distribution, there is no strong evidence of natural selection.

The gene genealogy of ND2 also separates the populations of Cucaita and Manzano from the rest populations with a bootstrap value of 92%. The internal clade includes Villa de Leyva, that geographically belongs to the northern populations, but it is grouped with the southern populations. I found two different haplotypes for the Cucunubá population and they were placed on different branches of the cladogram. This separation is doubtful since one of the individuals from Cucaita had the longest sequence of all the ND2 gene, probably affecting his place in the cladogram. The tree topology approximately agrees with the geographic distributions of populations and there is no significant relationship between altitudinal and genetic distances. This contrast again support that phylogeographic variation (neutral) is stronger than ecogeographic (adaptive) variation.

Conclusions

1) The CO1 and ND2 genes were variable enough to make population-level and phylogeographic studies on *Hyla labialis*.
2) The genetic structure shown by the CO1 and ND2 genes agrees with the model of isolation by distance, meaning that phylogeographyc, not ecogeographic processes mainly structured this genes.
3) The genetic data supports a high degree of differentiation between northern and southern populations.
4) The genetic pattern of variation agrees with the pattern of vocal variation, but not with the pattern of altitude-related size variation.

Literature cited

Amézquita, A. 1995. Estudio experimental sobre la variación interpoblacional en el crecimiento post-metamórfico y el fenotipo de la madurez sexual de *Hyla labialis* (Anura : Hylidae). MSc Thesis, Universidad de los Andes. Bogotá.

Amézquita, A. 1999. Color pattern, elevation and body size in the high andean frog *Hyla labialis*. Re. Acad. Colomb. Cienc. 23 (suplemento especial: 231-238)

Amézquita, A., 2001. Signal diversity and the evolution of the vocal communication system in the high Andean frog *Hyla labialis*. PhD Thesis, Universidad de los Andes. Bogota.

Amézquita, A., Lüddecke, H. 1999. Correlates of intrapopulational variation in size at metamorphosis of the high Andes frog *Hyla labialis*. Herpetologica. 55: 295-303.

Avise, J. C. 1994. Molecular Markers, Natural History and Evolution. Chapman & Hall. New York.

Avise, J.C., 2000. Phylogeography, the History and Formation of Species. Harvard University Press. Cambridge, Massachusetts.

Futuyma D. 1998. Evolutionary Biology. Third Edition. Sinauer. USA.

Hartl, D. L., Clark, A.G. 1997. Principles of Population Genetics. Third edition. Sinauer Associates. Massachuetts.

Hillis D.M., Moritz C., Mable B., K. 1996. Molecular Systematics. Second Edition. Sinauer. USA.

Hutchinson, D. W., Templeton, A. R. 1999. Correlation of pairwise genetic and geographic distance measures: inferring the relative influences of gene flow and drift on the distribution of genetic variability. Evolution, 53: 1898-1914

Legendre, P., Vaudor, A. 1991. The R package: Multidimensional analysis, spatial analysis. Departament de sciences biologiques. Université de Montreal. 142p

Lewin, R., 1997. Patterns in Evolution, the New Molecular View. Scientific American Books. England.

Lüddecke, H. 1995. Intra and interpopulational comparison of temperatures selected by *Hyla labialis*. Pp. 192-196. Eds. Llorente, G.A., Montori, A., Santos, X., Carretero, M.A. Scientia Herpetologica. Barcelona.

Lynch, J.D. 1999. Ranas pequenas, la geometría de la evolución, y la especiación en los Andes colombianos. Rev. Acad. Colomb. Cienc. 23: 143-159.

Macey, J. R., Larson, A., Anajeva, J. B., Fang, Z., Papenfus, T. J. 1997. Two novel gene orders and the role of light strand replication arrangement of the vertebrate mitochondrial genome. Mol. Biol. 14: 91-104.

Maddison, W.P., Maddison, D. R. 1992. Mac Clade analysis of phylogeny and character evolution, version 3.0. Sinauer, Saunderland, M.A.

Moody, P.A. 1962. Introduction to Evolution. Second Edition. A Harper International Student Reprint. London.

Navas, C.A. 1996. Thermal dependency of field locomotor and vocal performance of high elevation anurans in the tropical Andes. J. Herpetol. 30: 478-487.

Palumbi, S. R., Martin, A. P., Romano, S., McMillan, W. O., Slice, l., Grabowsky, G. 1991. The simple fool guide to PCR. Special Publ. Dept. Zoology. University of Hawaii. Honolulu.

Posada, D., Crandall, K. A. 1998. MODELTEST. Testing the model of DNA substitution. Bioinformatics 14: 817-818.

Rand, A., S. 2001. A History of Frog Calls Studies, 405 B.C. to 1980. In Anuran Communication. *Ed.* M. J., Ryan. Smithsonian Institution Press. USA.

Ruiz-Carranza, P.M., Ardila-Robayo, M.C., Lynch, J.D., 1996. Lista actualizada de la fauna Amphibia de Colombia. Rev. Acad. Colomb. Cienc. 20: 365-415.

Sambrook, J., Fritch, E. F., Maniatis, T. 1989. Molecular Cloning. A Laboratory Manual. Second Edition. Cold Spring Harbor (NY). Laboratory Press. Vol. 3.

Swofford, D. 1999. PAUP* Ver. 4.0b2. Smithsonian Institution. Washington, D.C.

Annexed figures

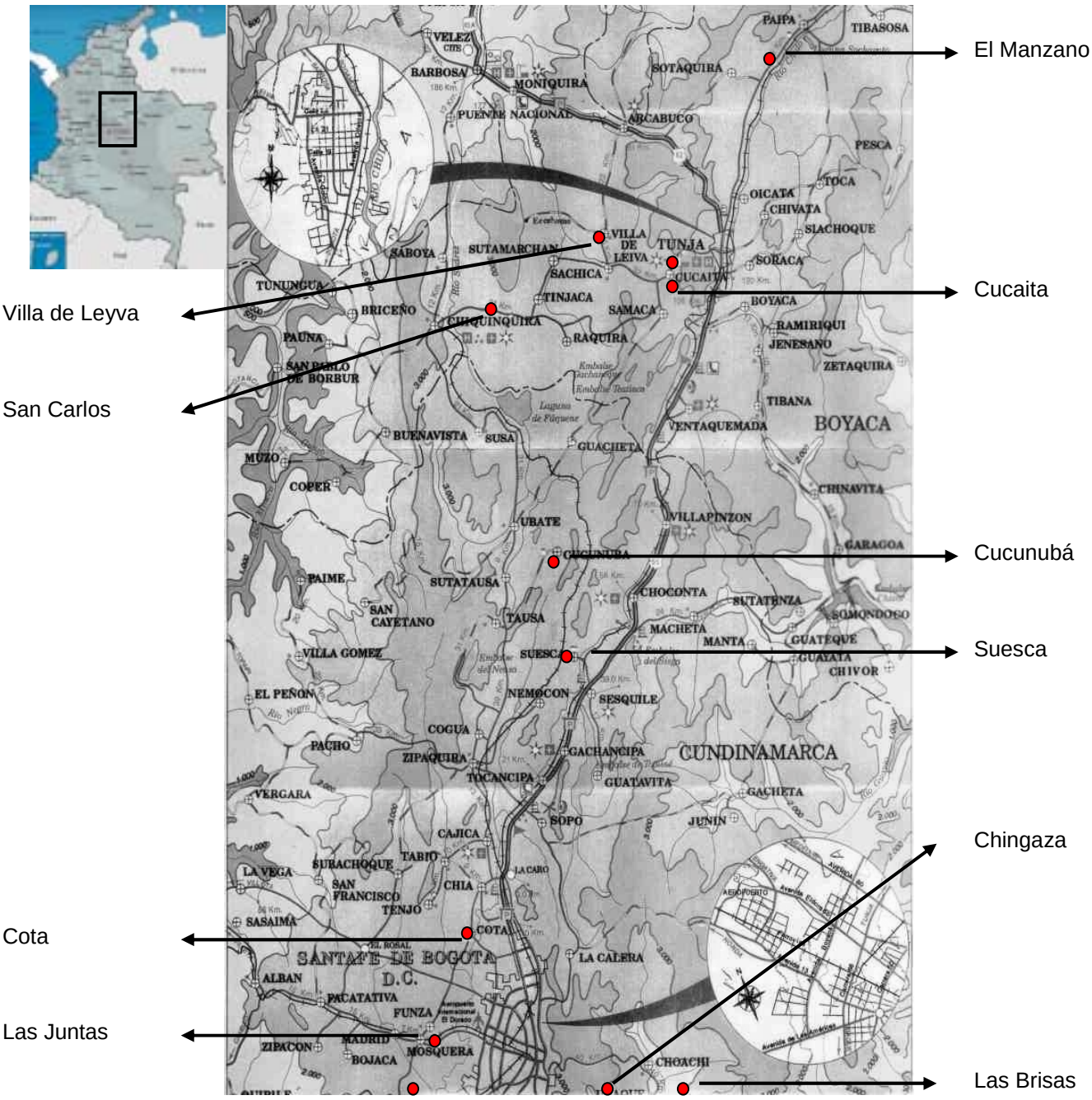


Fig. 1 Geographic distribution of populations of the frog *Hyla labialis* in this study.

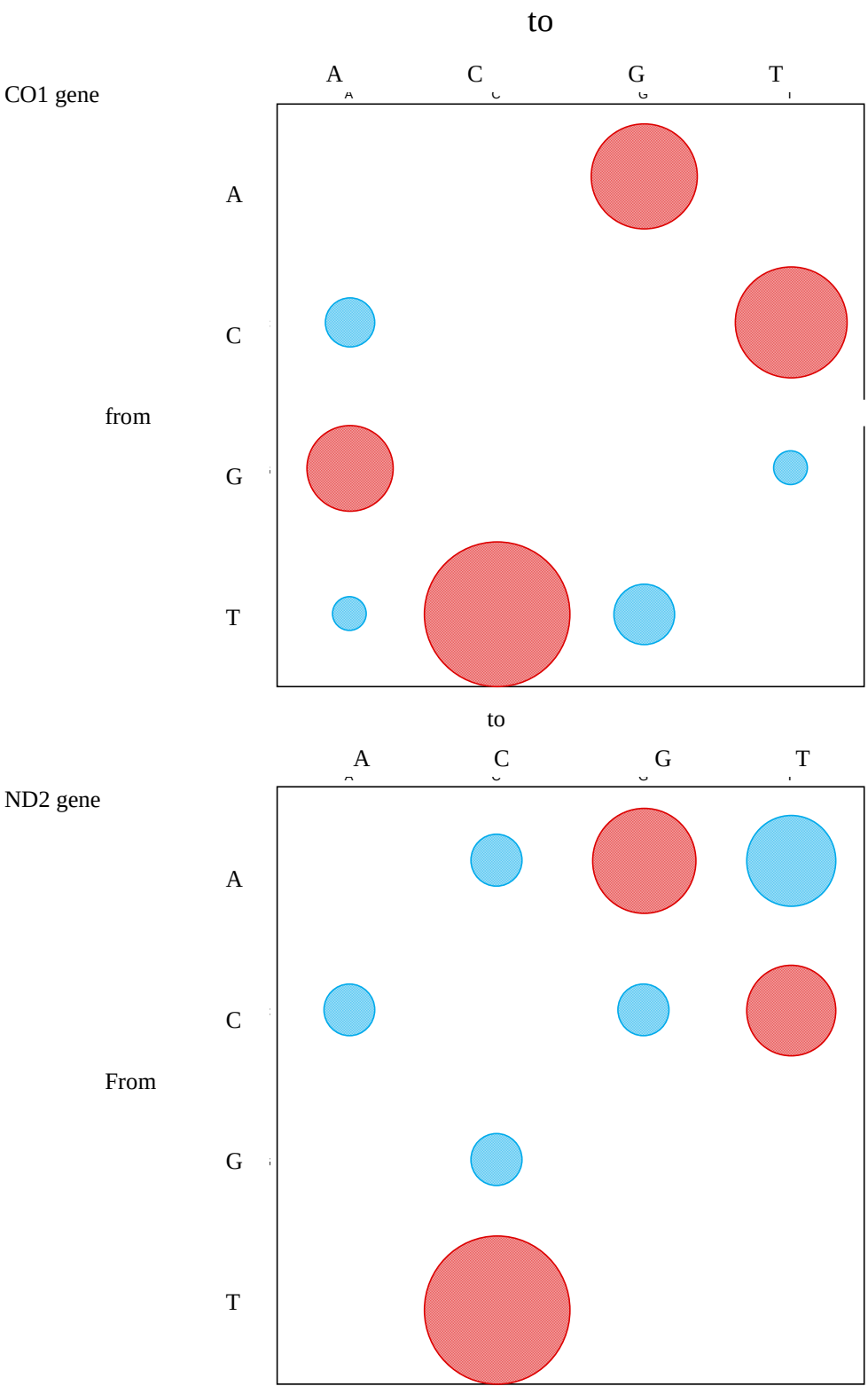
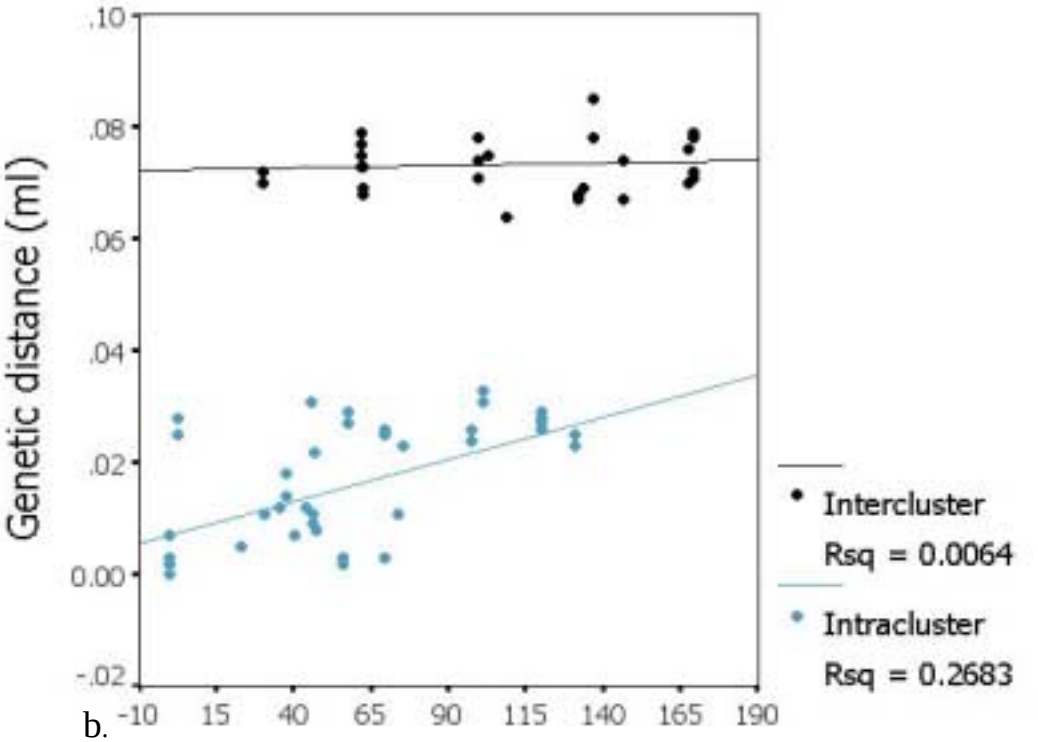
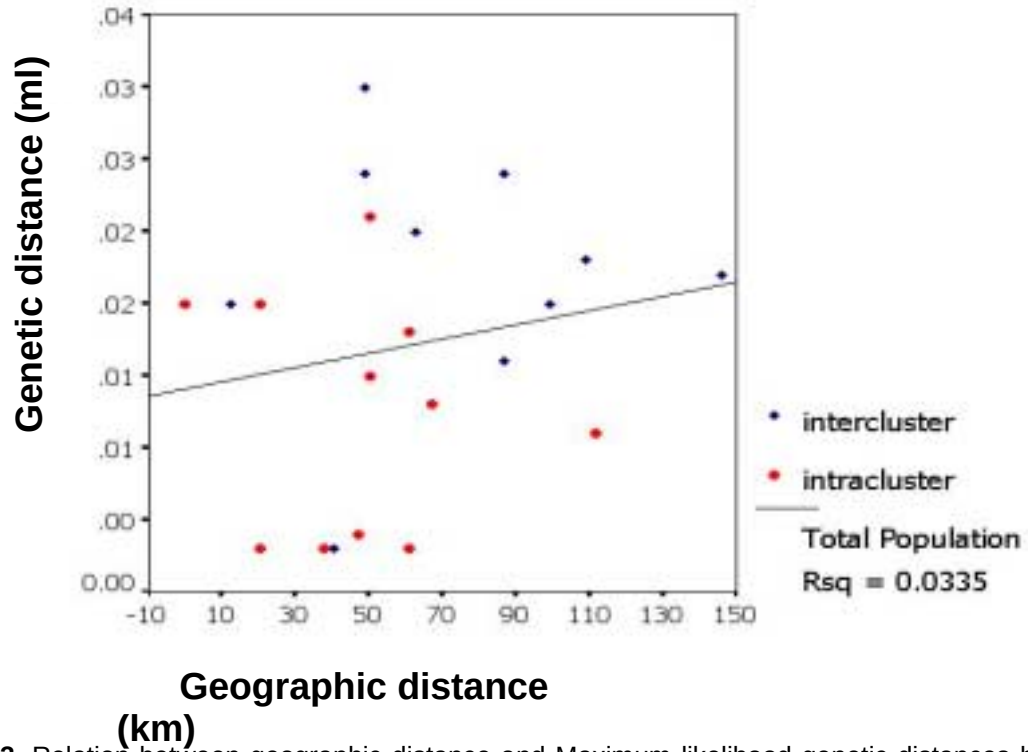


Fig. 2. Proportion of transitions (red dots) and transversions (blue dots) on mitochondrial CO1 and ND2 genes.

a.



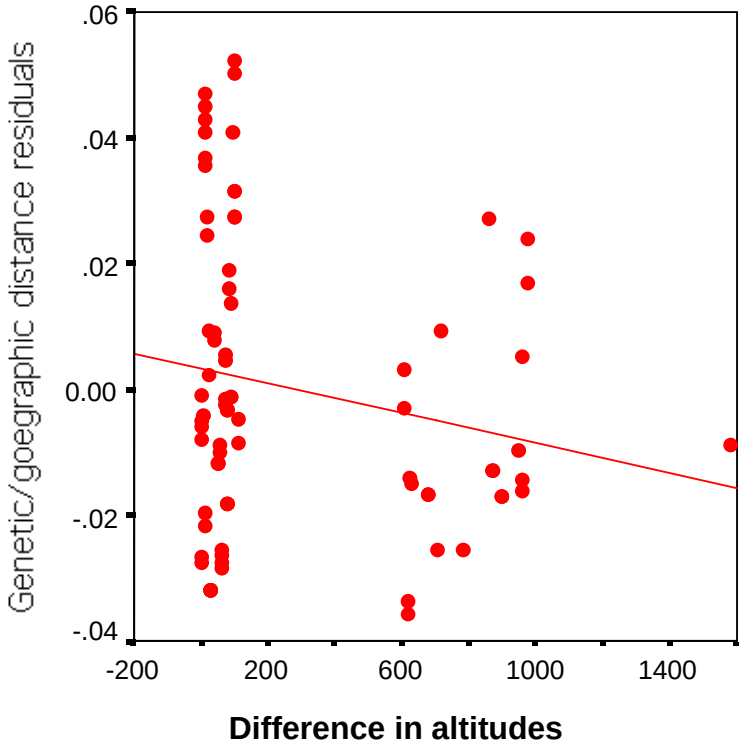
b.



Geographic distance (km)

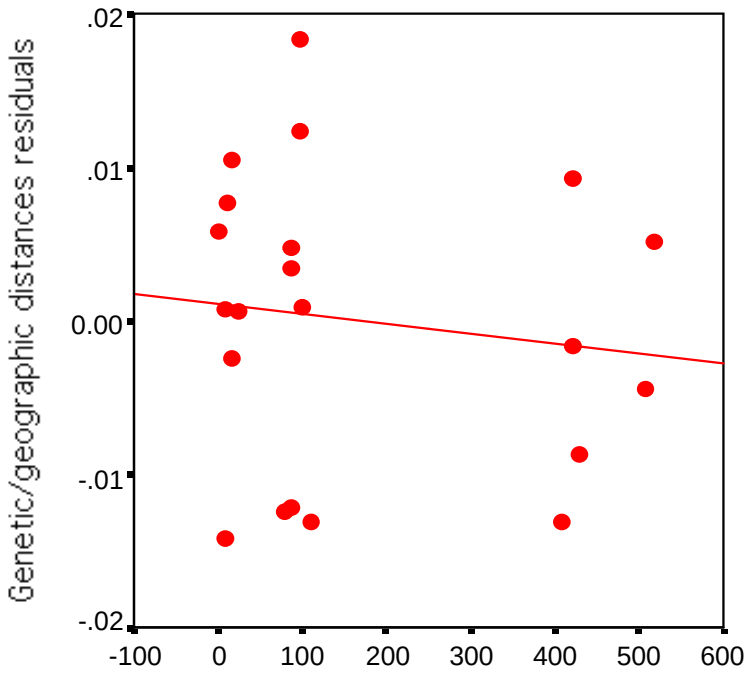
Fig. 3. Relation between geographic distance and Maximum likelihood genetic distances between all pairs of *Hyla labialis* populations for a. CO1 gene and b. ND2 gene. Blue dots denote intracluster (south-south, north-north) pairs of populations, black dots denote intercluster (north-south) pairs of populations.

a.



Rsq=0.0633

b



Rsq=0.0169

Fig. 4. Relationship between all pairs of *Hyla labialis* populations altitude and genetic distance on a. CO1 gene and b. ND2 gene. In figure a are discriminated the intra a intercluster pairs of populations.

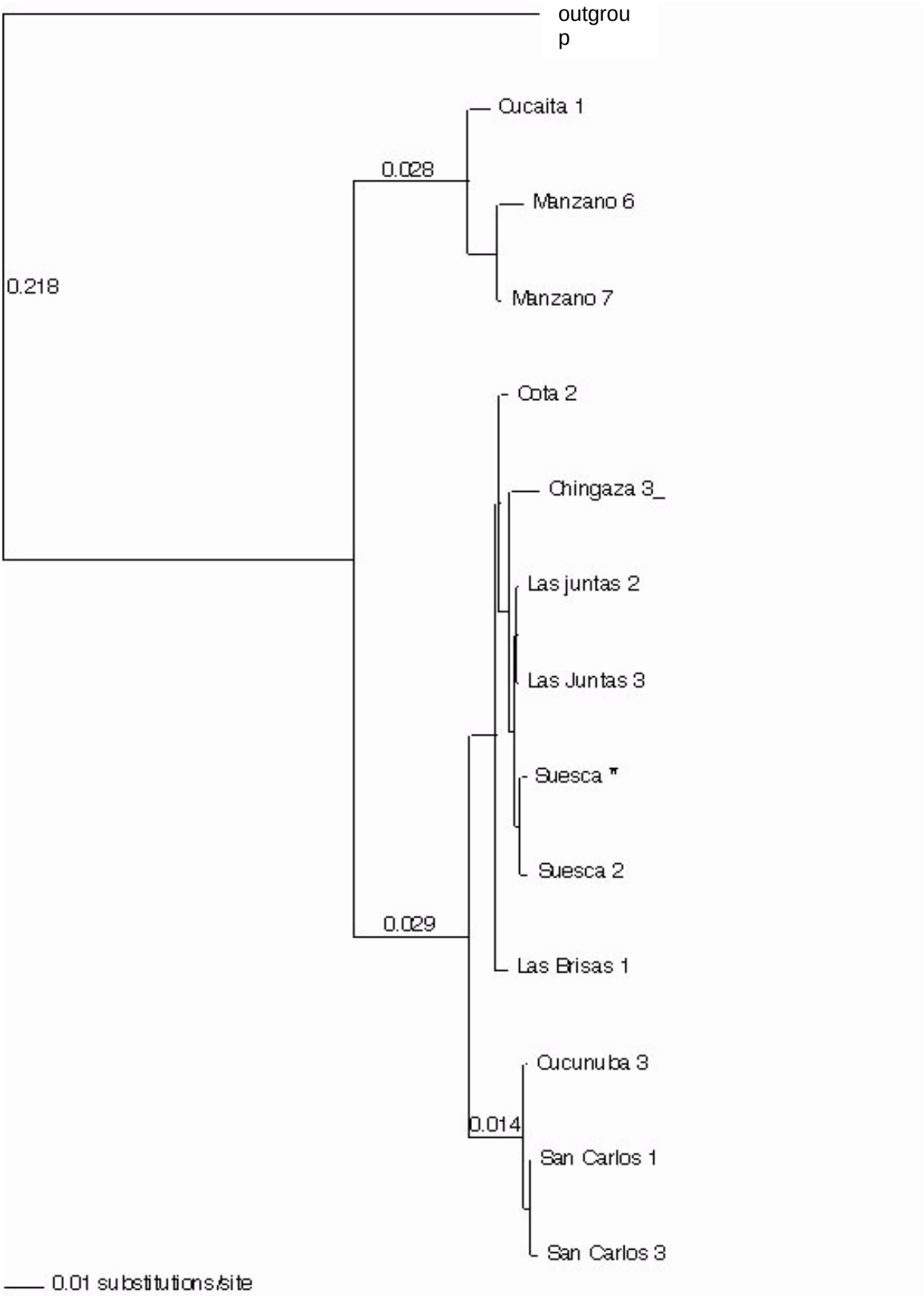


Fig. 5. Maximum likelihood phylogram (HKY85 + G distribution distances) for the CO1 gene, showing the relationship between 9 populations of *H. labialis*. Numbers are proportions of nucleotide substitutions.

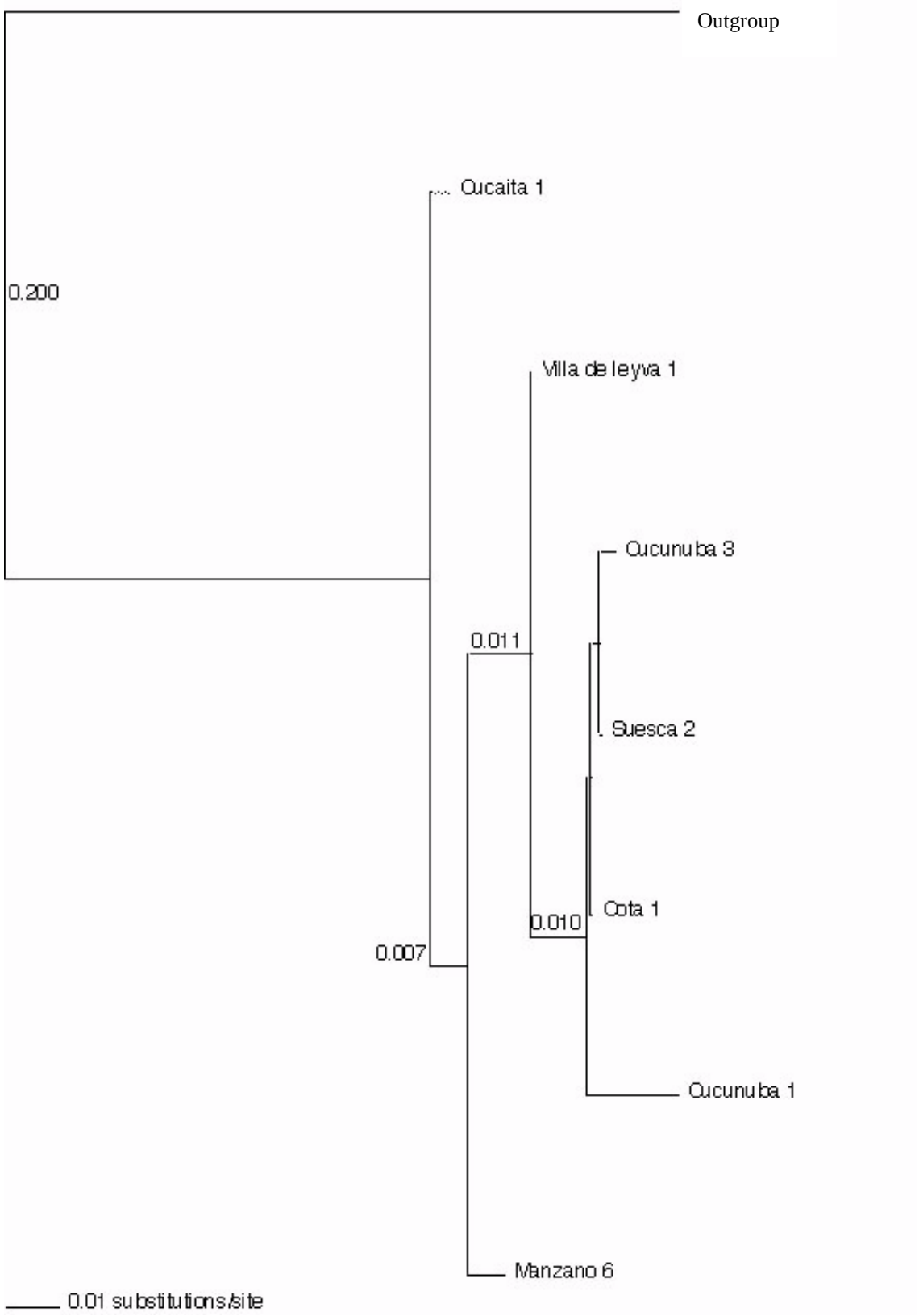


Fig. 6. Maximum likelihood phylogram (HKY85 + G distribution distances) for the ND2 gene, showing the relation between 6 populations of *H. labialis*. Numbers are proportions of nucleotide substitutions.

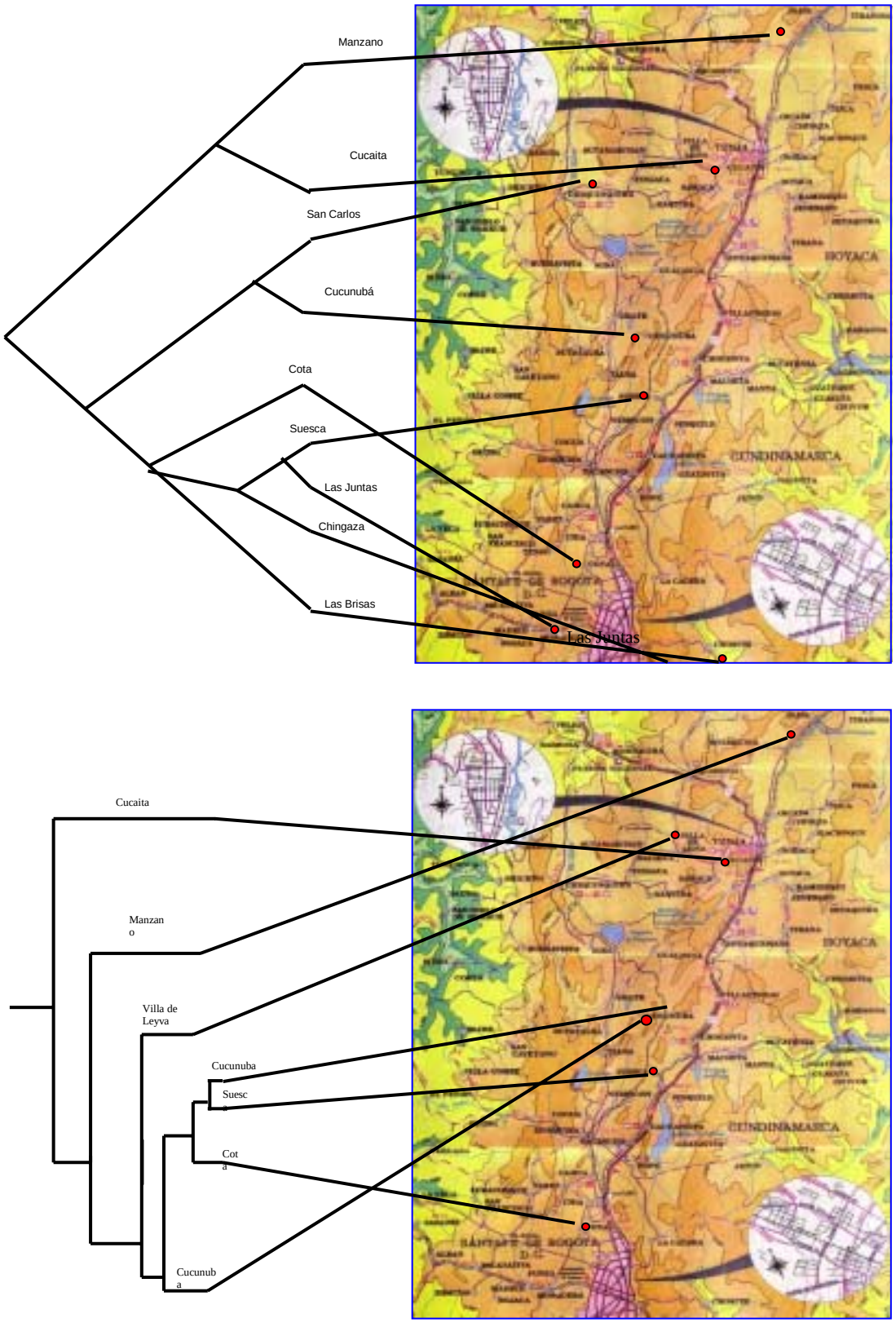


Fig. 7. Relationship between CO1 (a) and ND2 (b) gene genealogy topology and the geographic distribution of *Hyla labialis* populations.