

## Chapter 2

# Coat Color is not an Indicator of Subspecies Identity in Colombian Woolly Monkeys

Sergio Botero and Pablo R. Stevenson

**Abstract** Woolly monkeys are severely threatened, and disagreement on their taxonomic status complicates conservation strategies. Two subspecies of woolly monkeys inhabit Colombia, but the genetics of their populations have not been studied. Using mitochondrial DNA sequences, we set out to estimate the level of gene flow between populations, and to corroborate their taxonomic position. We found two separate evolving units with limited levels of gene flow. However, their separation does not correlate with the existing subspecies distinction, which is based on pelage color. We, therefore, propose a genetic differentiation of the woolly monkey taxa and emphasize the importance of the detected inconsistency in subspecies differentiation based on coat color.

**Keywords** *Lagothrix lagothricha lugens* · Platyrrhini · Pleistocene refugia · Primate conservation · Neotropical primates · Atelidae · Atelinae

## 2.1 Introduction

Woolly monkeys (*Lagothrix*, Atelidae) are important seed dispersers in the ecosystems they inhabit (Di Fiore and Rodman 2001; Nishimura 1999; Peres 1996; Stevenson 2000), dispersing over one third of the effectively dispersed seed biomass in forests where they are abundant (Stevenson 2007). Woolly monkeys are threatened by habitat destruction and hunting, both for their meat and local pet trade (Boubli et al. 2008; Stevenson et al. 2008; Peres and Palacios 2007). The Andean populations are the most threatened, and this situation is further aggravated by uncertainty regarding the taxonomic status of its taxa (Defler 2004).

---

S. Botero (✉) · P. R. Stevenson

Laboratorio de Ecología de Bosques Tropicales y Primatología (LEBTYP),  
Universidad de los Andes, Carrera 1 No 18a-10, Bogotá, Colombia

S. Botero

Rockefeller University, 1230 York Avenue, Box 18, New York, NY 10065, USA

e-mail: sbotero@rockefeller.edu

P. R. Stevenson

e-mail: pstevens@uniandes.edu.co

T. R. Defler, P. R. Stevenson (eds.), *The Woolly Monkey*, Developments in Primatology: Progress and Prospects, DOI 10.1007/978-1-4939-0697-0\_2, © Springer New York 2014

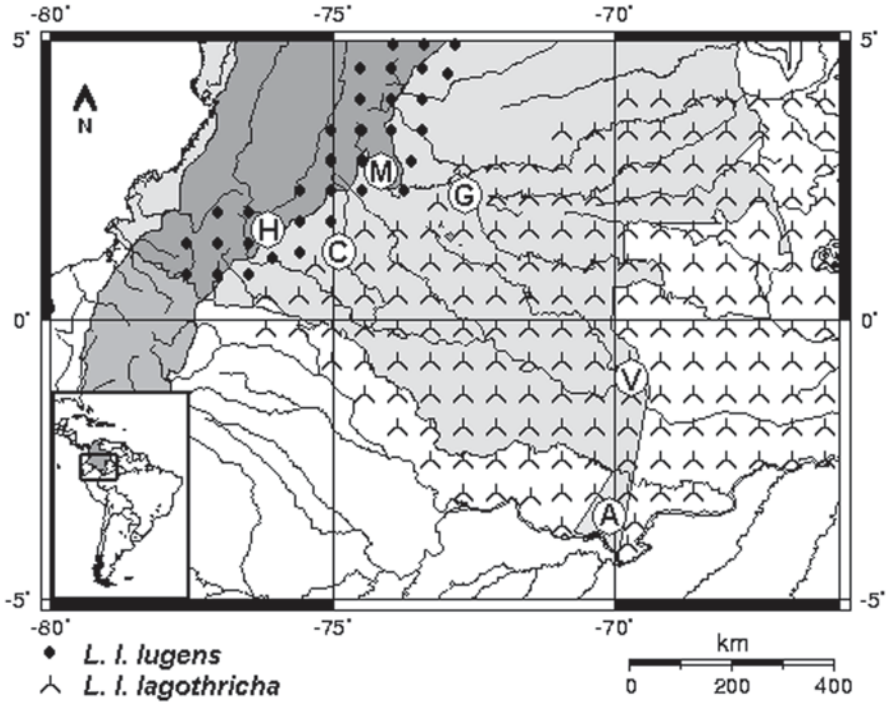
The taxonomy of the genus was originally revisited in 1963, when it was determined that woolly monkeys comprised two species, *Lagothrix flavicauda* and *Lagothrix lagothricha*. The latter was also subdivided into four subspecies: *L. l. lugens*, *L. l. lagothricha*, *L. l. cana*, and *L. l. poeppigii* (Fooden 1963). This taxonomy remained unchanged until 2001 when, based on morphological characters of museum specimens, all the taxa were raised by a level, separating the group into two genera: *Oreonax* and *Lagothrix*. *Oreonax* is a monotypic genus with *Lagothrix flavicauda* as its only species, and *Lagothrix* is now composed of four species: *L. lugens*, *L. lagothricha*, *L. cana*, and *L. poeppigii* (Groves 2001). The proposal of *Oreonax* has been shown to be an artifact of sampling (Matthews and Rosenberger 2008), and it is likely that the subspecies status is a better description of the remaining taxa (Defler 2004). To avoid ambiguity, we will use Fooden's (1963) taxonomy for the rest of the paper.

There are two taxa of woolly monkeys in Colombia (Fig. 2.1). *L. l. lagothricha* inhabits the south of the country in southern llanos, the Amazon region, and its range extends well into the Amazonas of neighboring countries. It has been classified by the International Union for Conservation of Nature (IUCN) as Vulnerable due to low natural densities, hunting pressure, and habitat degradation, although its wide range and presence in some pristine areas make it of minor concern when compared to other taxa in the group (Palacios et al. 2008). *L. l. lugens* inhabits the eastern and central cordilleras in the northern Andes and adjacent lowlands, showing the smallest distribution of all subspecies. It is likely extinct in Venezuela and is considered Critically Endangered due to both habitat degradation and hunting (Stevenson and Link 2008). The diagnostic difference between the taxa is a uniform brown color for *L. l. lagothricha* and a black to gray color for *L. l. lugens*, although significant variation in coat color is described in the classic revision (Fooden 1963) and recent reviews (Defler 2004).

With the use of molecular and cytogenetic markers, *L. l. lugens* and *L. l. lagothricha* were previously shown to be nonreciprocally monophyletic, rejecting their previously proposed species status (Botero et al. 2010). However, these analyses did not incorporate geographical information, precluding a detailed interpretation of the demographical processes between the subspecies, and the sample origin was assigned based exclusively on coat color. Here, we set out to corroborate these results, while including geographical information by sampling several of the Colombian populations of woolly monkeys.

## 2.2 Methods

Fecal samples of six populations of woolly monkeys were collected in 99% ethanol during a period from June 2008 to August 2009. We only report the samples that effectively amplified during the polymerase chain reactions (PCRs) performed, as the number of samples collected per population varied significantly. Sampling sites are indicated in Fig. 2.1. Sampling site coordinates, taxa, phenotype observed, and



**Fig. 2.1** Map showing the location of sampling sites in Colombia (country limits outlined in light gray). A, Amazonas; V, Vaupés; C, Caquetá; H, Huila; G, Guaviare; M, Meta. The major rivers (black) and the 700 m above sea level isocline (darkest gray inside of Colombia, dark gray outside) are outlined. Map obtained with the online GMT implementation. (Wessel and Smith 1991)

sample size are provided in Table 2.1. For the Estación Biológica Caparú population, samples were collected for a previous study unrelated to the current one. Coordinates should only be used as an approximation since the samples were gathered during extensive follow-ups and the coordinates given correspond to the location of the housing facilities in each site. In all cases, the coat color of the individuals sampled corresponded to that of the subspecies assigned based on their geographical location.

DNA was extracted using QIAamp DNA Stool Mini Kit (Qiagen) or UltraClean Fecal DNA Isolation Kit (MoBio) according to manufacturer protocols. After extraction, we amplified the hypervariable region I of the mitochondrial D-loop region in a total volume of 50  $\mu$ l using 1X reaction buffer, 2.5 mM magnesium chloride, 0.8 mM dNTP (Bioline) each, 0.5  $\mu$ M primer each, 2.5 uBiolase DNA Polymerase (Bioline), and 0.1  $\mu$ g/ $\mu$ l BSA (Promega). Thermal profile consisted of an initial denaturation at 94°C, 10 min; followed by 47 cycles of 94°C, 30 s; 60°C, 45 s; 72°C, 45 s; and a final extension at 72°C for 7 min, with primers L15400 (5'-TC-CACCATTAGCACCCAAAG-3') and H15940 (5'-CCTGAAGTCGGAACCA-GATG-3'), which had been used previously on atelines (Collins and Dubach 2000) and the nomenclature by Kocher et al. (1989) was followed. We checked successful

**Table 2.1** Sampling sites coordinates and sample sizes

Site	Coordinates	Taxa	Coat color	N	Code in figures
Serranía de la Macarena Parque Nacional Natural (PNN) Tinigua in the eastern cordillera piedmont in Meta	2° 37'N, 74° 4'W	<i>L. l. lugens</i>	Gray	17	M
PNN Cueva de los Guacharos, at the base of the eastern cordillera in Huila	1° 36'N, 76° 6'W	<i>L. l. lugens</i>	Gray	11	H
Caquetá on the western bank of the Caguan river, in the eastern cordillera piedmont. This is a small remnant population in a heavily degraded landscape	1° 20'N, 74° 53'W	<i>L. l. lugens</i>	Gray	9	C
PNN Amacayacu, on the northern bank of the Amazonas river, near the city of Leticia, in Amazonas	3° 23'S, 70° 9'W	<i>L. l. lagothericha</i>	Brown	16	A
Granja de Investigaciones "El Trueno," Instituto SINCHI, at the border between the Orinoquia and Amazonia region in Guaviare	2° 22'N, 72° 41'W	<i>L. l. lagothericha</i>	Brown	9	G
Estación Biológica Caparú in Vaupés	1° 4'S, 69° 30'W	<i>L. l. lagothericha</i>	Brown	10	V

amplification of a ~490-bp band on 1 % agarose gels, and sequenced positive reactions though Macrogen Korea commercial service. Only the sequencing reactions were outsourced; every other step in the processing of samples was performed in our laboratory. Sequences have been deposited in GenBank under accession numbers: GU212746-GU212756 Huila, GU212728-GU212736 Caquetá, GU212774-GU212783 Vaupés, GU212757-GU212773 Meta, GU212712-GU212727 Amazonas, and GU212737-GU212745 Guaviare.

As a minimum, three different sequencing reactions, from different PCRs, were performed for each primer. We only report results on samples that showed high-quality amplifications to avoid including nuclear integrations of mitochondrial DNAs (NUMTs) in the study. Since the region studied does not code for a protein, it is not possible to translate the DNA sequence into an amino acid sequence to detect premature stop codons as possible indicator of a NUMT. However, due to the probabilistic nature of PCR, and the relatively high amount of mitochondrial genome copies per cell (about 1,000–10,000; Shadel and Clayton 1997), it would be highly unlikely to find only homozygous individuals should the primers used anneal on one or more NUMTs. This assumption should hold even if the primers preferentially align on the NUMTs, a scenario that must be contemplated given that they were derived to anneal on conserved regions of the D-loop (Collins and Dubach 2000). Considering the relatively high number of haplotypes found in our study, heterozygous individuals should be expected if NUMTs were included. We thus believe our data to be free of NUMTs, but the inclusion of NUMTs cannot be completely ruled out.

Sequences were aligned using ClustalW (Larkin et al. 2007), as implemented in Bioedit (Hall 1999), and trimmed at the ends for a final sequence length of 431 nucleotides which included no gaps. For phylogenetic analyses, we performed maximum parsimony (MP) and maximum likelihood (ML) heuristic searches in PAUP\* 4.0b10 (Swofford 2003). Parameters of the searches were 1,000 random addition sequence replications keeping ten trees per cycle for MP, and ten random addition sequence replications keeping ten trees per cycle using the TPM1uf + I model, as selected using jModelTest (Posada 2008; Guindon and Gascuel 2003) with the Bayesian information criterion (BIC). Bootstrap support was obtained with 1,000 replicates for MP and 100 for ML. We performed a Bayesian phylogenetic inference analysis in MrBayes 3.2.1 (Ronquist and Huelsenbeck 2003). The parameters used were seven Markov Chain Monte-Carlo (MCMC) runs, 1,000,000 steps long each, a burn-in fraction of 0.5, the TPM1uf + I model, and all others set to default. The use of seven chains ensures that swapping between them is high, effectively providing good sampling of the space. For comparison of the tree topologies obtained, we used a Shimodaira–Hasegawa (SH) test (Shimodaira and Hasegawa 1999) with 1,000 bootstrap replicates, as implemented in PAUP\* (Swofford 2003). We did not use out-groups for the analysis because the high mutation rate of the mitochondrial marker used would confound the analysis since saturation is likely to be a problem (Pesole et al. 1999). This was evident when an alignment was made using *Ateles* as an out-group, requiring several gaps and showing regions in which it was impossible to find an unequivocal alignment.

For demographic analysis, we used the program DNAsp v5.00.07 (Rozas et al. 2003) to estimate basic statistics as well as  $F_s$  (Fu 1997) and  $R_2$  (Ramos-Onsins and Rozas 2002), with a coalescent confidence interval calculated with 10,000 replicates. We used the program Arlequin v3.1 (Excoffier et al. 2005) to perform an analysis of molecular variance (AMOVA), with 10,000 permutations, to determine the population structure. Pairwise  $F_{st}$  between populations were also calculated to estimate the effective number of migrants per generation ( $Nm$ ) with the equation  $F_{st} = (2Nm + 1)^{-1}$ , with 1,000 permutations to obtain a significance estimator. We tested the hypothesis of isolation by distance between populations with a Mantel test in the same program, using a distance matrix calculated in hundreds of kilometers (100 km) and 10,000 permutations. We constructed a haplotype network in the program Network (Fluxus engineering), using a median-joining algorithm (Bandelt et al. 1999).

We used a coalescent approach to estimate the time of separation, and the level of gene flow between the subspecies. Specifically, we tested for an isolation with migration model (Nielsen and Wakeley 2001), as implemented in the program IM (Hey and Nielsen 2007). This model assumes each population is constant in size. For this analysis, we used ten MCMC chains of 70,000,000 steps, sampling every 100 steps, with a burn-in of 350,000 steps, and a geometric heating scheme with parameters  $h_1 = 0.8$  and  $h_2 = 0.9$ . After seven optimization runs, we defined values of 400 for the  $q_1$  parameter: 1 for the  $t$  parameter and maximum values of 2 and 1 for the  $m_1$  and  $m_2$  parameters, respectively. These parameters allowed us to achieve maximum sampling of the space near the peak values. We checked the probability plots and autocorrelation plots to evaluate the convergence of the parameters.

The mutation rate used for our time of separation analysis,  $19.4 \times 10^{-9}$  mutations per site per year, comes from the average D-loop extended termination-associated sequences (ETAS) domain (which encompasses the hypervariable region I) as estimated for closely related pairs of mammals (Pesole et al. 1999). The highly variable D-loop region is ideal to resolve the phylogeny of very closely related taxa (Whittaker et al. 2007), but its high mutation rate makes its use for a molecular clock complicated, and it can vary significantly between similar taxa (Pesole et al. 1999). Amplification of other mitochondrial markers, for which more precise mutation rates are known or could be determined, failed for our samples (for COI, COII, and cytB a large proportion of samples did not amplify or showed very noisy chromatograms suggesting amplification of NUMTs, data not shown). Given the lack of better data, we opted for using the average mutation rate determined by Pesole and coworkers, which is also quite close to that of the *Homo–Pan* comparisons (while smaller than that of the *Pan paniscus–Pan troglodytes* comparison; Pesole et al. 1999). The estimated values of separation time are thus an approximation and should be treated with caution.

We repeated the AMOVA and IM analysis (which are the only ones done between subspecies and not populations), grouping the Guaviare population with *L. l. lugens*, since we propose this as a better grouping (see results and discussion). We changed the IM  $t$  maximum parameter to 35, and the number of steps to 100,000,000, with a burn-in of 500,000 to achieve convergence and a large effective sample size. These parameters were determined after four optimization runs.

## 2.3 Results

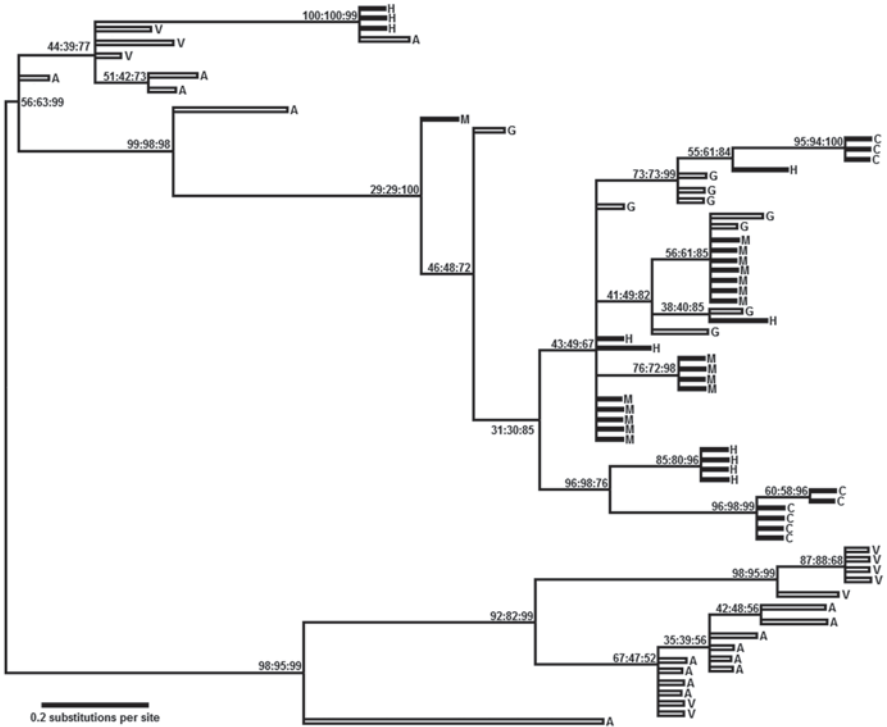
The number of haplotypes, the nucleotide diversity ( $\pi$ ), and the number of segregating sites for each population and subspecies are given in Table 2.2. Phylogenetic analyses did not show reciprocal monophyly between the subspecies and coat color does not differentiate any clades. The obtained tree topologies were similar and depicted alternative relationships were not rejected (SH test  $p > 0.05$  for all possible combinations). The phylogram obtained for the Bayesian inference analysis is shown in Fig. 2.2, indicating node support for the other analysis when nodes were shared.

Demographic analyses showed no evidence of population size changes for *L. l. lagothericha* ( $F_s = -2.19$ , 95 % confidence interval (CI):  $-7.73$ – $5.96$ ,  $P = 0.24$ ;  $R_2 = 0.18$ , 95 % CI:  $0.06$ – $0.17$ ,  $P = 0.99$ ), or *L. l. lugens* ( $F_s = 1.34$ , 95 % CI:  $-6.25$ – $6.54$ ,  $P = 0.73$ ;  $R_2 = 0.12$ , 95 % CI:  $0.06$ – $0.18$ ,  $P = 0.56$ ). This also holds for *L. lagothericha* as a species ( $F_s = -2.14$ , 95 % CI:  $-10.1$ – $9.51$ ,  $P = 0.31$ ;  $R_2 = 0.16$ , 95 % CI:  $0.05$ – $0.16$ ,  $P = 0.98$ ). AMOVA results showed no significant separation between the subspecies ( $F_{ct} = 0.28$ ,  $P = 0.30$ ), while significant population structuring did exist ( $F_{st} = 0.63$ ,  $P < 0.000001$ ). When the analyses are performed, including the Guaviare population in the *L. l. lugens* taxon instead of considering it as *L. l. lagothericha*, significant population structure ( $F_{st} = 0.70$ ,  $P < 0.000001$ ) as well as significant separation between the taxa ( $F_{ct} = 0.62$ ,  $P = 0.047$ ) are detected.

**Table 2.2** Descriptive genetic diversity statistics for the sampled populations

Population	Phenotype	N	Pi	Number of haplotypes	Number of segregating sites
Amazonas	Brown	16	0.029	11	38
Guaviare	Brown	9	0.007	7	9
Vaupés	Brown	10	0.029	6	27
Caquetá	Gray	9	0.015	3	13
Huila	Gray	11	0.025	6	26
Meta	Gray	17	0.005	4	7
<i>L. l. lagothericha</i>	Brown	35	0.043	24	50
<i>L. l. lugens</i>	Gray	37	0.017	12	32
<i>L. l. lagothericha</i> <sup>a</sup>	Brown	26	0.031	17	43
<i>L. l. lugens</i> <sup>a</sup>	Gray and brown	46	0.015	17	33

<sup>a</sup> Results are given for the conventional subspecies as well as for the new grouping we propose (see discussion)



**Fig. 2.2** Phylogram obtained in the Bayesian inference analysis. Node support is indicated when nodes are shared between trees obtained in the analysis (MP|ML|Bayesian inference posterior probability \*100). A, Amazonas; V, Vaupés; C, Caquetá; H, Huila; G, Guaviare; M, Meta. *Solid thick* terminals represent gray phenotypes, *while hollow* ones represent brown phenotypes. Note that the tree is unrooted

**Table 2.3** Estimated values of effective number of migrants per generation ( $N_m$ ) between populations included in the study

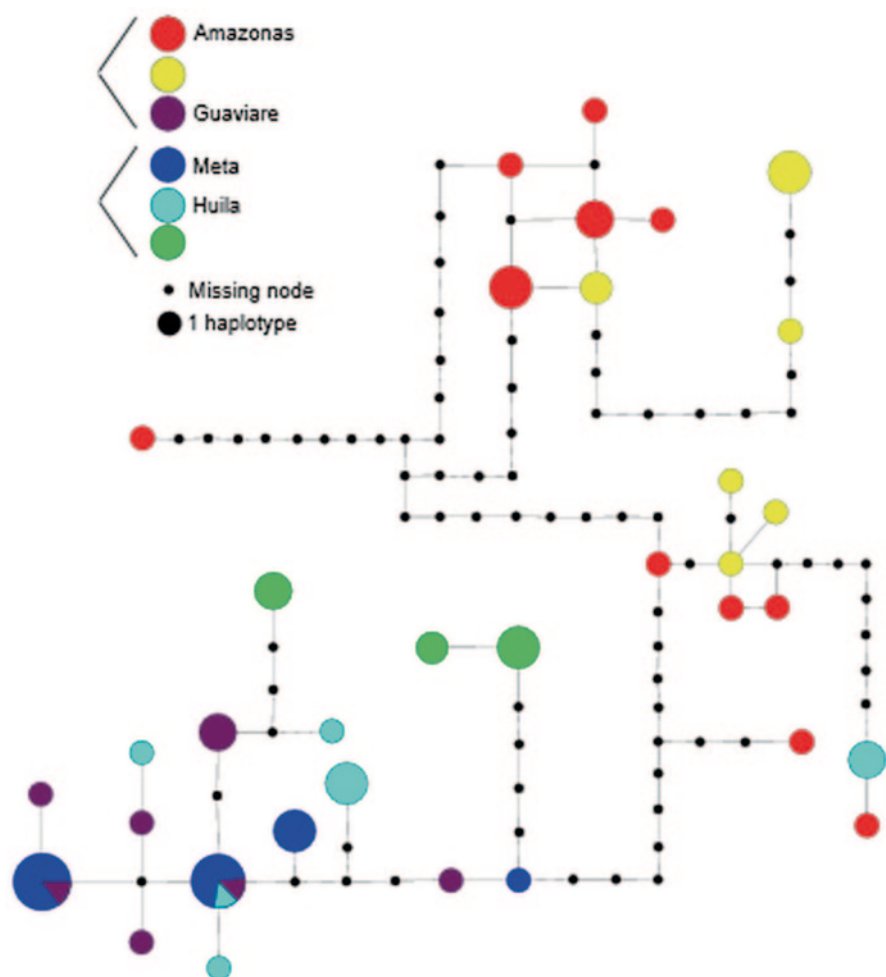
Population	Caquetá	Meta	Huila	Guaviare	Amazonas
Meta	0.53***	—	—	—	—
Huila	2.41*	1.28***	—	—	—
Guaviare	0.92**	6.29	2.20*	—	—
Amazonas	0.30***	0.18***	0.43***	0.23***	—
Vaupés	0.26***	0.14***	0.39***	0.19***	3.13*

Note the P values are derived from  $F_{st}$  calculations (see “Methods”), thus a nonsignificant value indicates enough migration between the populations as to make them undistinguishable genetically when analyzing their D-loop sequences ( $F_{st}$  nonsignificantly different from 0)

\*  $P < 0.05$ ; \*\*  $P < 0.01$ ; \*\*\*  $P < 0.001$

The obtained values of  $N_m$  are given in Table 2.3. These values come from the estimated  $F_{st}$  between pairs of populations, which showed significant results in all cases ( $P < 0.05$ ) except for the Meta–Guaviare populations, which would be undistinguishable. This approach to calculate  $N_m$  has drawbacks, but it is adequate when distance is assumed to be the main factor for population structure and  $N_m$  values are small (Holsinger and Weir 2009). There is considerable gene flow in populations of the same subspecies, with the exception of the *L. l. lagothricha* Guaviare population, which shows limited gene flow with other *L. l. lagothricha* populations while showing very high gene flow with *L. l. lugens* populations. Inside *L. l. lugens*, only the Meta population shows reduced gene flow with the Caquetá population. Mantel test showed a marginally nonsignificant correlation of the genetic distance matrix with the geographic distance matrix ( $r = 0.78$ ,  $P = 0.056$ , explained variance = 61%). The IM analysis estimated  $N_m$  value of 0.22 (95 % CI: 0.19–13.5) effective migrants per generation from *L. l. lagothricha* to *L. l. lugens* and of 5.6 (95 % CI: 0.91–51.4) effective migrants per generation in the opposite direction, which would indicate intermixing of the subspecies with a preferential gene flow from *L. l. lugens* to *L. l. lagothricha*. This is consistent with the levels of gene flow between pairs of populations estimated though the  $F_{st}$ . The relatively large confidence intervals obtained should not change the overall conclusion from the analysis since the trend of the values allows for a directional estimate of the migration, and a value of  $N_m$  higher than 0.5 is considered to indicate relatively high levels of gene flow (Hey 2010). Effective sampling size was above 800 for all parameters of the IM analysis.

When the Guaviare population is included as part of *L. l. lugens*, the estimated values of  $N_m$  also change, being now 0.016 (95 % CI: 0.05–4.1) effective migrants per generation from *L. l. lagothricha* to *L. l. lugens* and of 0.018 (95 % CI: 0.09–9.3) migrants per generation in the opposite direction. The distribution of the  $N_m$  values showed an ever-decreasing tendency, and this is reflected in the estimated value being outside of the 95 % confidence interval. Although this fact seems odd, it is to be expected, given that the confidence interval is calculated based on the posterior distribution of the values found, so the maximum value would always be outside of the confidence interval if the distribution is constantly decreasing. These values in-



**Fig. 2.3** Haplotype network for the sequences included in the analysis

dicate a very limited level of gene flow between the two subspecies if the grouping we propose is applied. Effective sampling sizes were above 2,500 for all parameters of the analysis.

After the optimization of the parameters, the probability plots and autocorrelation plots showed no abnormal behavior, suggesting a good sampling of the space in both IM analyses.

The constructed haplotype network is shown in Fig. 2.3. There is a general segregation of the two taxa, but the *L. l. lagothericha* Guaviare population is clearly mixed with the *L. l. lugens* populations. There are also some *L. l. lugens* individuals from the Huila population that are more closely related to the *L. l. lagothericha* populations. Overall, the network is consistent with the levels of gene flow and structure calculated.

The IM analysis estimates the time of separation of the subspecies including populations of Amazonas, Vaupes, and Guaviare in *L. l. lagothericha* and populations Meta, Caqueta, and Huila in *L. l. lugens* to be 50,649 years ago (95% CI: 15,368–116,547) and 1,831,332 years ago (95% CI: 546,260–4,095,904) when the Guaviare population is included in *L. l. lugens*.

## 2.4 Discussion

The absence of detectable demographic changes in the data represents a good starting point for other analyses, such as the Nm calculations, that assume a constant population size. It should be noted that this reflects an absence of population size changes several generations ago. Recent anthropogenic effects would not be detected under this analysis and are beyond the scope of this project. Our estimates of the migration level between populations seem to be realistic since we used a mitochondrial marker and, although there is some evidence of limited male dispersal in the species (Di Fiore and Fleischer 2005; Maldonado and Botero 2009), dispersal in the species is almost entirely by females (Nishimura 2003).

The observed absence of monophyly between the taxa evaluated is consistent with previous results that support a subspecies scheme (as opposed to separate species) in which there is some level of intermixing (Botero et al. 2010). However, that observation alone cannot distinguish between incomplete lineage sorting and true intermixing between the populations. The isolation with migration model used is ideal to approach this problem (Hey and Nielsen 2004; see below).

Estimated migration values show a high level of gene flow between the Caquetá and Huila populations, which is to be expected given their geographical proximity. It was unexpected to find *L. l. lugens* as far into the lowlands as we did in the Caquetá population, given the previously defined boundaries (Fooden 1963; Hernández-Camacho and Cooper 1976). A detailed study to determine the precise limits between the taxa is thus necessary. The Huila population also shows a high level of gene flow with the Meta population, while the Caquetá population shows reduced gene flow with the latter. High levels of gene flow are observed between the *L. l. lagothericha* Amazonas and Vaupés populations, but the Guaviare population shows limited migration with these two populations while showing high levels of migration with the *L. l. lugens* populations. This is unexpected since the distribution of *L. l. lagothericha* appears to be continuous (Hernández-Camacho and Cooper 1976), and the phenotype of the Guaviare population is that of that subspecies.

The Mantel test results show a marginally nonsignificant relationship, which would agree with a proposed limited area of contact only in the Caquetá cordillera slope (Hernández-Camacho and Cooper 1976; Fooden 1963), since geographic distance would not be a good indicator of gene flow between the populations. However, the high levels of gene flow between the Guaviare and Macarena populations strongly suggest that this is not the case. A detailed analysis including several populations from Orinoquia and northern Amazonia regions would be necessary

The Woolly Monkey

Behavior, Ecology, Systematics, and Captive Research

Defler, Th.; Stevenson, P.R. (Eds.)

2014, X, 302 p. 70 illus., 33 illus. in color., Hardcover

ISBN: 978-1-4939-0696-3