MtDNA investigation of the Cayapa (Chachi), an indigenous population of Ecuador with high record of the American founder lineage D4h3a

Martin Bodner[1*], Gabriela Huber[1], Bettina Zimmermann[1], Catarina Xavier[1], Reinhard Würzner[2], Walther Parson[1,3]

[1]Institute of Legal Medicine, Innsbruck Medical University, Innsbruck, Austria
[2]Division of Hygiene and Medical Microbiology, Department of Hygiene, Microbiology and Social Medicine, Innsbruck Medical University, Innsbruck, Austria
[3]Penn State Eberly College of Science, University Park, PA, USA

Introduction

Native American mtDNA diversity is characterized by four common "pan-American" haplogroups (A2, B2, C1, D1) and several minor lineages with limited, sometimes enigmatic dispersal [1-4]. After the rough picture of American colonization has been clarified in the last decades, the focus of Paleo-Indian mtDNA population genetics has moved towards high-resolution analyses of geographically restricted areas, single tribes, and minor or local lineages that may convey more detailed insights on (additional) migratory waves and routes in the peopling of the Americas [cf. 5].

We here pursue this approach by investigating the mtDNA composition of the Cayapa (Chachi) people from Northern Ecuador, for which an Amazonian origin, migration into the Andean highlands and later the Cayapa River basin in Esmeraldas province (Fig. 1), where they live today, to evade Incas and Europeans have been suggested. The Cayapa have remained stable in size (today: >3000 individuals), and little admixture is reported despite close proximity to African-American population groups since the 19th century [6-8]. MtDNA haplogroup D4h3a, first reported from partial control region data in 1999 as the "Cayapa lineage" [8], present at a high percentage in this population but rare in others, has meanwhile been confirmed as a 'minor' American founder lineage [3,9].

With 120 mtDNA sequences generated according to highest forensic standards, we present the first large complete mtDNA control region sample for the Cayapa, thereby also augmenting the scarce Ecuadorian mtDNA reference data.

Material and Methods

Total DNA was extracted from 40-50 µl blood plasma, collected and prepared in 1990-91, using the QIAamp Circulating Nucleic Acids Kit (QIAGEN, Hilden, Germany). After simultaneous verification of integrity and quantitation of mtDNA in a real-time quantitative assay [10], we completely Sanger-sequenced the mtDNA control region of 120 Cayapa samples applying ten primers on single PCR amplicons, according to highest forensic laboratory and data analysis quality standards [11,12]. We interpreted the haplotypes considering the established mtDNA phylogeny [13, build 15].

References

[1] Tamm E et al., PLoS ONE 2007; 2(9):e829.
[2] Achilli A et al., PLoS ONE 2008; 3(3):e1764.
[3] Perego UA et al., Curr Biol 2009; 19(1):1-8.
[4] Perego UA et al., Genome Research 2010; 20(9):1174-1179.
[5] Bodner M et al., Genome Research 2012; 22(5):811-820.
[6] Barrett SA, The Cayapa Indians of Ecuador. Heye 1925.
[7] DeBoer W, *in* Stahl, PW: Archaeology in the lowland American tropics. Cambridge University Press 1995: 243-262.
[8] Rickards O et al., Am J Hum Genet 1999 ;65(2):519-530.
[9] Kemp BM et al., Am J Phys Anthropol 2007;132(4): 605-621.
[10] Niederstätter H et al., FSI:Genetics 2007;1(1):29-34.
[11] Parson W & Bandelt HJ, FSI:Genetics 2007;1(1):13-19.
[12] Bandelt HJ & Parson W, Int J Legal Med 2008;122(1):11-21.
[13] van Oven M & Kayser M, Hum Mutat 2009; 30(2):E386-E394.
[14] Cardoso S et al., Heredity (Edinb) 2012; 108(6):609-615.
[15] Merriwether DA et al., Philos Trans R Soc Lond B Biol Sci 1999; 354(1379):121-129.

Figure 1. Origin of the Cayapa sample
The asterisk indicates the sampling area in the Cayapa river region of Esmeraldas province (marked yellow) in Northern Ecuador. The Cayapa river is indicated in blue.

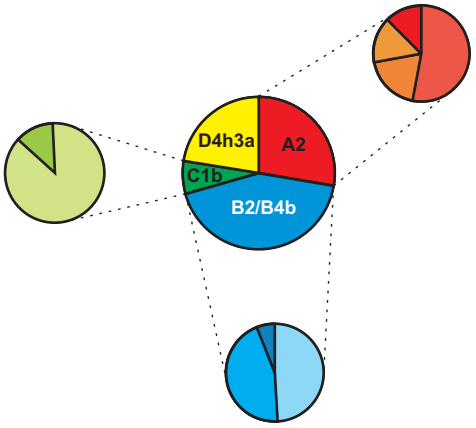

Figure 2. MtDNA diversity of the Cayapa
The central pie chart indicates the proportions of the four mtDNA lineages found in the Cayapa sample (n=120). The three peripheral pie charts depict the proportions of the haplotypes found within the segment they are assigned to. There is only one D4h3a haplotype. Range analyzed: control region. Point heteroplasmies, cytosine insertions after nps 309, 573, and 16193, and $(AC)_n$-variation around nps 523-524 not considered.

Table 1. Diversity parameters of the Cayapa sample calculated disregarding point heteroplasmies, cytosine insertions after nps 309, 573, and 16193, and $(AC)_n$- variation around nps 523-524. Range analyzed: control region.

	number of samples	haplotypes	RMP[a]	POD[b]
entire dataset	**120**	**10**	**0.159**	**84.10%**
subset haplogroup A2	34	4	0.344	65.60%
subset haplogroup B2/B4b	50	3	0.447	55.30%
subset haplogroup C1	9	2	0.654	34.60%
subset haplogroup D4h3a	27	1	1.000	0.00%

a... random match probability
b... power of discrimination

Results

Our samples classified into four groups. We found three of the 'major' Native American founder lineages - A2 (28.3%), B2 (including B4b) (41.7%), and C1b (7.5%), while D1 was absent. The "Cayapa" lineage [8], D4h3a, was confirmed in the expected high proportion (22.5%). We revealed yet undescribed variation possibly indicating novel sub-clades within all four founder lineages. However, variation was limited: just four haplogroup A2, two B4b (likely B2), one B2c2, two C1b and a single D4h3a haplotype(s) were present (Fig. 2), which is also reflected in a high random match probability of overall 1:6.3 and 1:1 to 1:3 for the four subsets. The resulting forensic parameters are shown in Table 1. Despite ongoing contact to European conquestors and African-American populations that inhabit the same area, neither Eurasian nor African maternal contribution was detected in the Cayapa sample.

Discussion and Outlook

While the Cayapa mtDNA pool is characterized by much more diversity than that of another Ecuadorian Native tribe recently described [14], still only the rather limited number of ten control region haplotypes was revealed. This can possibly be explained by bottlenecks in population size that occured during or as consequence of the migrations reported in the history of the Cayapa, where only a limited number of women may have acted as founders for the population in the new settlement area.

This dataset from a scarcely described region and population contributes to the refinement of the phylogeny and phylogeography of the ubiquituos 'major' American founder lineages, and the widespread, but usually rare founder haplogroup D4h3a that is frequent in the Cayapa. Further phylogenetic studies using complete mitogenome sequences will show whether the previously undescribed control region polymorphisms detected truly represent novel lineages and where these lineages are dispersed.

Comparative mtDNA (control region) studies, including other Native American and mixed urban populations from regions (postulated to have been) inhabited by the Cayapa, will likely reveal additional details of human migration history into and within (Northern) South America. Our findings confirm the necessity of ongoing sequencing efforts at a regional scale to yield the complete picture of American pioneer colonization. Still, major sampling and sequencing efforts are mandatory for uncovering all of the most basal variation in the Native American mtDNA haplogroups.

Our results confirm the meaning and usefulness of "old" samples [15], when the necessary care in storage and analyses is applied.

***contact:** martin.bodner@i-med.ac.at

Acknowledgements
The authors wish to thank the donors for giving their blood for research and Christiane Maria Bauer, Beate Beer and Harald Niederstätter (all: Institute of Legal Medicine, Innsbruck Medical University) for valuable technical assistance and discussion. This work received support from the Austrian Science Fund (FWF): L397 and the Theodor Körner Fonds zur Förderung von Wissenschaft und Kunst.