

Tracing Arawakan Southern Dispersal: Clues from Mitogenome Sequencing of Southern Amerinds



C. Xavier¹, F. Simão², M. Bodner¹, C. Strobl¹, D. Silva², C. Barletta³, A. Casas-Vargas⁴, W. Usaquén⁴, D.H. Tineo⁵, L. Gusmão^{2,6} and W. Parson^{1,7}



¹Institute of Legal Medicine, Medical University of Innsbruck, Innsbruck, Austria
²DNA Diagnostic Laboratory (LDD), State University of Rio de Janeiro (UERJ), Rio de Janeiro, Brazil
³Laboratorio de Genética Humana, Universidad Nacional Mayor de San Marcos, Lima, Peru
⁴Group of Population Genetics and Identification, Genetics Institute, National University of Colombia, Bogotá, Colombia
⁵Instituto de Medicina Legal, Universidad Nacional Mayor de San Marcos, Lima, Peru
⁶Institute of Pathology and Immunology of the University of Porto (IPATIMUP), Porto, Portugal
⁷Forensic Science Program, The Pennsylvania State University, PA, USA
 catarina.gomes@i-med.ac.at

P03-202
170320

Introduction

South America's colonization and internal migration routes remain unsolved issues for the scientific community. The wide variety of different environmental settings and the amplitude of climate variations at the time of colonization might explain the big variation of different tribes and groups in South America, in contrast to its northern counterpart [1].

The Arawakan are the largest and most widespread linguistic family in South America, being distributed at time of European arrival from the Caribbean Islands to the Brazilian wetlands. The Arawakan were also a distinct Amazonian group which would stand out from other Amazonian tribes by their advanced socio-cultural societies, the agricultural and fishing economic practices as well as complex commercial and political alliances [1, 2]. Linguistic and anthropological studies situate western Amazonia (in the border between Amazonian Peru and Brazil) as the most probable origin of the Arawakan language (Fig. 1, [3]). Afterwards Arawakan are thought to have migrated alongside riverbanks throughout the continent, spreading their cultural and agricultural traditions, naming their diaspora the South America's Neolithic (Fig. 1, [3]).

For this study we analysed samples from three different Arawakan communities: the Wayúu in northern Colombia, the Ashaninka in pre-Andine Peru and the Terena in the Brazilian wetlands. Furthermore we gathered a sample set from a Shipibo-Conibo tribe, inhabiting the shores of the Ucayali river in Peruvian Amazon. The Shipibo are a Panoan speaking tribe which engaged in several warfare encounters and intermarriage with Arawakans (Piro) living in the same region [4]. All samples have been typed for the full control region of the mitochondrial DNA (mtDNA), the information of the found lineages allowed a triage of samples to undergo the full mitogenome sequencing.

A deeper study of under-described populations as well as the increase of the study's resolution allows a better comprehension of the Arawakan migratory movements at a minor scale as well as increase the size of a forensic and population mtDNA database - EMPOP [5].

Materials & Methods

The study was approved by Ethics Commissions. Sampling was performed under informed consent. All samples were extracted using a phenol-chloroform protocol or a Chelex based protocol and quantified using an in house real time qPCR method [6]. Libraries were built using the Precision ID mtDNA Whole Genome Panel (Thermo Fisher Scientific) and sequenced using the IonChef and IonS5 system (TFS). A validation study of this method is presented in P01-23.

The sequences were compared to the revised Cambridge Reference Sequence (rCRS) [7] using the Ion Torrent Variant Caller v. 5.2.1.38 plugin (TFS) and all these positions have been posteriorly checked with IGV (Integrative Genome Viewer). Haplotypes were determined by two separate analysts and haplogroups were assigned according to Phylotree (build 16) [8]. Furthermore, all haplotypes have been submitted to EMPOP4 (P06-422) for alignment and haplogroup check.

In order to understand the distribution of the lineages found in this study and close neighbors, literature research and database queries were done.

Discussion

The previous sequencing of the full CR of the mtDNA helped us detecting one lineage (B2+16051G+16360T+152C) that was shared between the Terena and the Shipibo-Conibo (Fig. 2). The deeper investigation of this particular lineage at the highest resolution possible (the full mtDNA genome) verifies the lineages present in Peru and Brazil are very similar (Fig. 2). A thorough look into literature identified the existence of the same lineage in Peru (Fig. 1). If only the CR or partial CR is investigated the same lineage appears along the Peruvian lowlands, the Titicaca basin and the Bolivian Moxos.

Furthermore, we identified a closely related lineage present in the Peruvian Ashaninka (B2+16051G+16129A+152C) that was further sequenced for the full mtDNA genome (Fig. 2). The same lineage (full genome) was found in other Peruvian samples and previously described (only CR) in Matsiguenga and Yanasha tribes, all pre-Andine Arawakans (Fig. 1). This might indicate a regional diversification of this specific lineage which also agrees with the linguistic differentiation between pre-Andine Arawakan and Proto-Piro-Apurina-Baure-Ignaciano Arawakan (western Amazonian lowlands and Southern Arawakans) in which Terena and the neighbor groups of Shipibo (the Piro) are included.

Other lineages have been chosen for further full mtDNA genome sequencing, either because they have not yet been described in the literature or presented a typical motive in their geographical panorama. For example, B2d lineages in Colombia have been identified previously and seem to be restricted to northwestern South America and Mesoamerica territories [9].

These results reinforce the fact that intense trading systems and alliances with neighbor groups with strong intermarriage policies have shaped the Arawakan genetic panorama. However, particularly in western Amazonia / Bolivian Llanos and Wetlands the B2 + 16051G + 16360T + 7786C + 152C stands out as a possible reminiscence of an Arawakan southern migration. Characterized as riverine groups with developed watercraft techniques, a southern migration alongside the Ucayali and Urumbamba rivers until reaching Madre de Dios River and finally the Pantanal wetlands in Brazil would not be surprising.

Results

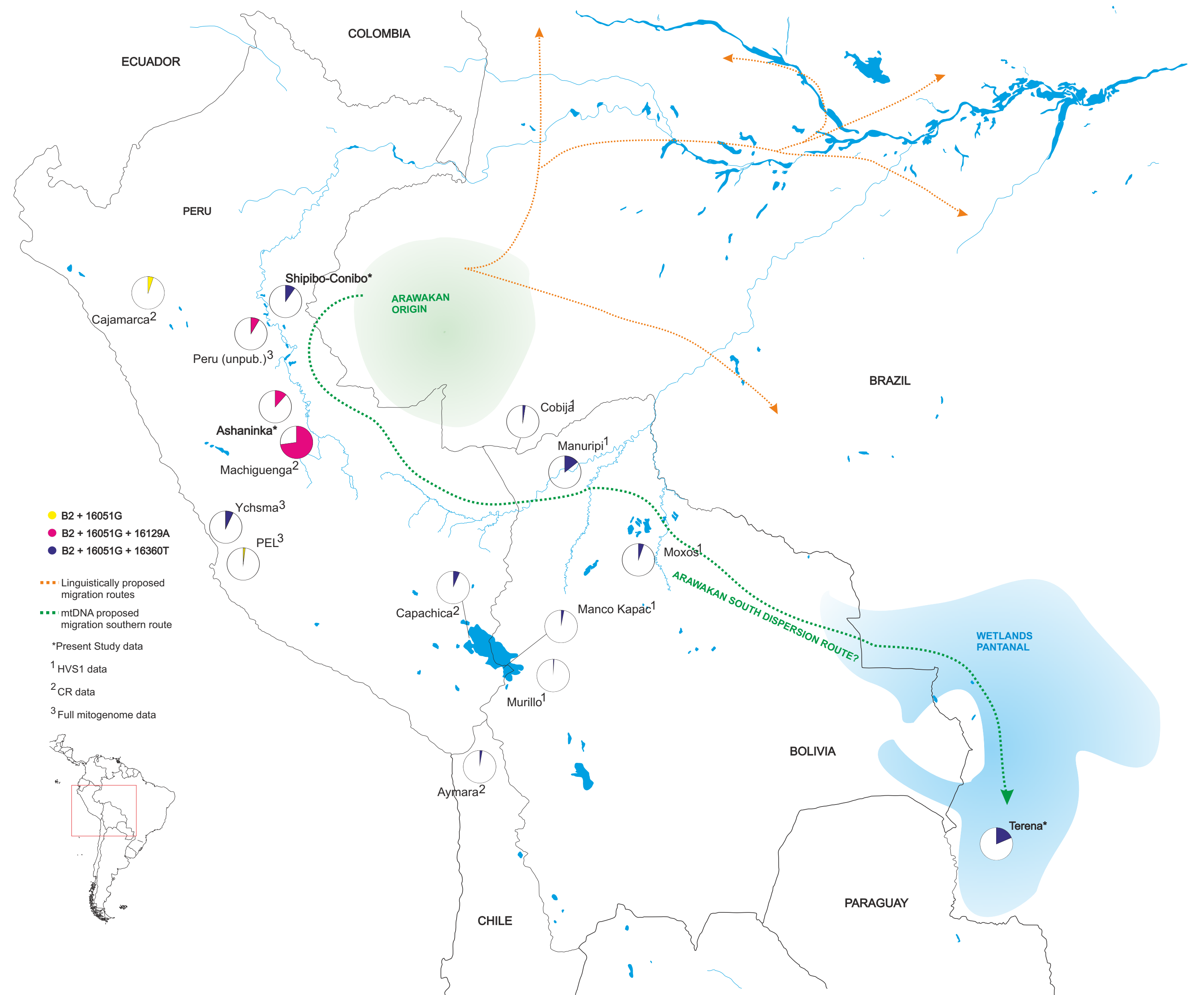


Fig. 1 B2+16051G (yellow), B2+16051G+16129A (pink), B2+16051G+16360T (dark blue) lineages distribution across the Peruvian lowlands, lake Titicaca, Bolivian Moxos and Brazilian Wetlands. Samples in bold and with a * are the present study data, other groups have been gathered from literature [all of them]. Orange arrows represent the linguistically proposed migration routes by the Arawakan, the green arrow is the migration route proposed by the mtDNA results. Shaded green area is the linguistic proposed location of origin of the Arawak language, shaded blue area represents the wetlands covered area. Other populations' data retrieved from [10-15].

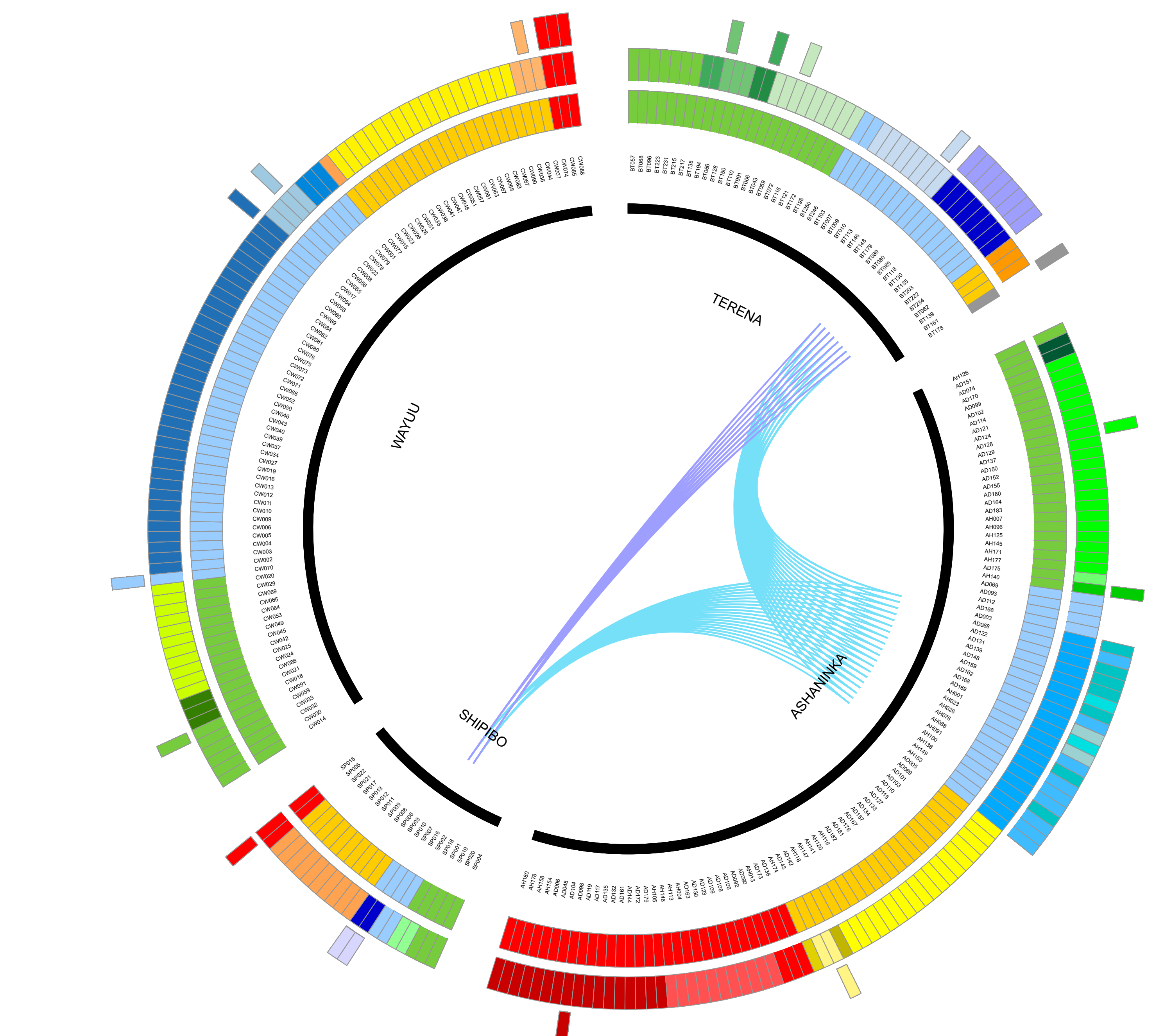


Fig. 2 Graphical representation of the mtDNA lineages found in the four studied groups. Inner black circular segments represents each of the groups. Inner graphic shows the composition of mtDNA lineages for the four major Native American haplogroups and L2. Middle graphic shows variation within each macrohaplogroup and outer graphic depicts the samples chosen for full mtGenome sequencing and diversity found. Links between sample groups reflect close lineage(s) of B2+16051G in dark (+16360T+7786T) and light blue (+16129C). Plot designed using Ciroos [16].

References & Acknowledgements

- [1] Rothammer F & Dillehay TD (2009) Ann Hum Genet 73:540-549
- [2] Hill JD & Santos-Granero F (2002) University of Illinois Press
- [3] Walker RS & Ribeiro LA (2011) Proc. of the Royal Society (online pub.)
- [4] Lathrap D (1970) Thames & Hudson
- [5] Parson W & Dür A (2007) FSI:Genetics 1(2):88-92 empop.online
- [6] Niederstätter H et al (2007) FSI:Genetics 1(1):29-34.
- [7] Andrews RM et al (1999) Nature Genetics 23:147
- [8] Van Oven M et al (2009) Hum Mutat 30(2):386-394 www.phylotree.org
- [9] Xavier C et al (2015) PlosONE 10(3)
- [10] Sandoval et al (2013) PlosONE 8(9)
- [11] The 1000 Genomes Project (2015) Nature 526(7571):68-74
- [12] Valverde G et al (2016) PlosONE 11(6)
- [13] Tito RS et al (unpub)
- [14] Taboada-Echalar P et al (2013) PlosONE 8(3)
- [15] de Saint Pierre M et al (2012) PlosONE 7(9)
- [16] Krzywinski, M et al. (2009) Genome Res 19:1639-1645

C. Xavier had, at the time of analysis, a PhD grant (SFRH/BD/90873/2012) from Fundação para a Ciência e Tecnologia (FCT). M. Bodner was supported by Theodor Körner Fonds zur Förderung von Wissenschaft und Kunst. The authors would like to thank Bettina Zimmermann, Gabriela Huber and Simone Nagl for technical help and analysis of the Sanger sequencing data previous to this work.

