

3.3

Genetic exchanges in the highland/ lowland transitional environments of South America

Chiara Barbieri

Introduction

Geneticists have often evoked the contrast between the Andean and Amazonian environments to explain the major patterns in the genetic structure of South America. Major differences, as already described in [Chapters 1.3](#) and [3.2](#), revolve around the ratio between the diversity within a given population, and around the diversity between different populations. In the Central Andes, populations are characterized by high genetic similarity to each other, but high genetic diversity between the individuals within a population; populations from the Amazon basin, meanwhile, are characterized by high differentiation between each other but low diversity across the individuals within a population. These contrasts have been interpreted in the light of different social dynamics playing out in the two environments: small isolated populations in the Amazon basin, and larger populations connected by gene-flow in the Andes (Tarazona-Santos et al. 2001; Fuselli et al. 2003; Wang et al. 2007; Dillehay 2009; Sandoval et al. 2016). Genetic contrasts between populations of the Andes and Amazonia include also a different composition of characteristic genetic lineages, such as uniparental haplogroups (on which see [Chapter 1.3](#), and the review in Bisso-Machado et al. 2012). These differences have been critical to demographic studies, which have proposed separate routes for the first settlement of the continent (Keyeux et al. 2002; Yang et al. 2010). Finally, genomic differences between populations of high and low altitude play a fundamental role in functional studies on how environmental constraints may have driven selection for specific biological adaptations (Beall 2014).

Few genetic studies, however, have addressed the circumstances of contact and exchange in regions transitional between the two major environments of Andes and Amazonia, which constitute the focus of this chapter. It presents four genetic

case-studies on the effect of contact and exchange between different ecological and cultural domains, highlighting limitations imposed by the respective population samples available and by the different genetic data chosen for the analyses.

Demographic studies that include genetic profiles of native populations have been focusing above all on uniparental markers, the DNA markers that are inherited on either the maternal (mitochondrial DNA, or mtDNA) or the paternal (Y-chromosome DNA) side (Chapter 1.3). Due to their transmission pattern they are suitable for reconstructing genealogies, and they are regarded as the gold standard for investigating phylogeography (that is, the distribution of phylogenetic lineages in specific regions of the world) and human migration and contact (Underhill et al. 2001; Pakendorf and Stoneking 2005; Torroni et al. 2006; Kundu and Ghosh 2015). For these markers, a large amount of data are available for inter-population comparisons. As a downside, when looking at the mtDNA or Y-chromosome we are limiting ourselves to a small fraction of the total DNA information carried by each individual, and we are considering only one ancestry line among the many that an individual bears. Deeper resolution is achievable with the use of autosomal data, which is still more demanding in terms of monetary and labour costs. As explained in Chapter 1.3 of this book, with the term autosomal we consider all the genetic material of our chromosomes (except the sex chromosomes) that is not transmitted solely on either the maternal or paternal side, but by virtually all our ancestors. Autosomal genomic data are more informative for fine-scale demographic reconstructions, but published data are still very few and far between for the populations of the Americas (Bustamante et al. 2011; Wall et al. 2011). Recent publications are improving the genomic coverage of the continent, revealing new sources of genetic diversity (Raghavan et al. 2015; Skoglund et al. 2015; Harris et al. 2018; Gneccchi-Ruscione et al. 2019).

The first two recent studies I examine here have made use of high-resolution autosomal data (SNP chip data). While the first employs a dataset that consists of only two populations, it is targeted towards research questions very much in line with the theme of this section. The second draws on a larger dataset, although oriented primarily to research questions on functional adaptation. Broader comparative datasets are included in the third and fourth case studies, based on mtDNA and Y-chromosome data. This chapter also includes a novel targeted comparative analysis that yields further insights into the questions already debated in the four case studies proposed. In its conclusions, it recapitulates the emerging major trends in the genetic make-up of populations inhabiting these transitional environments.

Autosomal data: A fine-grain resolution

A single Andean origin for Arawakan speakers of central Peru

The first case-study, based on autosomal data, looks into the origins of the Shimaá, a small population living in the transitional environment of the eastern slopes

of the Andes (or *Yungas* in some interpretations of that term) in central Peru. The Shimaa speak Machiguenga, a language of the Arawak family, and present cultural features typically found in neighbouring Amazonian regions (on Arawak and its distribution, see [Figure 1.2.1](#) in [Chapter 1.2](#), and [Chapter 3.4](#)). Scliar et al. (2014) compare the diversity within fragments of autosomal DNA across ten Shimaa individuals and 11 Quechua speakers from Tayacaja, 300 km to the west, chosen to represent a population from the highlands. The authors apply Bayesian statistical analysis and model-testing to explore the nature of the relationship between the two populations. The results provide strong support for a split between the two populations that would have taken place not more than 5,300 years ago. The authors conclude by suggesting that the ancestors of the Shimaa were a small group who separated from a wider Andean population: this inference is based on the lower diversity of the Shimaa individuals, who harbour only a subset of the genetic variants found in the Quechua sample. The authors therefore evoke a scenario in which the Shimaa migrated from the Andes to the lower slopes towards Amazonia and underwent a cultural/linguistic shift after coming into contact with Arawak speakers. Unfortunately, the lack of any other comparative autosomal data prevents the authors from evaluating the contact dynamics of the Andes–Amazonia transition zone more widely; in fact, with just two samples available, only one divergence model could be tested. It would be crucial to test an alternative model where the Shimaa diverged from an Amazonian population, but this would require additional data from neighbouring regions, and from other Arawak speakers in particular.

Extremely high altitude influences genetic differentiation

The second case-study, again based on high-resolution autosomal data, focuses on functional adaptation. The Andes make for a good scenario for testing the effects of natural selection, given the increasingly hostile environment at higher altitudes. To survive at extreme elevations, humans developed a number of biological adaptations to hypobaric hypoxia (see review in Beall 2014). Yet altitudes above 4,000 m appear to have been settled from the late Pleistocene onwards (Rademaker et al. 2014), giving thousands of years for adaptations to high altitude to develop.

Studies on functional genetics suggest that highlanders are in part genetically differentiated from lowlanders. Eichstaedt et al. (2014), for example, found traces of selection on genetic markers associated with cardiac reinforcement when comparing two neighbouring populations of north-west Argentina: the Wichí of the Gran Chaco who live below 1,000 m, and the so-called ‘Colla’ who live in the highlands above 3,500 m. (This present-day population that goes by the name ‘Colla’ is not to be confused with the ethnic group immediately west and south of Lake Titicaca during the rise of the Incas.) This example serves as a useful reminder of the role played by factors other than demography (in this case, high-altitude environments) in shaping human genetic diversity.

In a second study, Eichstaedt et al. (2015) analyse whether a population living at intermediate altitudes might also be affected by moderate levels of hypoxia. The Calchaquíes of north-west Argentina live at 2,300 m in a region intermediate between the Altiplano and the Chaco: this region served as a migration corridor during late Inca expansion. Both studies from Eichstaedt and colleagues compare autosomal SNP data from their target populations with other available South American populations. These are taken from the public databases of HGDP-CEPH and from Reich et al. (2012) and Mao et al. (2007), for a total of 19 populations; eight of these, however, have fewer than ten individuals each, making it difficult to represent the genetic make-up of the whole target population. In the population analysis by Eichstaedt and colleagues, the Calchaquíes present an ancestry component commonly found in the neighbouring ‘Colla’, as well as in other (Quechua- and Aymara-speaking) populations of Peru and Bolivia. The Wichí, meanwhile, present an ancestral component widely found in other populations of the Gran Chaco, such as the Toba and, to a lesser extent, the Guaraní. The marked genetic difference between the Calchaquíes, who appear similar to other Andean highlanders, and the Gran Chaco populations, who all harbour (albeit at varying percentages) an ancestral component exclusive to their region, was not unexpected (Frank 2008). The Calchaquíes were also interacting intensely with populations from higher altitudes, as Inca allies and colonists were moved into this territory from various regions including the Titicaca basin (Lorandi and Boixadós 1988). Finally, the Calchaquíes present a subset of the genetic adaptations to high altitude found in the Argentine ‘Colla’, although the origin of this genetic signal is difficult to assess: it could be a mild response to environmental stress, or simply the result of gene flow from inter-marriage with the ‘Colla’.

Uniparental markers: Larger comparative datasets

Turning to uniparental markers (mtDNA and Y-chromosome), there are certainly more South American populations for which we have data, especially in the Andes (Bisso-Machado et al. 2012). Only recently, however, have studies begun to abandon a compartmentalized ‘Andes or Amazonia’ vision, to focus on exchanges *between* the two environments, that is, both the contribution of Andean genetic lineages to Amazonia and vice versa (see Chapter 1.3). New colonization routes have been proposed to account for the distribution and phylogeny of certain characteristic maternal and paternal lineages (Perego et al. 2010, 2012; Bodner et al. 2012; Saint Pierre et al. 2012a, 2012b). In some cases, the migration hypotheses are justified by historically attested population movements that offer plausible explanations for the patterns observed today (Bodner et al. 2012), but more often these phylogeographic studies are focused on the genealogy of specific lineages, rather than on the prehistory of specific populations.

Here I report on two case-studies based on uniparental markers, which investigate the origin and demographic history of two populations who live(d) in transitional environments, geographically close to those covered in the autosomal studies above. These are the ancient populations of the Quebrada de Humahuaca in north-west Argentina, and the present-day Yanasha, another Arawak-speaking population in the Andes–Amazonia transition in central Peru.

A window into the past: aDNA from Argentina shows maternal (but not paternal) connections with the Gran Chaco

Mendisisco et al. (2014) analysed mtDNA and Y-chromosome data from archaeological remains found in the Quebrada de Humahuaca (Jujuy province) and in the neighbouring Calchaquí valley (Salta province). Ancient DNA (aDNA, see [Chapter 1.3](#)) was obtained from teeth dated AD 1000–1450, corresponding to the Regional Development Period (RDP). The Quebrada de Humahuaca is a valley in a strategic location between the Andean highlands (the Bolivian Altiplano and Argentinean Puna) and the eastern edges of the lowland forests and the Chaco. The region has been inhabited for at least 10,000 years and has long been characterized by a significant level of cultural, economic and social interactions, with relatively highly developed societies and dense populations (Nielsen 2001). The relationships between the ancient population of the Quebrada de Humahuaca and other ancient and present-day South American populations were explored through both the maternal (mtDNA) and paternal (Y-chromosome) lines. The mtDNA profile of the Quebrada de Humahuaca shows a high percentage of haplogroup A2, a lineage otherwise frequent in populations of *northern* South America, in the Guianas, and in some scattered populations of the Amazon basin (Bisso-Machado et al. 2012). This high frequency is unusual for this region, found neither in surrounding contemporary populations nor in ancient Andean samples. In fact, ancient and contemporary Andean highland samples are instead characterized by high frequencies of haplogroup B2 (Bisso-Machado et al. 2012; Fehren-Schmitz et al. 2014). Other analyses are also possible from mtDNA: not just comparing haplogroup frequencies per population, but analysing parts of the mtDNA sequence, which allows for finer resolution. The Quebrada de Humahuaca female-line mtDNA profile is overall genetically intermediate between the Andean and Gran Chaco population clusters (the latter represented by the Wichí and Guaraní), possibly suggesting a mix of the two genetic components.

For the Y-chromosome, meanwhile, a set of STR (Short Tandem Repeat) markers was analysed and compared to similar data retrieved from the literature. In this male-line STR data, the Quebrada de Humahuaca profile is closer to that found in populations speaking languages of the Aymara, Quechua, Guaraní and (formerly) Uro linguistic lineages (for the latter, see [Chapter 4.1](#)). The authors therefore conclude that the pre-Hispanic populations of the transitional region

of the Quebrada de Humahuaca may have evolved locally without a significant genetic contribution from preceding or contemporary highland Andean cultures. Nevertheless, some exchanges could have occurred on the maternal side, towards the Gran Chaco, as a consequence of patrilocal exogamy (that is, the tendency for men to remain in their home region while women from elsewhere ‘marry in’ to it).

Layers of genetic and linguistic contact in Arawak speakers of central Peru

Our fourth and final case-study here addresses the origin of the Yanesha, a population in the Selva Central of Peru (provinces of Junín and Pasco), on the eastern slopes of the Andes (Barbieri et al. 2014). The Yanesha speak an Arawak language, like the Machiguenga in the first case-study by Scliar et al. (2014). The Yaneshas’ form of Arawak, however, betrays especially heavy influence from Quechua, and indeed certain other language sources. In other words, their Arawak base language has been impacted by a series of different contact strata. The single most significant impact was from the Yaru dialect of central Quechua, spoken in neighbouring areas of the highlands. Other sources of borrowing are southern Quechua, nearby Amazonian languages and other unidentified languages (Adelaar 2006). These contacts affected aspects of the sound system, grammar, and above all the lexicon, in the form of a large number of loanwords. The home territory of the Yanesha, situated along a trade route towards the Cerro de la Sal (for which see Chapter 5.4) already established before the Incas (Lumbreras 1974), may have been marked by conspicuous population movements influencing the linguistic and genetic diversity of the Yanesha. For more on the Yanesha or Amuesha language, see Chapter 3.4.

To understand the origins of the Yanesha and the genetic impact of these several waves of contact over the centuries, samples were collected from communities at different altitude levels, in the high selva (*selva alta*), from altitudes between 1,200 m and 1,800 m, and in the intermediate selva (*selva media*) at c. 300 m, for a total of 214 individuals. Both uniparental markers were analysed, to compare the genetic profile of the Yanesha against the rest of the continent, in particular the neighbouring Andean and Amazon populations. The Y-chromosome comparative dataset includes 62 populations, the mtDNA dataset 77 populations.

A first observation is that for both uniparental markers, genetic diversity between individuals is higher than the average across the other South American populations analysed. The high diversity values indicate that the Yanesha populations were not particularly isolated, and/or had a large effective population size (proportional to a lower likelihood of marrying a relative). This could indicate that the Yanesha were part of an exchange network that introduced a degree of gene-flow from other populations. In further pairwise comparisons with other South American populations, it became clear that the mtDNA dataset does not have enough resolution to be able to pinpoint any single major source of genetic contribution. The maternal profile of the Yanesha appears not particularly distinct, but

similar to other South American populations from the Andes, Amazonia and the north of the sub-continent.

The most informative results come from individual genetic profiles (haplotypes) composed of 15 Y-chromosome STRs. The haplotypes of individuals belonging to different populations were compared. Assuming that identical haplotypes are shared by genealogically related individuals (who inherited the haplotype from a common ancestor, not too many generations ago), then the share of identical haplotypes between individuals of two populations will be proportional to the strength of recent contact and intermarriage (in this case, of male lineages, in these Y-chromosome data). As the generations succeed each other after the contact period, there is a higher chance of accumulating mutations, which would make the haplotypes look increasingly different. Of the two Yanesha populations, the high selva Yanesha share more haplotypes with other populations than do the intermediate selva Yanesha. This may reflect an environmental factor: the intermediate selva villages are more densely surrounded by forest, and possibly less accessible from the exchange routes.

It is also possible to plot onto a map of South America the frequencies of identical and similar haplotypes in the source populations that potentially exchanged these haplotypes with the Yanesha. Notably, the main source of contact is found in areas that once fell within the southern half of the Inca Empire (from 1472 or earlier): from the shores of Lake Titicaca as far as central-western Bolivia. This pattern does not seem to be paralleled in the female line, however, so this predominantly male gene-flow might be best explained by movements of male traders (the result of deep-time processes of exchange), and/or military forces (associated in some way with the Inca Empire: short-lived but with apparently dramatic impacts upon populations). The results do not, however, allow us to detect any specific gene-flow from the population that had the most powerful contact impact on their language, namely Yaru Quechua-speakers from central Peru. This may be attributed to two factors: a) that our database lacks populations suitably representative of Yaru speakers (the closest geographic proxy would be the sample of Quechua speakers from Huancavelica), or b) more recent contact masking the earlier inputs from the Yaru. A third scenario would simply imply that the linguistic contact was not accompanied by any substantial gene-flow.

The genetic composition of the Yanesha, then, would appear to result from intense exchanges with Andean populations. The genetic data alone neither support nor refute a potential Amazonian origin for the Yanesha population, but that is inferred from the nature of the Yanesha language as Arawak in origin, with later strata of contact influence from highland languages (Adelaar 2006). On the other hand, an Amazonian genetic component was not contemplated for the Shimaa, a population similarly living in the eastern Andean slopes and speaking an Arawakan language of Amazonian origin. In the first case-study above, in fact, the authors exclude any Amazonian genetic input to this population, which would thus have to

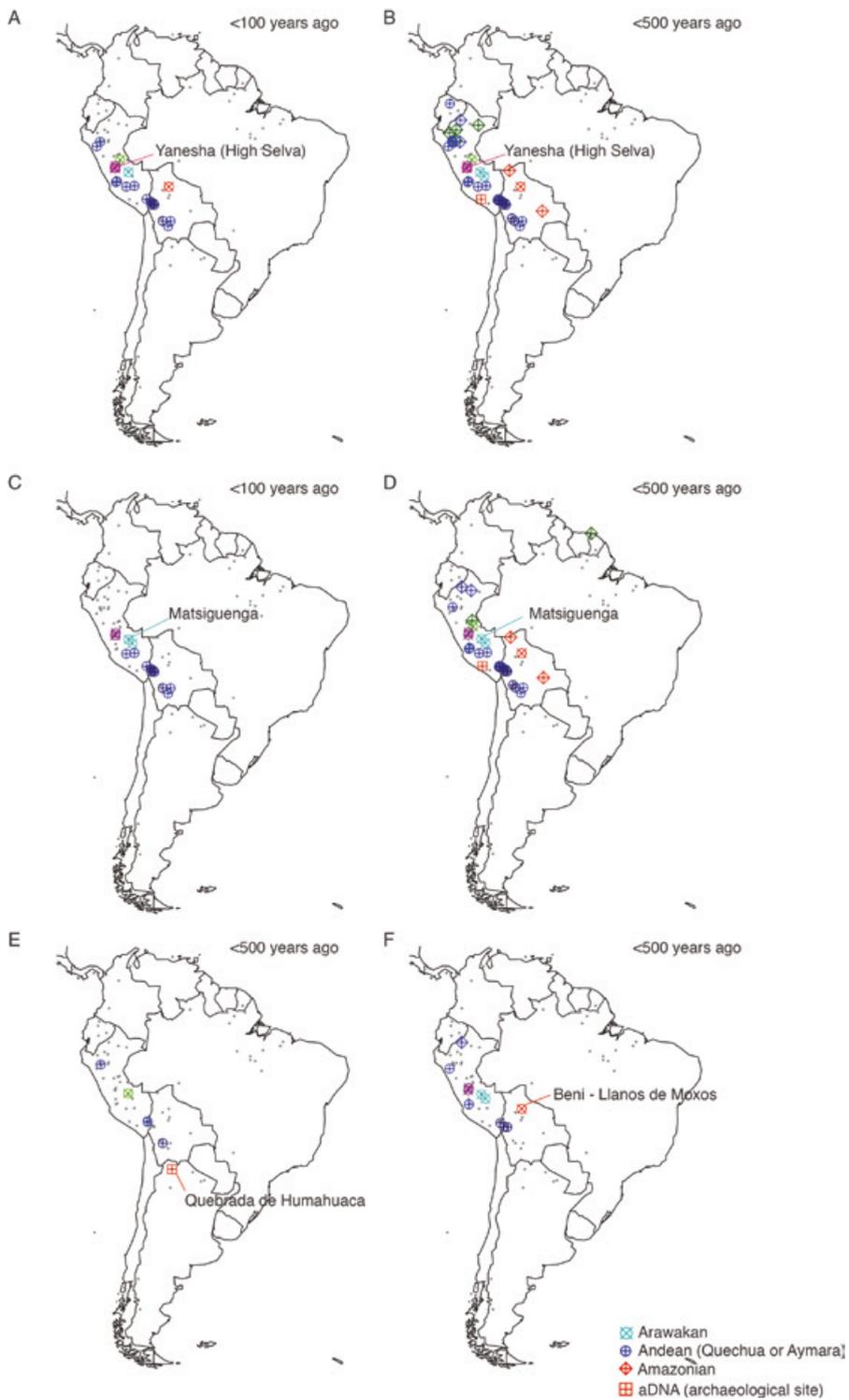
have completely shifted language and culture away from their Andean genetic origins (Scliar et al. 2014). Comparisons obtained from autosomal data and a reduced comparative dataset are probably not exhaustive for testing the dual Andean–Amazonian component in these transitional environments of the Andean eastern slopes. Would the Shimaá share the same genetic profile of their linguistic neighbours the Yanéscha, if one looked at their uniparental markers?

Additional Y-chromosome comparisons shed light on the genetic make-up of populations living in the highlands/lowlands transitional environment

To clarify the factors that could have contributed to the genetic make-up of the Shimaá, and of other populations from transitional environments, I have performed further comparisons using the one genetic marker that can provide both maximum availability of comparative population data and a satisfactory level of resolution. STR markers are positions on the Y-chromosome characterized by a high mutation rate between generations, and which are thus highly variable. Roewer et al. (2013) reports STR data for 17 loci in a wide set of South American populations. This dataset, merged with other available population data, proved to be highly informative for the Yanéscha case-study (Barbieri et al. 2014). The time-depth of isolation and contact reachable with such comparisons was formally tested with Bayesian simulations (Barbieri et al. 2017). I have therefore used the most updated dataset from Barbieri et al. (2017) and compared patterns of haplotypes shared with the chosen target populations. Comparisons of the amounts of shared haplotypes within pairs of populations were evaluated on two levels: haplotypes that are either identical or very similar, and which therefore reflect divergence times within the last 100 years; and less similar haplotypes, with divergence times calculated to fall within the last 500 years.¹ For further technical details on the genetic data and analyses followed, see Barbieri et al. (2017).

The results in [Figure 3.3.1](#) show the amount of very similar haplotypes (those that could be derived from a common ancestor within approximately 100 years, or 3 to 4 generations) and of less similar haplotypes (an approximate divergence time range of 500 years) for various populations: high selva Yanéscha (A and B), Machiguenga (C and D), Quebrada de Humahuaca (E) and Llanos de Moxos (F).

The Machiguenga populations analysed in Mazières et al. (2008) and Sandoval, Lacerda et al. (2013) were considered as a linguistic proxy for the Shimaá, who also speak a Machiguenga language. These population samples share very similar haplotypes (and therefore recent common ancestors) with the neighbouring Quechua speakers of Cuzco, the Aymara and Quechua speakers of Lake Titicaca, and the Quechua speakers of Potosí, as well as with the Yanéscha populations with whom they share the same Arawak language lineage (C). The paternal ancestry of the Arawakan Machiguenga is therefore similar to that of the



Arawakan Yanesha, with gene-flow from the highlands until even more recent times. At a deeper timescale, however, the Machiguenga share similar haplotypes with populations of the Llanos de Moxos in Bolivia, and with the Kalina of French Guiana (D), as well as with Quechua speakers of the Amazonian regions of north-eastern Peru. Another point of difference with the Yanesha is that the Machiguenga have very low genetic diversity (that is, the Machiguenga individuals share very similar haplotypes between each other). Low genetic variance (as a measure of diversity) is correlated with a high degree of isolation, which prevents the genetic component from being admixed and thus prevents its diversity being enriched by introduced non-local haplotypes. The value of internal diversity (here calculated as haplotype variance) is only 0.36 and 0.24 in the two Machiguenga samples, while it reaches 0.67–0.70 in the Yanesha and in the ancient Quebrada de Humahuaca sample. The low values of the Machiguenga are more compatible with those found in prototypical isolated Amazonian populations, as explained at the beginning of this chapter, while the Yanesha and the Quebrada de Humahuaca seem more in line with levels of mobility and exchange found in the Central Andes. See also [Figure 3.2.1](#), in [Chapter 3.2](#), and a list of diversity values for different populations in [Supplementary Table 2](#) in [Barbieri et al. \(2014\)](#).

In the next target population in this analysis, the ancient sample from Quebrada de Humahuaca, we do not see any haplotypes shared with living populations over the last 100 years, as expected given the time elapsed since the death of the individuals recovered from the site. With a deeper time frame, less similar haplotypes are found shared in present-day Quechua-speakers from Taquile and Amantaní islands in Lake Titicaca, and in Aymara-speakers from Pampa Aullagas (Bolivia), as well as in a population from Amazonia (Yine) and in the northern Andes (near Chachapoyas) (E). No similarities are found with the other ancient DNA samples included in the analysis, from the site of Tompullo. So while a connection with the ancestors of living Andean populations seems plausible, the evidence is sporadic at best, and historical contact appears difficult to reconstruct.



Figure 3.3.1 Chronological chart showing the time-depth of the major archaeological divergences between Amazonia and the Andes prior to c. AD 1500. Maps indicate the populations in the South American dataset that share haplotypes with the selected target populations, within approximate timeframes of 100 and 500 years. The small dots locate each of the populations included in the comparative dataset (for details, see [Barbieri et al. 2017](#)). On each map, the target population is indicated with a line. Maps A and B: sharing patterns for the high selva Yanesha. Maps C and D: sharing patterns for the Machiguenga (averaged between the two samples available from [Mazières et al. 2008](#) and [Sandoval et al. 2013b](#)). Map E: sharing patterns for the ancient DNA from Quebrada de Humahuaca. Map F: sharing patterns for the Llanos de Moxos, Beni department. Map built in R with dedicated packages ([Becker et al. 2018](#)). © Chiara Barbieri.

Finally, further comparisons are shown for a sample from the Llanos de Moxos (Beni department, Bolivia: [Chapters 4.3](#) and [4.4](#)), to add a perspective from another transitional environment: the Bolivian piedmont. This sample, analysed by Cárdenas et al. (2015), consists of a mix of individuals from various rural localities with good representation of the province of Moxos, where the Moxo languages of the Arawak family are spoken (Aikhenvald 1999). Interestingly, this Moxos population shares identical or very similar haplotypes only with the Yanesha population (data not shown) and less similar haplotypes with a set of populations slightly different to those plotted in B and D: Aymara- and Quechua-speakers from the shores of Lake Titicaca, but also people from Cajamarca in northern Peru, and above all with the Yanesha and Machiguenga. It is tempting to suggest a genetic connection between Arawak speakers of the eastern slopes of the central Andes (such as Yanesha and Machiguenga) and the Bolivian lowlands of the Moxos, which would be in line with the (controversial) hypothesis that the Arawak language family originated in the western Amazon basin (Walker and Ribeiro 2011), and that its expansion was associated with that of domesticated manioc in southern Amazonia, again where it reaches into Bolivia (Olsen and Schaal 2001). Nevertheless, these speculations are difficult to prove without a more complete dataset, which would need to include other populations representative of Amazonian Arawak speakers.

Overall genetic trends in the Andes–Amazon transition, and conclusion

In conclusion, genetic data support various different structures between Andean and Amazonian populations, with both uniparental markers and autosomal data showing different ancestral components and different patterns of diversity (Tarazona-Santos et al. 2001; Fuselli et al. 2003; Bisso-Machado et al. 2012; Barbieri et al. 2014; Eichstaedt et al. 2014). Many factors played a role in building the Andes–Amazonia genetic divide: demographic, historical, but also environmental, as shown by Eichstaedt et al. (2014, 2015); see also [Chapters 2.2](#) and [3.2](#). Nevertheless, the dynamics between these two major regions have only recently begun to be addressed from a genomic perspective (Gnecchi-Ruscione et al. 2019).

Population contact can translate into gene-flow, the direction of which generally comes from the culturally dominant population. In most of the recent case studies reviewed, the authors reported the sharing of genetic motifs with current populations living at high altitude: the global picture therefore seems to agree on a predominant influence of the Andean highlands. This happens in particular with the Calchaquí of north-west Argentina (Eichstaedt et al. 2015), but also with the two Arawakan populations on the eastern slopes of the Central Andes, the Yanesha and the Machiguenga, who received a major paternal contribution from the southern highlands (from Lake Titicaca to Potosí), plausibly when these regions all came under the Inca Empire. For the Shima Machiguenga, Scliar et al. (2014) propose

an exclusively Andean origin and a subsequent complete linguistic/cultural shift to Arawak; this scenario may not hold once more neighbouring populations from both highlands and lowlands are included in their models. In the above section, for instance, I suggest evidence for possible connections with the Bolivian piedmont which merit further inquiry.

Other regional patterns of exchange are also detected: the ancient population of the valley of Quebrada de Humahuaca, in north-west Argentina, shows a maternal contribution from the Gran Chaco, in line with a patrilocal marriage practice. Finally, the patterns of sharing between the Llanos de Moxos, the Yanéscha and the Machiguenga reveal the possibility of a connection between Arawakan speakers of the Andes–Amazonia divide and the lowland fringe of the Andes.

Our perspective is strongly biased towards the data available: the choice of populations sampled, and the choice of genetic data analysed. In some cases, the results are non-informative (for example, the maternal ancestry of the Yanéscha looks very similar to the one found in the majority of the South American populations analysed, see Barbieri et al. 2014). With the latest publications releasing further fine-scale genetic data (full mtDNA genomes, high-resolution Y-chromosome SNP and STR data, and, in particular, high-coverage autosomal data), and more coverage of case-study populations, we will be able to shed further light on population dynamics across the Andes–Amazonia transition.