DISSERTATIONES
BIOLOGICAE
UNIVERSITATIS
TARTUENSIS
387

LINDA ONGARO

A genomic portrait of American populations

DISSERTATIONES BIOLOGICAE UNIVERSITATIS TARTUENSIS
387

DISSERTATIONES BIOLOGICAE UNIVERSITATIS TARTUENSIS
387

LINDA ONGARO

A genomic portrait of American populations

UNIVERSITY OF TARTU
Press

Institute of Genomics, University of Tartu, Estonia

Dissertation was accepted for the commencement of the degree of Doctor of Philosophy in Gene Technology on 18th of June 2021 by the Council of the Institute of Genomics and the Institute of Cell and Molecular Biology, Faculty of Science and Technology, University of Tartu.

Supervisors: Francesco Montinaro, PhD; Research Fellow, Department of Biology, University of Bari, Italy; Research Fellow of Population Genetics, Institute of Genomics, University of Tartu

Luca Pagani, PhD; Associate Professor, Department of Biology, University of Padova, Italy; Senior Research Fellow of Population Genetics, Institute of Genomics, University of Tartu, Estonia

Mait Metspalu, PhD; Director and Professor of Evolutionary Genomics, Institute of Genomics, University of Tartu, Estonia

Opponent: Cesar Fortes-Lima, PhD; Researcher in Human Population Genetics, Uppsala University, Sweden

Commencement Room No 105, 23B Riia St, Tartu, on the 31st of August 2021, at 11:15.

Publication of this thesis is granted by the Institute of Molecular and Cell Biology and the Institute of Genomics, University of Tartu.

ISSN 1024-6479
ISBN 978-9949-03-656-1 (print)
ISBN 978-9949-03-657-8 (pdf)

University of Tartu Press
www.tyk.ee

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF ORIGINAL PUBLICATIONS

I

Grugni V, Raveane A, **Ongaro L**, Battaglia V, Trombetta B, Colombo G, Capodiferro MR, Olivieri A, Achilli A, Perego UA, Motta J, Tribaldos M, Woodward S, Ferretti L, Cruciani F, Torroni A, Semino O. 2019. **Analysis of the human Y-chromosome haplogroup Q characterizes ancient population movements in Eurasia and the Americas**. *BMC Biology* 17 (1), 3.
doi: 10.1186/s12915-018-0622-4

II

Capodiferro MR, Aram B, Raveane A, Rambaldi-Migliore N, Colombo G, **Ongaro L**, Riviera J, Mendizábal T, Hernández-Mora I, Tribaldos M, Perego UA, Li H, Scheib CL, Modi A, Gòmez-Carballa A, Grugni V, Lombardo G, Hellenthal G, Pascale JM, Bertolini F, Grieco GS, Cereda C, Lari M, Caramelli D, Pagani L, Metspalu M, Friedrich R, Knipper C, Olivieri A, Salas A, Cooke R, Montinaro F, Motta J, Torroni A, Martín JG, Semino O, Malhi RS, Achilli A. 2021. **Archaeogenomic Distinctiveness of the Isthmo-Colombian Area.** *Cell*, 184(7), 1706–1723.e24.
doi: 10.1016/j.cell.2021.02.040

III

Ongaro L, Scliar MO, Flores R, Raveane A, Marnetto D, Sarno S, Gnecchi-Ruscone GA, Alarcon-Riquelme M, Patin E, Wangkumhang P, Hellenthal G, Gonzalez-Santos M, King RJ, Kouvatsi A, Balanovsky O, Balanovska E, Atramentova L, Turdikulova S, Mastana S, Marjanovic D, Mulahasanovic L, Leskovac A, Lima-Costa MF, Pereira AC, Barreto ML, Horta BL, Mabunda N, May CA, Moreno-Estrada A, Achilli A, Olivieri A, Semino O, Tambets K, Kivisild T, Luiselli D, Torroni A, Capelli C, Tarazona-Santos E, Metspalu M, Pagani L, Montinaro F. 2019. **The genomic impact of European colonization of the Americas.** *Current Biology*, 29(23), 3974–3986.e4.
doi: 10.1016/j.cub.2019.09.076

Author's contributions to the listed articles are as follows:

Ref. I: I designed primers and performed laboratory work by genotyping samples, reconstructed and dated the phylogenetic tree, analysed the data, provided some of the figures and co-wrote the manuscript.

Ref. II: I performed haplotype-based analyses, participated in interpreting the results and providing input for writing the paper.

Ref. III: I participated in all data analyses except sex-biased estimation, interpreted results, provided main figures and co-wrote the manuscript.

ABBREVIATIONS

aDNA	ancient DNA
AS-PCA	Ancestry-Specific Principal Component Analysis
BGI	Beijing Genomics Institute
BSP	Bayesian Skyline Plot
CE	Common Era
chrY	Y chromosome
Cm	centimorgan
CP	CHROMOPAINTER
EM	Expectation-Maximization
FS	fineSTRUCTURE
F_{ST}	fixation index, mean pairwise genetic distance
GT	GLOBETROTTER
GWAS	Genome-wide association studies
Hg	haplogroup
HMM	Hidden Markov Model
IBD	Identical by descent
ISOGG	International Society of Genetic Genealogy
kb(p)	Kilobase (pair)
ky(a)	kilo years (ago)
LA	Local Ancestry
LD	Linkage Disequilibrium
LGM	Last Glacial Maximum
Mb	Megabase
MCMC	Markov Chain Monte Carlo
ML	Maximum Likelihood
MP	Maximum Parsimony
MSY	male-specific region
mtDNA	mitochondrial DNA
NGS	Next Generation Sequencing
Ne	effective population size
NNA	North Native American
NNLS	Non-Negative Least Square
NRY	Non-Recombining region of Y chromosome
PAR	PseudoAutosomal Region
PCA	Principal Component Analysis
SNA	South Native American
SNP/SNV	Single nucleotide polymorphism/variant
SRY	Sex-determining region of Y
ssDNA	single-strand DNA

sub-Hg	sub-haplogroups
(T)MRCA	(Time) to the Most Recent Common Ancestor
TVD	Total Variation Distance
UPopA	Unsampled Population A
UPopI	Unsampled Population from the Isthmus
YCC	the Y Chromosome Consortium

1. INTRODUCTION

The evolution of American populations has been the subject of several multidisciplinary studies. Our knowledge regarding the formation of the genetic diversity of the Americas is still incomplete, although genetic studies are constantly adding new details on this topic. The development of new technologies, such as Next Generation Sequencing (NGS), enormously increased the number of worldwide genomic data, including those from the Americas. Moreover, these technologies together with other technical improvements, lead to the possibility of extracting and analysing DNA from ancient specimens, making "ancient genomic" (aDNA) one of the many fundamental tools to understand our ancestor's past.

Although the Americas were the last continents to be reached by our sapiens ancestors, the processes shaping their genetic variation have been extremely complex, and their studies have been the topic of many genetic surveys for more than three decades. In the beginning, uniparental systems dominated the population genetics research of American populations. It started with mitochondrial DNA (mtDNA) and soon included the Y chromosome (chrY) analysis. The latter revealed that the two founding Native American chrY haplogroups (Hg) probably were Hg C and Hg Q, accounting for about 5% and 75% of Native American males, respectively. However, the resolution of these haplogroups did not undergo substantial improvements until a few years ago.

The first publication included in this dissertation (Ref I) aims to investigate from a male perspective the genetic history of the Americas through a fine dissection of the Pan-American haplogroup Q and to reconstruct a comprehensive and detailed haplogroup Q phylogeography and that of its sub-lineages.

The uniparental systems could be considered as two loci that are used to understand the female and male perspective of human history. They can describe only two ancestors of the thousands involved in shaping the genetic legacy of modern populations. The genomic representation of a more significant number of ancestors is encrypted in the autosomal markers. Therefore, autosomal markers are crucial to understanding the timing and the dynamics of population movements in the Americas. Thanks to archaeological and genetic evidence, it is now accepted that the first people arriving in North America came from Siberia, passing through Beringia after late Glacial times. Initial settlements were followed by widespread people movements that reached southern South America relatively fast, as early as ~15 thousand years ago. Several recent studies have provided new information about this subject, reconstructing the genomic history of indigenous groups from different regions of the Americas, but the Isthmo-Colombian area is still lacking.

Hence, the second publication of this thesis (Ref II) employed both ancient and modern DNA data analysis to reconstruct the genomic history of the Isthmo-Colombian area. It aims to define the genomic background of Panamanian indigenous populations to evaluate the intra-Isthmus variability and shed light on pre-

Columbian Americans' genomic history assessing the connection between the Isthmo-Colombian area and the rest of the Americas.

Besides the first migrations, American populations result from several admixture events since the colonial era and the Trans-Atlantic slave trade. Moreover, many waves of migration followed by local admixture occurred in the last two centuries, the impact of which has been largely unexplored.

The third reference in this thesis (Ref III) explores how more recent migrations shaped the genomic background of admixed American populations. In particular, this study aims to reconstruct the fine-scale ancestry composition, estimate the time of admixture, examine the demographic evolution of different continental ancestries after the admixture and assess the extent and magnitude of sex-biased gene-flow dynamics in admixed American populations.

2. LITERATURE OVERVIEW

This literature overview first gives a general background of the history of the American continents from a multidisciplinary perspective. It then follows with a genomic history of the American population to set the background for the more specific questions addressed in the research work that constitute this thesis.

2.1. History of the Americas

The study of human migrations has been approached from many different perspectives, exploiting the knowledge obtained from several disciplines. We might say that is one of the subjects that need to be observed from many angles, in order to be able to get a thorough understanding of it. Investigating people's movements and trying to assess what happened since the appearance of Humans in Africa until the present has been a challenging voyage, of which only the main stops have been uncovered. Although each continent and geographical area on Earth has its history and everywhere, there are still topics that remain unclear or unexplored the history of the Americas is perhaps the one that remains even further from being fully unveiled.

2.1.1. The American continents before the arrival of the Europeans

2.1.1.1. The first peopling of the Americas from a multidisciplinary perspective

To reconstruct the first peopling of the Americas, scientists from many different disciplines have investigated different pieces of evidence with four main questions: who were the first Americans? Where did they come from? When did they arrive, and how? And what routes did they follow?

Nowadays, thanks to archaeology and genetics, it is widely accepted that the first people arriving in North America came from Siberia, passing through Beringia sometime between ~17.5 and ~14.6 kilo years ago (kya) (Waters 2019). Although very recent works show that there are archaeological and undirected genetics evidence of human presence down to Mexico around the Last Glacial Maximum (LGM, ~26–18 kya) (Ardelean et al. 2020; Becerra-Valdivia and Higham 2020) and thus this would suggest that the first entrance into the American continents occurred as early as 33–31 kya, further studies are needed to confirm this new hypothesis.

To understand how Humans could colonize the Americas through the Bering Sea, it is crucial to consider the climatic conditions at the time the first humans entered the American continents. During the period of interest, the climate was dominated by the ice age, with the LGM at around ~26–18 kya. The Beringian land bridge was present from 65–36 kya and then again from 30–13 kya, and the

land in Beringia was probably similar to tundra. It must also be considered that in this frozen territory, there were niches where climatic conditions made human survival possible (Sikora et al. 2019). For much of the Pleistocene, further south, there was an ice sheet covering a large part of North America, blocking the movements of people. At the end of the Pleistocene, the temperature started to rise, and the ice sheet was divided into eastern (Laurentide) and western (Cordilleran) sections by an ice-free inland corridor stretching from present-day Yukon, through Canada, to Montana. While the environment in this corridor would still have been inhospitable, it constitutes a possible route into the rest of the Americas. It is thought that this corridor was open before the LGM until 24 kya and again after 14 kya (Goebel et al. 2008). Moreover, a Pacific coastal ice-free corridor became available for passage at that time, approximately 15 kya. Therefore, two possible migratory routes have been proposed: an interior route, by land from Siberia across the Beringian land bridge and down an ice-free corridor in North America; and a coastal route, by sea along the coast of Siberia, Beringia, and North America (Figure 1).

The oldest human presence in the inland corridor is documented at Charlie Lake Cave (British Columbia, Figure 1), where stone tools are associated with bison bone remains radiocarbon dated to 12.35 ± 0.5 kya (Driver et al. 1996). On the other hand, the opening of the ice-free coastal corridor is tied to the ice margin retreat that started ~17 kya (Lesnek et al. 2018).

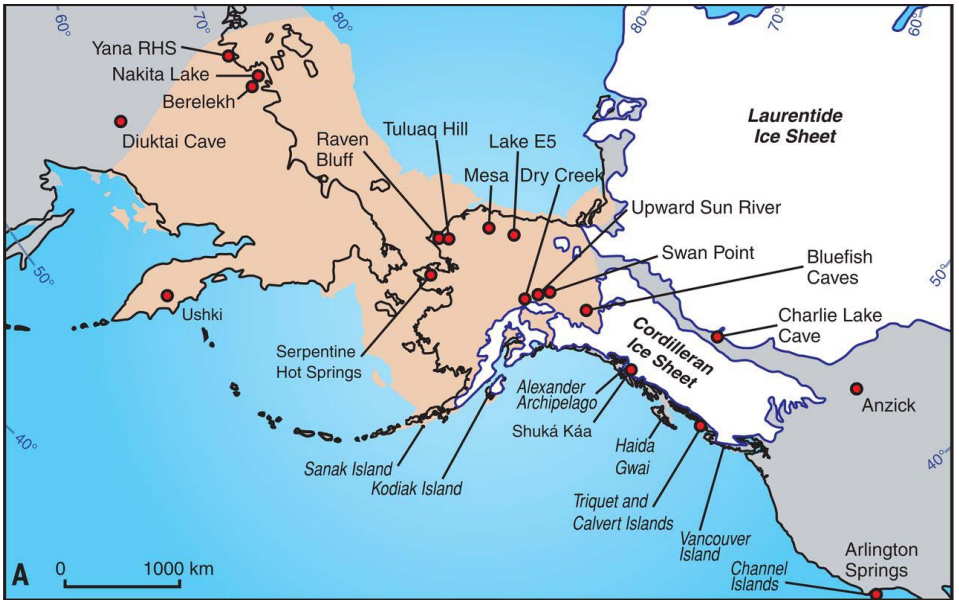

Figure 1. Representation of Beringia (light brown), Cordilleran and Laurentide ice sheets (white), the ice-free corridor between the ice sheets, the Pacific coastal route, and the location of important archaeological and geological sites. Reprinted with permission from AAAS from (Waters 2019).

From an archaeological point of view, south of the ice sheets, pieces of evidence of first human presence are provided by several sites dated between ~15.5 and ~13.3 kya. For example, the sites of Hebior and Scheider in Wisconsin, in which excavations recovered disarticulated bones of mammoths, dated around 14.85 and 14.65 kya, respectively (Overstreet 2005; Joyce 2013). Further evidence of human presence, dated ~ 15.5 kya, were unearthed in Mexico and Texas (Anda et al. 1953; Waters et al. 2011) and South America, as demonstrated by the human remains found in the site of Monte Verde in Chile (14.2 kya) and Huaca Prieta in Peru (14.15 kya). For decades, scientists were convinced that the Clovis complex was the oldest homogeneous group of artefacts, found mainly in North America, to arise in the Americas from 13.5 kya (Waters and Stafford 2007; Goebel et al. 2008). This archaeological complex found mostly in North America, is named after the town of Clovis in New Mexico, and it is characterized by Clovis points, fluted projectile points constructed from a variety of stone types. Clovis artefacts were found exclusively south of the continental ice sheets, and their densest concentration lies east of the Mississippi River, but they have also been discovered west of the Mississippi River toward the eastern side of the Rocky Mountains and south into northern Mexico (Waters and Stafford 2007). Therefore, Clovis was probably absent in the Pacific coast (Beck and Jones 2010) and also in Central or South America (Pearson 2017; Acosta-Ochoa et al. 2019), where other types of artefacts were detected. Lanceolate points with basal stems characterize the Western Stemmed Tradition found in Oregon, Idaho and Nevada with the earliest date related to the first site in Oregon (~13–12.7 kya) (Jenkins et al. 2012; Davis et al. 2014; Goebel and Keene 2014). These studies show that the Western Stemmed Tradition was at least contemporaneous with Clovis, or even prior; therefore, it did not derive from Clovis.

As for the southern continent, another type of artefacts was observed: the fishtail projectile points. In Argentina and Chile, different archaeological sites were dated between ~12.8 and ~12.2 kya, which could indicate a human presence that coexisted with Clovis (Jackson et al. 2007; Prates et al. 2013); these fishtail points also occurred in Central America and other parts of South America (Acosta-Ochoa et al. 2019). It is not clear whether this type was invented in South America or if it descended from the ones present in the Northern continent.

Another discipline that is crucial to reconstruct the history of the first peopling of the Americas is linguistics. In 1986, Greenberg and colleagues published a research paper in which they proposed a three-migration hypothesis (Greenberg et al. 1986). They classified languages into three families: Eskimo-Aleut spoken in the Arctic, Na-Dene in Canada and south-western US and Amerind, which groups all the other indigenous languages in North American and all languages in Central and South America. Thanks to linguistic and genetic data available at that time, they hypothesized that the first dispersal, earlier than 11 kya, was related to Amerind, while around 9 kya there was the one of the Na-Dene, and at last, the Eskimo-Aleutian dispersal was around 4 kya. Greenberg et al. examined the genetic data available to them, characterised mainly by classical polymorphisms, and concluded that they were supporting their three-migration hypothesis.

The American continents' population dispersal happened relatively quickly, perhaps in no more than a thousand years; these people could settle and adapt in a highly heterogeneous geographic environment and develop different lifestyles during the expansion. However, their lives were suddenly devastated. William M. Denevan wrote: "The discovery of America was followed by possibly the greatest demographic disaster in the history of the world" (Denevan 1992). In fact, at the time of the conquistadores' arrival after 1492, the pre-contact Americas' population estimates were expected to be as high as 112 million, while others estimate the population to have been lower, around eight million. However, native populations declined to less than six million in less than 200 years (Sánchez-Albornoz and Others 1974; Thornton 1987; Denevan 1992). Moreover, the distribution and density of Indigenous people throughout the continent were uneven, probably influenced by geography and environmental changes, which conditioned dispersal dynamics and social configuration, ranging from hunter-gatherers to complex hierarchical civilizations (Salzano and Bortolini 2005). Native American populations are thought to have collapsed during the first century of the colonial period, with a reduction of approximately 90% of the population size, meaning that in several colonized areas with relatively small populations, the natives were essentially eradicated (Thornton 1987). Even if the violence was extreme, this catastrophe was also caused by famines and infectious diseases brought from the "Old World".

2.1.1.2. The Isthmo-Colombian area demographic scenario

The Panamanian isthmus lies between the Atlantic and Pacific Oceans and connects the two American continents. This area might appear as a small one if it is compared to other vast regions, but it nevertheless represents a crucial region for the understanding of the Americas' demography. First of all, it was the only land bridge between the two continents, during the initial peopling of South America and it has then remained a crossroads of goods, technologies, ideas and peoples throughout history, including more recent colonial times (Cooke 2005).

Evidence from archaeology and paleoecology show that human occupation in the Isthmo-Colombian area has been continuous since at least 11 kya (Cooke et al. 2013; Martín et al. 2016; Redwood 2020). The earliest well documented cultural remains in Panama refer to Clovis dated to ~11 kya (Cooke et al. 2013). In general, archaeological evidence and lake sediments concurred that some descendants of the initial colonizing populations remained in the region adapting to the changing environmental conditions of the Late Glacial-Holocene transition, while others expanded southward (Dillehay et al. 2017). Thanks to the land bridge position and its orography, influenced by the proximity of multiple plate junctions, a multitude of isthmian and insular landscapes was created that favoured the development of a high degree of diversity, also among human groups (Barrantes et al. 1990; Cropp and Boinski 2000; Anderson and Handley 2002; Herlihy 2003).

After ~8 kya, agriculture started to prosper across the Isthmo-Colombian Area; although most cultivars, such as maize (*Zea mays*) and squash (*Cucurbita spp*)

together with few domesticated animals, like muscovy duck (*Cairina moschata*), were first domesticated north or south of Panama, there is no clear evidence that their initial introduction has been accompanied by major population displacements (Cooke 2005; Piperno 2011). However, this area was certainly subjected to a series of migrations from the north, the south and the east (with the Caribbean islands), along both overland and maritime routes, as testified by agriculture and pottery variations and other cultural changes. The maritime route has been documented since 6.2–5.6 kya (Martín et al. 2016). Gold and cacao trades with Mesoamerica are documented by the first Europeans, but a more ancient connection with Aztec and Maya populations is also possible (Lothrop 1942). Connections with continental Colombia occurred at least until 1.4 kya, and perhaps also during the last 500–700 years (Mason and Johnson 1940).

In the last five centuries, the arrival of Hispanic colonialists wiped out some autochthonous cultures, but Indigenous resistance was more effective in the extensively forested western Panama, thus allowing for a higher degree of survival of pre-Hispanic cultures (Quilter and Hoopes 2003). Nowadays, the Panamanian population comprises Indigenous and mixed individuals. According to the last census, the Indigenous ethnic groups make around 12% of the total population (2010). The main Indigenous ethnic groups are Ngäbe (62.3%), Guna (19.3%) and Emberá (including Wounáan) (7.5%) followed by smaller cultural groups: Bribri and Naso Djërdi (or Teribe) (Perego et al. 2012; Grugni et al. 2015). Nuclear Chibchan (also Chibchan, Chibchano) languages are the most widely spoken linguistic group in Panama today. This language family includes the majority of surviving or recently extinct languages spoken in the Isthmo-Colombian area, including Honduras, Nicaragua, Costa Rica, Panama and northern Colombia. The Chibchan name derives from an extinct language, called Chibcha, once spoken in Colombia. However, genetic and linguistic data now indicate that this family's root is dated about 10 kya in a "core area" between southern Costa Rica and western/central Panama, where the widest variety of Chibchan languages is still found.

2.1.2. The American continents after the arrival of the Europeans

2.1.2.1. The European colonization

On the 3rd of August 1492, Cristoforo Colombo, an Italian explorer and navigator, departed from Palos de la Frontera, Spain, with three ships: the Santa María, under Colombo's direct command, the Pinta and the Niña. He had the support of Queen Isabella I of Castile and King Ferdinand II of Aragon, intending to discover and acquire islands and mainlands in the Ocean Sea (Elliott 1984). The night of October 12, they landed on an island in the Bahamas, called Guanahani by the natives, named *San Salvador* by Colombo. Colombo undertook other three voyages towards these new lands, reaching many other islands in the Caribbean Sea, including Puerto Rico and Cuba. Soon after the discovery of the Americas by

Colombo, Spain and Portugal began to explore and settle the continents. Throughout the XVI century, both countries started to establish a string of colonies also on the coastal mainland, including the Pacific coast (Adhikari et al. 2017).

These two kingdoms fought themselves over the new territories until 1494 when they signed the Treaty of Tordesillas, ratified by the pope, in which the new continents were divided into two areas of exploration and colonization along a meridian 370 leagues west of the Cape Verde islands, off the west coast of Africa. This conferred Portugal all colonies in the Atlantic Ocean and West Africa, but only a part of the south-east of South America (Elliott 1984). This division still stands today, especially linguistically, with Brazil being the only Portuguese-speaking country in Latin America. The spanish justification for their claims to the new continents was based on the Ideals of the Reconquista; they believed it was their moral duty to convert the natives to Christianity and they were supported by the Church and the Pope (Bonch-Bruevich 2008).

Other European kingdoms, in particular the French and the British, with others like the Dutch having a smaller role, followed the Iberian expansion to the Americas. The French established colonies in eastern North America, several Caribbean islands, and small coastal parts of South America. During the XVI century they also explored and colonized a region of Canada that was subsequently called New France. Almost a century after the Iberians, England started to colonize the new continents and they eventually conquered much of Eastern North America, the Caribbean, and parts of South America. They began to place new settlers in new lands wishing to create many different colonial societies, differently from the Spanish that were mainly interested in conquering new territories (Elliott 1984). Instead, the Dutch colonized some islands in the Caribbean establishing the Netherlands Antilles, together with other territories that they will lose in the future.

One of the causes of the widespread admixture in the early stages of the colonization era was the male-biased immigration from Europe (i.e. the Iberian Peninsula), favouring the intermixing with Native and sometimes Sub-Saharan African women in a patriarchal way (Mörner 1967; Lavrin 1989; Kamen 2003). For example, the first Iberian immigrants were mostly (>80%) males (Boyd-Bowman 1976), while the proportion of females was initially 5–6%, and they were up to 28% in the 1560s and 1570s (Bethell 1984). Therefore, the initial shortage of Iberian women naturally encouraged mixed marriages. Another element favouring higher rates of interethnic mating was the fact that the establishment of the colonies usually matched with the presence of Native American settlements, precisely because of the nature of the colonization process, finding in these peoples sources of labour and taxation (Sánchez-Albornoz 1994; Salzano and Bortolini 2005).

Although most American territories occupied during the European colonial expansion have become independent countries, some continue to be dependencies, especially in the Caribbean.

As already mentioned before, the indigenous populations suffered a major breakdown caused by the arrival of the colonizers. A big part of this was caused by the epidemics brought from the Old World; diseases like smallpox, typhus,

influenza, diphtheria and measles were not present in the Americas before and when they spread up to 95% of the indigenous population perished. It is important to underline that not only the diseases were responsible for this huge decrease in populations, but also famine, slavery and overwork contributed greatly (Thornton 1987).

2.1.2.2. The Trans-Atlantic Slave Trade

Slavery had existed in the Iberian Peninsula throughout documented history. After the end of the Western Roman Empire, different systems of slavery remained in the successor Islamic and Christian kingdoms of the peninsula. Moreover, before the Trans-Atlantic slave trade, slavery was common in many African areas for many centuries. Evidence indicates that enslaved people from some parts of Africa were exported to Africa, Europe, and Asia before the Americas' European colonisation (Klein 1999).

Upon discovering new lands through their naval explorations, European colonists soon migrated to and settled in lands outside their continent. Under the Kingdom of Castile, Europeans invaded and colonised the Canary Islands during the XV century, where they turned much of the land to the production of wine and sugar. At the time, primarily Portuguese traders used the Canary Islands as a naval base, and they began to move their activities down Africa western coast, capturing people to be later sold as slaves in the Mediterranean. By 1494, the Portuguese king had opened dealings with the leaders of several West African states that would allow trade between their respective peoples (Thornton and Thornton 1998).

Therefore, between the XVI and XIX century, Spanish and Portuguese began the so-called Trans-Atlantic Slave Trade to mitigate the loss of labour work due to the breakdown of the indigenous population. Until the first decades of the XVII century, African slaves arrived predominantly in Spanish colonies in the Americas (Klein 1999). The first region in Africa to be exploited for the slave trade comprised what nowadays are territories of Gambia and Senegal on the West Coast, and until 1640 this area remained the main site of the Spanish possessions (Barry 1998). At first, the Spanish deported more slaves than the Portuguese, but this trend shifted through time; by the beginning of the century, African slave arrivals to Brazil, a Portuguese colony, equalled the number arriving in Spanish America, and in a decade, they surpassed their numbers (Figure 2A–B). From the beginning of the XVII century, the Portuguese settled in Gabon, Congo and Angola, starting to trade slaves from these regions (Sánchez-Albornoz 2014). The volume of slaves from this area of Central Africa was the highest compared to other regions from the sixteenth until the late nineteenth century (Klein 1999; Fortes-Lima and Verdu 2021).

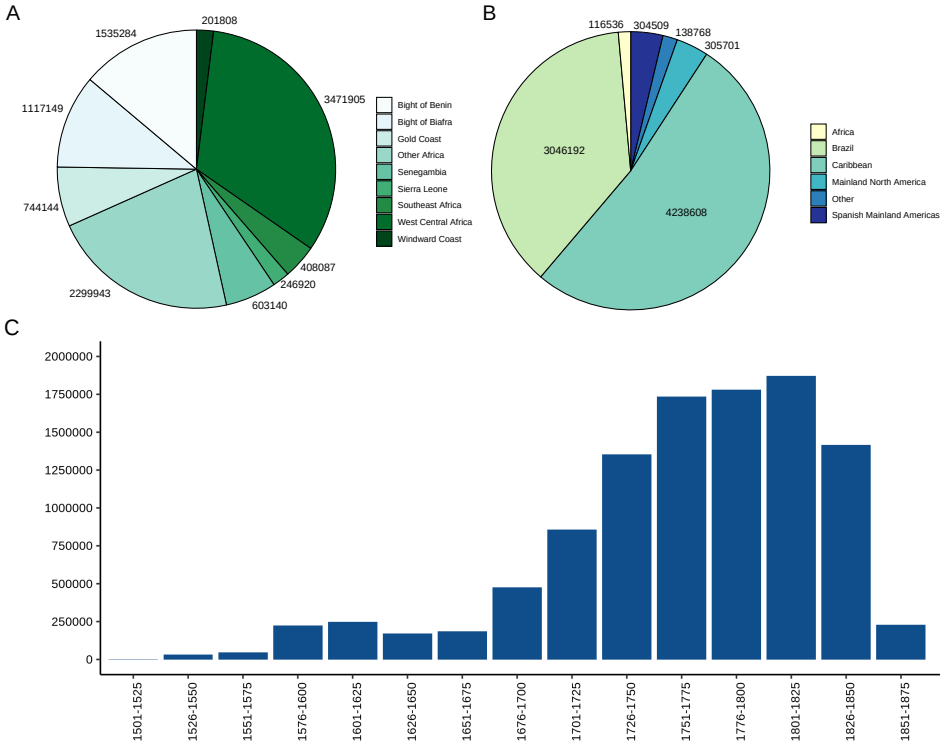

Figure 2. Trans-Atlantic Slave Trade data representation. **A)** Total number of African slaves by region of embarkation. **B)** Total number of African slaves by region of disembarkation. **C)** Timeline of the total number of African slaves embarked from Africa. Data downloaded from https://www.slavevoyages.org; permission conveyed through Creative Commons Attribution-Noncommercial 3.0 United States License.

When the British and the French discovered that sugar was one of the few crops that could profitably be exported into the European market on a mass scale, they started to exploit the African slaves too. By the 1720s, the Iberian colonies were getting less than half of all the slave arrivals, while the British, French and Dutch colonies were trading more than 50% of all Africans arriving in America. By 1680 in Barbados, a British colony, over 38,000 slaves were arriving, especially from the Bight of Benin and the Bight of Biafra, two of the primary source areas reported for the British-mediated slave trade (https://www.slavevoyages.org/). Barbados became the most populous and wealthiest English colony, being also the one with more slaves per square mile (Klein 1999).

The magnitude of the slaves' deportation was not the same during the centuries but underwent an increasing trend (Figure 2C). In the first two or three centuries of the trade, the people deported mainly lived within fifty miles of the African coast. Nevertheless, the rise in the demand in America boosted the increased movement of slaves, changing the trade qualitatively and quantitatively in the eighteenth and nineteenth centuries. In West African different regions, such as the Bights of Benin

and Biafra and the Congo-Angola area, the source of slaves now moved further inland and more of the West and East African populations became involved in the trade. Indeed, the Atlantic coast of Africa was not the only region involved in slave deportation. The Portuguese infiltrated East-Central Africa, in particular Mozambique, together with Angola, but they needed more time to exploit this area at the same rate. They were first attracted by the gold, but then they began to trade people. In the XIX century, Mozambique represented the third-largest supplier of slaves (Klein 1999).

Finally, the phenomenon of the slave trade had a huge impact in shaping the history and the diversity of American populations, without neglecting the devastation and suffering that the African continent and its populations suffered; in total, the number of Africans deported to the Americas could be estimated approximately to 12 million people (https://www.slavevoyages.org, (Fortes-Lima and Verdu 2021)).

2.1.2.3. Recent migrations from other continents

By the XIX century, the massive deportation of Africans to the Americas was interrupted. However, voluntary immigrants from other parts of the world continued to flow into the new continents; many people continued to arrive, contributing to the increase in the variability of American populations, especially from Southern and Eastern Europe. For instance, it has been estimated that more than 32 million people migrated to the United States at the end of the XIX century and the beginning of the XX century (Bayor, 2016); while more than 6 and 5 million people moved to Brazil and Argentina, respectively (Baily & Míguez, 2003). These individuals arrived mostly from the Iberian Peninsula, but also a considerable number of Italians and Germans made the journey.

The so-called "Italian diaspora" has been described as one of the largest migrations of the XIX century (Bonaffini & Perricone, 2014; Giunta & Sciorra, 2014; Wong, 2006). Although it started soon after the European discovery of the Americas, in the beginning, the numbers were not significant. However, they reached high proportions in the second half of the XIX century, with more than 11 million individuals migrating towards the continents, mainly to the US, Brazil and Argentina. Moreover, between 1866 and 1916, approximately 4 million Italians were admitted to the United States. In Brazil, the phenomenon was also promoted politically; in fact, the society offered subsidies for the promotion of immigration. After the first two decades of the XIX century, almost half of all immigrants were Italians. Their annual arrival rate became higher than the one from Portugal in 1876. These migrations continued until 1902 when the Italian government signed a decree to end all subsidized emigration to Brazil (Tosi 2002). In Argentina, millions of individuals migrated from Northern (earlier) and Southern (later) Italy from the second half of the XIX century throughout the 1950s (Baily and Míguez 2003; Meade 2011). At the beginning of the XX century, it has been recorded that Italian immigration was the highest (39.4%) compared to other countries (Solberg 1970; Brown 2010).

Instead, the German immigration to the Americas took place between 1820 and World War I, when nearly six million Germans migrated to the United States, representing the largest group of immigrants from 1840 to 1880 (Wittke 1952).

In addition to European immigrants, a high number of migrants from other parts of the world moved to the Americas. They settled mainly in the United States, but a proportion also migrated to South America. Chinese migrants worked in Cuba primarily on sugar plantations, while in Peru, they contributed to the construction of the Andean Railroad or in the fields. As for the Japanese, they reached Brazil and Peru, where many worked as servants and low wage workers for planters. In both cases, the influx of Asian workers to certain parts of the Americas was to fill the void left in the labour forces after the abolition of slavery (Meade 2011).

2.2. The genomic history of the American populations

Historically, the American continents have been primary receivers of migrants where different cultures and populations have mixed during the centuries, continuing to evolve. Studying the genomic background of these populations has been fundamental to understand their history.

2.2.1. The first peopling of the Americas from the uniparental point of view

2.2.1.1. Uniparentally inherited loci: mtDNA and Y chromosome

Uniparental markers proved to be particularly useful for population genetics studies, especially when sequencing technologies and computational capabilities were not yet at the level we can observe now. They are related to the mitochondrial DNA (mtDNA) and the Y chromosome (chrY), which share a significant feature, the uniparental transmission, respectively maternal and paternal.

The mtDNA is a circular double-stranded DNA molecule stored in the mitochondria, and it is around 17 kilo base pair (kbp) long in mammals. It is found in a great abundance in cells because each cell contains tens to thousands of copies of the mitochondrial genome, depending on the cell type. As already mentioned, the main features of mammalian mtDNAs, for population genetics, is that they are inherited through the maternal line, from mother to offspring, without any paternal contribution as already ascertained in 1980 (Giles et al. 1980). Therefore, the maternal inheritance, the lack of recombination, a high copy number per cell, and a fast mutation rate are the characteristics that made the mtDNA a fundamental tool for evolutionary genetics studies. Because of its higher mutation rates and multiple copies in a cell, mtDNA was the first choice before the PCR advent and when the sequencing technologies were mainly based on the Sanger method (Sanger and Coulson 1975), which was very expensive. This allowed the scientists

to cost-effectively scan a higher proportion of variation when compared to autosomal chromosomes. However, despite their high informativeness, fast mutation rates also increase the frequency of recurrent mutations, which are very well known in many mtDNA regions, such as the hypervariable one.

The structure of the Y chromosome consists of a significant region called Male Specific region (MSY, 37Mb long), also known as the Non-Recombining region of the Y chromosome (NRY), and three PseudoAutosomal Regions (PAR) (Veerappa et al. 2013; Hughes and Page 2015). This chromosome is of crucial importance for the determination of sex as in it is placed the SRY gene (Sex determining Region of Y), which is responsible for the male differentiation of the embryo; chrY also includes other genes responsible for male fertility.

As briefly described above, the mtDNA and chrY loci have the unique feature to be transmitted to the descendants only by mothers and fathers, respectively. This aspect gives many advantages that many researchers exploited. For example, mtDNA and the MSY do not undergo homologous recombination; therefore, their sequences may vary only because of the accumulation of random mutational events, and there is no need for gametic phase reconstruction, as explained later in this thesis. This accumulation process generated over time groups of monophyletic chromosomes (haplogroups, Hg) characterized by the same combination of markers, equal position, and order, inherited from a Most Recent Common Ancestor (MRCA) (Torroni et al. 1993).

Therefore, all modern mtDNAs and MSY regions coalesce back to one ancestral sequence at some point in the past (Jobling et al. 1997; Jobling and Tyler-Smith 2003; Behar et al. 2012). At the same time, variants belonging to different haplogroups showed a separation at a specific time in the past, from which a series of mutations occurred independently and accumulated, differentiating various haplogroups and sub-haplogroups (sub-Hg).

The study of the origin and diffusion of haplogroups was fundamental to outline a basis for studying the genetic history of humans, which was then deepened with the possibility, in recent decades, of studying recombination events with autosomal markers. The genealogies reconstructed from the two uniparental systems might reveal similarities and/or differences among the cultural, social, and evolutionary histories of different species and population groups.

One of the main advantages of uniparental systems is that it is possible to apply two basic concepts of population genetics, the coalescence time and the mutation rate, without taking into account the confounding effect of recombination events, in order to date the MRCAs age precisely. The coalescence allows us to calculate when two alleles converge in the past in terms of generations; with the mutation rate, it is possible to calculate the probability that a mutation occurs every generation, usually measured as the probability of a neutral substitution per site per generation. Assuming a mutation rate constant over time, the molecular clock delineating the coalescence times without the intervention of recombination, it is possible to date when the MRCAs arise in the past.

Uniparental systems have another advantage, they allow to analyse the dynamics of male and female effective population sizes (Ne) through time. For

instance, Karmin and colleagues (2015) reported that the analyses of the dynamics of Ne's male-specific aspects in worldwide individuals revealed a sharp decrease of chrY Ne around 8–4 kya without a counterpart in the mtDNA data. At the extreme, the female Ne was estimated to have been up to 17-fold higher than the male Ne at the same time. This drastic decline in male Ne can be explained by culturally inherited variance of reproductive success among males, potentially related to wealth accumulation during the Neolithic revolution (Hammer et al. 1998).

Moreover, their specific features have attracted scientists' interest because it represents an essential tool for various studies; for example, there are practical implications in forensic DNA analysis and genetic genealogy.

Despite all the reported advantages, the uniparental systems could be considered as two loci that are used to understand the female and male perspective of human history. Therefore, they can describe only two ancestors of the thousands involved in shaping the genetic legacy of modern populations, and for this reason only a small part of the evolutionary history of a population. The genomic representation of a more significant number of ancestors is encrypted in the autosomal markers.

Among the two uniparental systems, only the Y chromosome variability has been investigated in this PhD research.

2.2.1.2. Phylogeny and phylogeography of the Y chromosome

The word "phylogeny" refers to studying the history of the evolution of a species or group, especially about lines of descent and relationships among groups of organisms or a population (Gittleman 2014). Moreover, the phylogeny, represented by a tree, is the result of a study called phylogenetics; the latter is a part of systematics that addresses the inference of evolutionary history and the relationship among or within groups of organisms. Phylogenetic analyses have become crucial to understanding biodiversity, evolution, ecological genetics, and genomes.

Another approach to the study of Human evolution through the use of chrY is phylogeography, which analyses the phylogenetic position of each group, its geographical distribution, and its internal variation, allowing to reconstruct the routes followed by many of the migrations in human history. It is an interdisciplinary approach that integrates phylogenetic data with climatologic, demographic, anthropological, archaeological, linguistic, and historical data.

Modern chrY haplogroups can be organized in a phylogenetic tree representing the evolution of the human Y chromosome (Jobling and Tyler-Smith 2017). There are different approaches to build a tree, and the most used are character-based: the Maximum Parsimony (MP) generates a tree that uses the smallest number of evolutionary changes; while, the Maximum Likelihood (ML) evaluates different tree topologies and assuming an evolutionary model finds the tree explaining the data with the maximum likelihood (Sharma et al. 2018).

In 2002, the (Y Chromosome Consortium 2002) (YCC) established a nomenclature for the phylogenetic tree: the main haplogroups are indicated by an uppercase letter of the alphabet and differ in turn in sub-haplogroups (Y Chromosome Consortium 2002). These can be identified by the corresponding letter followed by the name of the terminal mutation or by an alphanumeric code where the first number indicates the sub-haplogroup, the following letter the subgroup of sub-haplogroup, and so on. Also, internal lines to a given group, which are not characterized by any downstream markers, are defined as para-groups, and they are indicated by the name of their haplogroup followed by an asterisk. The nomenclature is constantly updated with novel markers (Jobling and Tyler-Smith 2003; Karafet et al. 2008; van Oven et al. 2014). Besides YCC, the International Society of Genetic Genealogy (ISOGG) (https://isogg.org/tree/) continually revises the nomenclature by adding new markers too.

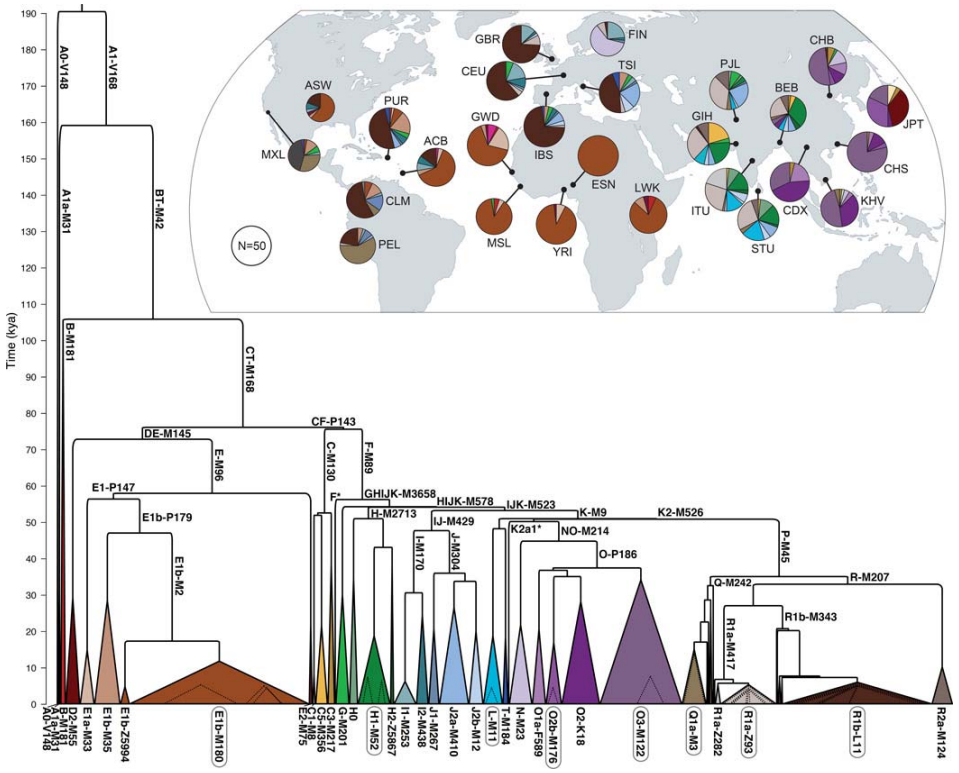

Figure 3. Representation of Y chromosome phylogenetic tree based on the 1000 Genomes Project data. The labels on the branches and below the triangles are haplogroup names. Labels outlined in grey ovals indicate haplogroups that have undergone recent rapid expansions. Haplogroups represented by many chromosomes are collapsed into triangles, and the triangle height represents the coalescence time, and the width represents the frequency in the sample. The world map indicates for each of the 26 populations: the geographic origin, sample size, and haplogroup frequencies according to the colours of the tree. Reprinted from Figure 2 (Poznik et al. 2016) with permission from Springer Nature conveyed through Copyright Clearance Center, Int.

The advancement in the High Throughput Sequencing (HTS) technology are providing a large amount of chrY SNPs (Xue et al. 2009; Rocca et al. 2012; Francalacci et al. 2013; Poznik et al. 2013; Wei et al. 2013; Scozzari et al. 2014; Poznik et al. 2016). This massive amount of SNPs allows to obtain a high resolution, but it is also a source of recurrent mutations that are consequently not informative at the population level.

The human Y chromosome's global phylogeny includes 20 main haplogroups (Figure 3), all derived from an ancestral chrY that is considered the last common ancestral chrY. The oldest subclade represents African variation, and it is followed by younger lineages describing variation outside of Africa (Karafet et al. 2008; van Oven et al. 2014; Hallast et al. 2015; Karmin et al. 2015; Poznik et al. 2016). The exclusive African origin of the deepest lineages (A00, A0, A1–3, B) in the phylogenetic tree supports the Out of Africa model (Wood et al. 2005; Batini et al. 2011; Cruciani et al. 2011; Mendez et al. 2013; Scozzari et al. 2014; Hallast et al. 2015; Barbieri et al. 2016; Poznik et al. 2016); according to it, anatomically modern humans arose in Africa, and a sub-group emerged from the continent, settling the rest of the world.

2.2.1.3. Phylogeny and phylogeography of the Native American Y chromosomes

The uniparental section of this research thesis will focus on the chrY phylogeny and phylogeography related to the American continents.

The current pool of Native American Y chromosomes is a mixture of haplogroups derived from pre-Columbian dispersals from Siberia and more recent gene flow from Europe and Africa (Grugni et al. 2015). Although the identification of the founding lineages has been complicated because of the high historical rate (about 16%) of male-mediated admixture in Native Americans (Bosch et al. 2003), two founding haplogroups, Hg C and Hg Q, were described early on (Underhill et al. 1996; Karafet et al. 1997; Karafet et al. 1999). On average, they account for about 5% and 75% of Native American males, respectively. Until a few years ago, the resolution of these haplogroups did not undergo substantial improvements; only a few studies based on full Y chromosome sequencing include samples belonging to haplogroup Q (Lippold et al. 2014; Karmin et al. 2015), and none of them focused specifically on its phylogeny.

Y haplogroup C is one of the most ancient and extensive haplogroups in Asia, with a wide distribution area extending from the eastern to the central parts of the continent and reaching Europe (Chiaroni et al. 2009; Karmin et al. 2015; Wei et al. 2018). The most common sub-haplogroup is M217 which probably originated in the southeast or Central Asia, where it spread towards Northern Asia and then into the Americas, represented mainly through its sub-clade C-P39 (Bortolini et al. 2003).

M242 is the signature marker of haplogroup Q, which seems to have arisen in Siberia (Altai regions), widely distributed (Zegura et al. 2004). It has been found in North Europe (Marjanovic et al. 2005; Battaglia et al. 2009; Di Gaetano et al. 2009), the Middle East (Regueiro et al. 2006; Grugni et al. 2012) and South Asia

(Deng et al. 2004; Sengupta et al. 2006; Fornarino et al. 2009). Furthermore, it is the predominant Y chromosome haplogroup among Native American populations. In 1996, the Stanford University research group headed by Dr. Peter Underhill discovered a SNP (Single Nucleotide Polymorphism) known as M3 (Underhill et al. 1996). Q-M3 sub-clade is the most prevalent one in both North and South Native American individuals; according to its distribution and to the associated haplotypes, it is likely that Hg Q originated in Central Asia and then spread from middle Siberia, where it is widespread nowadays, reaching Beringia where probably M3 arose.

Q-M3 and Q-M242 (xM3, that means not M3), subsequently classified as Q-L54 (xM3), have been identified as the two American founding lineages, both containing a complex arrangement of sub-clades (Dulik, Owings, et al. 2012; Battaglia et al. 2013; Geppert et al. 2015).

2.2.1.4. Ancient DNA and Y chromosome

In the last decade, the methodologies' progress has entirely changed the study of human Y chromosome variation providing new insights into human history. Ancient DNA (aDNA), the DNA extracted from archaeological remains (e.g. bones), gives a temporal and spatial dimension to genetic studies that would be unavailable with modern genomes' analysis alone. The widespread sequencing of aDNA was prevented for several years due to contamination problems, mainly because endogenous DNA, in archaeological remains, is scarce and fragmented. Most of the genetic material obtained from fossils tends to be exogenous, like from environmental microbes or individuals who handled the material. The latter is a very problematic issue, given that present-day human DNA and endogenous aDNA are practically identical and therefore challenging to discriminate, and so it can introduce biases in the subsequent analyses. However, researchers have developed new expedients to correct contamination in the aDNA samples allowing the study of previously unusable specimens; first of all, a standard extraction practice under strict clean-room conditions and several bioinformatics tools to analyse data (Knapp et al. 2012; Orlando et al. 2015). As already mentioned, the other main problem of aDNA is the low quantity of endogenous DNA, which has been partially overcome by the development of NGS. This new technology has transformed aDNA research, allowing the sequencing of all the DNA present in a substrate and subsequently extracting ancient individuals' whole genomes to characterise genetic variation (Margulies et al. 2005). The use of NGS techniques has favoured the sequencing of DNA found in small fragments that are difficult to amplify and sequence with traditional methods and the ability to distinguish ancient sequences from the modern contamination. The aDNA holds peculiar features, such as short sequences and deamination of cytosines to uracils at strand ends.

Despite the progress, contamination remains one of the main problems in the analysis of ancient sequences. Several methods have been developed, and most are based on identifying single or multiple haplotypes in haploid regions of the human genome. Among the most widely used software are Schmutzi (Renaud et al. 2015) and contaMix (Fu et al. 2013). They are based on identifying possible

contamination by modern DNA, comparing the consensus (or reads) to a panel of reference mitogenomes. In a nutshell, they calculate the probability that possible SNPs patterns are not inherited from a common ancestor but are instead due to the presence of more haplotypes, and therefore by contaminants. A similar strategy can also be applied on nuclear DNA, but only on males, considering possible contamination on the X chromosome. In males, this chromosome is haploid and therefore, the presence of multiple alleles indicates (excluding sequencing errors) the presence of contamination (Moreno-Mayar et al. 2018). In addition to technological development, the experimental part has also seen a significant improvement, considering where to search for endogenous DNA in human remains and how to treat the extracted DNA. The identification of the petrous bone as a rich source of DNA has allowed the possibility of finding reasonable amounts of DNA even in environments that instead favour the degradation of the molecule.

Moreover, once the DNA is extracted, the pre-treatment of DNA could be done with an enzyme (Uracil-DNA glycosylase) that corrects the deamination on the cytosine (Rohland et al. 2015). Then, the building of sequencing libraries starting from single-strand DNA (ssDNA) allows a large portion of ancient DNA to be sequenced, otherwise excluded from the classical methods based on the construction of libraries starting from double-stranded DNA (Gansauge and Meyer 2013). In addition, enrichment methods by capture have been developed to increase the amount of sequenced endogenous DNA. These approaches are based on the capture of the whole genome (Lindo et al. 2017) or specifically selected regions previously identified as informative for the human genetic history (Maricic et al. 2010; Fu et al. 2015).

Although sequencing ancient Y chromosomes and getting adequate coverage can be challenging, MSY sequences of ancient individuals have provided insights into the history of chrY variation. For example, ancient Y-DNA sequences of dated human remains can be used to see if they are consistent with conclusions made from phylogenetic trees built using modern data and calibrate the molecular clock. Additionally, ancient Y-DNA may also shed light on clades that have gone extinct and cannot be detected from contemporary chrY sequences (Kivisild 2017).

American ancient samples helped to unravel the history of the two Y chromosome Native American lineages, the haplogroups C and Q. These haplogroups have been involved in multiple migrations in the Americas from Siberia, starting around 15 kya (Zegura et al. 2004; Dulik, Owings, et al. 2012; Dulik, Zhadanov, et al. 2012; Battaglia et al. 2013; Grugni et al. 2015). As already mentioned, the two ancient sub-clades of hap Q, Q-M3 and Q-L54(xM3) were probably born somewhere in Siberia before the first dispersal into the Americas, and together they capture the overwhelming majority of extant Native American Y chromosomes today (Figure 4) (Zegura et al. 2004; Battaglia et al. 2013; Grugni et al. 2015; Jota et al. 2016). The first ancient specimen to be analysed was found in On Your Knees Cave (OYKC) on Prince of Wales Island, Alaska and radiocarbon dated 10.3 kya. His chrY was assigned to haplogroups Q-M3* by direct analysis of the signature marker M3, thus, placing a minimum date of 10.3 kya for the emergence of this haplogroup (Kemp et al. 2007).

From the shotgun sequences of two other ancient genomes from the Americas, the Anzick Boy associated with Clovis and dated ~11 kya, and the Kennewick Man dated ~9 kya, it turned out that their Y chromosomes belong to Q-M3 and Q-M971, respectively (Rasmussen et al. 2014; Rasmussen et al. 2015; Kivisild 2017). Thus, these results confirmed that the two sub-lineages Q-M3 and Q-M971, were already present in the Americas at least 10.3 kya. A fourth ancient American sample from Greenland, the first complete ancient human genome sequenced, is a 4 kya specimen referred to as Saqqaq from the site where it was discovered, and his chrY belongs to Q-B143 sub-haplogroup (Rasmussen et al. 2010). This lineage separated from other sub-groups of haplogroup Q, diverging more than 25 kya and has been associated with a founding lineage that spread from Asia through the Bering Strait toward Greenland, where it has been isolated, maybe in a population closely related to the Paleo-Eskimos.

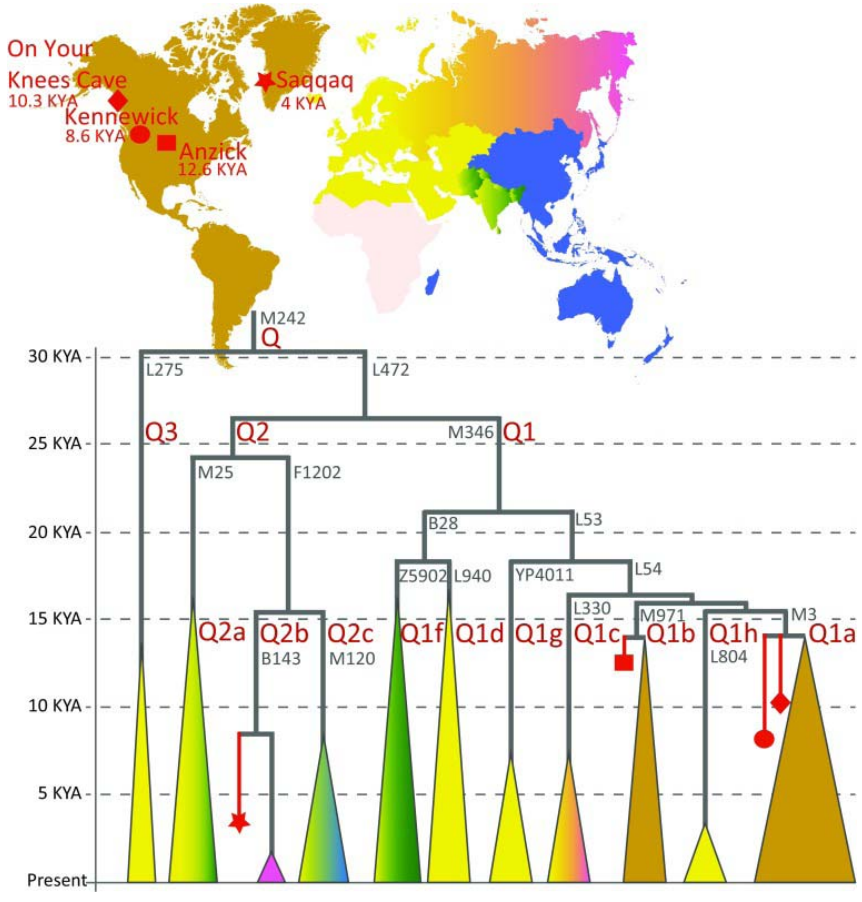

Figure 4. The phylogeographic and phylogenetic mapping of ancient Y chromosomes haplogroup Q. The structure of the major sub-haplogroups is drawn in proportion to time according to estimates from high coverage genomes of present-day populations. Reprinted from Figure 8 (Kivisild 2017) under the terms of the Creative Commons Attribution 4.0 International License (https://creativecommons.org/licenses/by/4.0/).

2.2.2. The first peopling of the Americas from maternal, autosomal and ancient DNA perspective

As previously mentioned, the first peopling of the Americas is a very complex topic to investigate; but during the last decade, a huge step forward has arrived from genomic studies, especially from the aDNA that has revolutionised the study of human evolution (Nielsen et al. 2017; Skoglund and Mathieson 2018). Data derived from different disciplines such as genetics, archaeology and linguistics, led to the hypothesis that the Americas were settled through three separate population movements, the so-called "tripartite migration model" (Greenberg et al. 1986).

The genetic study of the Americas began ~30 years ago, where uniparental systems, analysed with restriction enzymes, dominated the research of population genetics. The first mitochondrial haplogroups ever identified were the four Native American lineages called A, B, C and D, which have become A2, B2, C1 and D1 (Torroni et al. 1993). Subsequently, other founding lineages were identified, reaching the current number of sixteen (Tamm et al. 2007; Achilli et al. 2008; Perego et al. 2009; O'Rourke and Raff 2010; Brandini et al. 2018). As for the chrY also the mitogenome perspective on the Americas' first peopling indicates that Native Americans trace their ancestry to Asian groups who colonised north-east Siberia, including part of Beringia, before the LGM.

The mtDNA data summarised the tripartite migration model in this way: (i) the first dispersal following a pacific coastal route, ~18–15 kya, by the first people that eventually reached South America in few millennia along the Pacific coast, marked by nine Pan-American founders A2*, B2*, C1b, C1c, C1d*, C1d1, D1, D4h3a, and D4e1c; (ii) very close in time, Native American ancestors of C4c, X2a and X2g lineages entered North America through an interior route of dispersal toward the mainland of North America, Na-Dene expansion; (iii) around 5 kya there was the spread of Paleo-Eskimo D2a lineages along the Arctic through Northern Canada and Greenland, which were replaced, in the same region, by the spread of Neo-Eskimos carrying A2a, A2b, and D3 lineages.

The first study that employed a large panel of American genome-wide data published by Reich and colleagues in 2012 supported this tripartite model (Reich et al. 2012), and the people who went south that encountered a vast uninhabited land, which they colonised in groups, thus forming the genetic structure of the Americas. However, this model seemed too oversimplified to describe the American continents' peopling, and subsequent research papers provided a more detailed reconstruction. Even with the new kind of data, the structure of the Americas' first population remains that of the tripartite model; the first peoples that entered North America during and after the LGM came from eastern Asia and passed through Beringia. Then, in unglaciated eastern Beringia/northern North America, they split into two branches called Northern Native American (NNA or Ancestry-B) and Southern Native American (SNA or Ancestry-A). Where and when these two ancestral populations split is still unclear. The NNA is represented by the ancient genome of the "Kennewick man" (also known as "the ancient one") (~9kya) and also by a group of ancient genomes from South Ontario (ASO) in

Canada dated ~4 kya (Rasmussen et al. 2015; Scheib et al. 2018). Interestingly, as already reported, the Y chromosome of Kennewick man belongs to the sub-haplogroup Q-M3, the principal American founder lineage, and could be related to the SNA expansion, in contrast with what the genomic data revealed so far.

Instead, the most ancient representatives of SNA are individuals who were living on both sides of the Rocky Mountains: Anzick-1, associated with Clovis and dated ~11 kya, and the Spirit Cave individuals associated with Western Stemmed technology and dated ~10.7 (Rasmussen et al. 2014; Moreno-Mayar et al. 2018). These ancient individuals carrying SNA ancestries entered South America, and they spread quickly along the southern continent, diversifying into many culturally distinct populations and indigenous groups. This is suggested by the earliest archaeological human presence in the Southern Cone at 14.6 kya and by ancient human genomes dating more than 9 kya on both sides of the continent: at Cuncaicha (Peru) and Los Rieles (Chile) on the Pacific, Lapa do Santo and Lagoa Santa (Brazil) on the Atlantic.

More recently, two different migration waves hit only the northern area of North America. The first ancient genome ever published was from a 4 kya individual, linked to the Saqqaq culture in Greenland, that showed more genetic affinity with Siberian populations than modern Inuits (Rasmussen et al. 2010). A different origin for Paleo- and Neo-Eskimos has been later supported by analysing other genomes (Raghavan et al. 2014). Therefore, the final model proposes a dual migration in the arctic, where the arrival of Neo-Eskimos replaced the initial Paleo-Eskimos populations, as the Saqqaq culture in Greenland that showed more genetic affinity with Siberian pop also proposed by mtDNA studies. Moreover, the new data proposed that during the middle Holocene, together with an improvement of the climatic condition, other migrations from the North/Northern Central America reshaped the genetic structure of Central and South America, primarily admixing or replacing the first people (Figure 5, Moreno-Mayar et al. 2018; Posth et al. 2018). Moreno-Mayar and colleagues (2018) proposed that this migration brings an ancestry of an unsampled population (UpopA) that is neither SNA nor NNA. This ancestry probably originated in Beringia around 25 kya, then admixed in Mexico with SNA and subsequently migrated towards the south. The possibility of another unidentified American ancestry highlights the concept of the Beringia standstill, which states that from Asia, populations that were headed towards the Americas have stopped due to ice in Siberia, where climatic conditions and isolation gave rise to the origin of the typical American ancestry (Tamm et al. 2007). Furthermore, the new data highlighted movements within the American continents, both from north to south (i.e. from California Channel Islands to Peru, Figure 5 (Posth et al. 2018)), and from south to north (from South America to the Caribbean islands (Fernandes et al. 2021)), which have also produced changes at the micro geographical level. In conclusion, the subsequent migrations and admixtures/replacements led to a more complex situation than the tripartite model, both at the continental and micro geographical level.

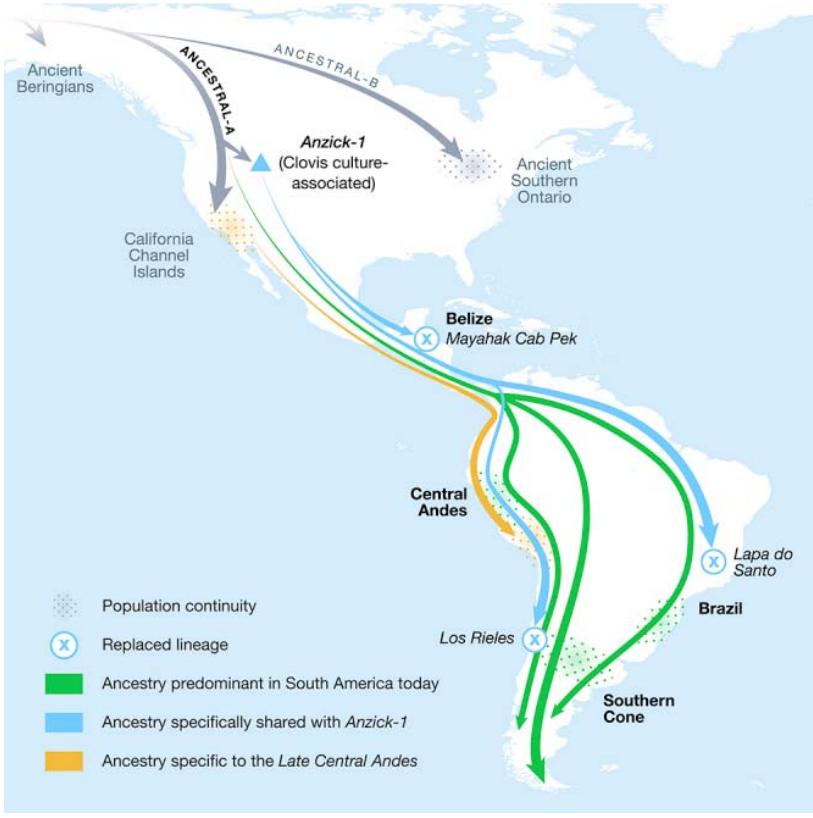

Figure 5. Migrations involved in the first peopling of the Americas and subsequent ancient replacements. Reprinted from graphical abstract (Posth et al. 2018). Licensed under Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), © 2018 Posth et al.

2.2.3. Admixture events shaped human genomic variability

Human population movements often lack historical records, or the available ones are not reliable for different reasons. Moreover, besides migrations and movements, history has seen many instances of forced deportation or displacements, often poorly documented, further complicating an unbiased reconstruction of demographic dynamics relying exclusively on historical evidence. Thus, genomic studies may provide an excellent opportunity to complement the understanding of the human past.

It is evident from archaeology, history and linguistics that populations are not discrete entities, and therefore, they are continuously evolving, producing "hybrid" populations and mixed ancestry individuals. The term "admixture" usually refers to creating a new deme from the mixing of ancestral populations that were previously relatively isolated from each other (Jobling et al. 2013).

During human history, the admixture events led to a subsequent shuffling of genetic ancestry through recombination, producing a variation in ancestry proportions between populations, among individuas in a population, and along the genome within an individual (Figure 6). Moreover, these processes also shape the phenotypic variation and may lead to disease risk differences between populations (Goetz et al. 2014). Because of the variation in the degree of admixture among individuals in an admixed population, the proportion of admixture can be associated with susceptibility to specific diseases that are more prevalent in one or other ancestral populations. One example could be related to the prevalence of type 2 diabetes, which positively correlates with the proportion of Native American ancestry (Gardner et al. 1984; Williams et al. 2000). For this reason, scientists put a significant effort into the fine-scale reconstruction of recently mixed populations to develop strategies and treatments tailored to the individuals, intending to set the basis to personalised medicine and pharmacogenomics approaches that are showing promising results every year.

Despite their ubiquity and importance, many admixed populations remain understudied both in population and medical genetics surveys (Marnetto et al. 2020); moreover, the large majority of genomic data available are from European individuals, and therefore, many populations from other parts of the world are critically underrepresented. This is a critical issue, especially in genome-wide association studies (GWAS), because of the lack of comparative data (Rosenberg et al. 2010; Popejoy and Fullerton 2016). The increase in diversity among study participants will improve our understanding of genetic structure in all populations and guarantee that genetic research is broadly applicable (Peterson et al. 2019).

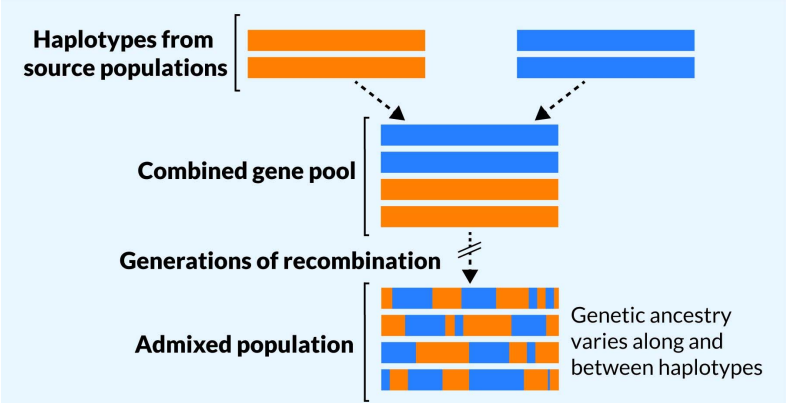

Figure 6. Schematic representation of admixture. A combined gene pool is generated when haplotypes from previously isolated populations come together because of large-scale movements of individuals. Then, thanks to recombination events between these haplotypes during generations, an admixed population with genetic ancestry that varies between individuals and along haplotypes originates. Modified from Figure 1 (Korunes and Goldberg 2021) under the terms of the Creative Commons Attribution 4.0 International License (https://creativecommons.org/licenses/by/4.0/).

2.2.4. The role of haplotype-based methods in population genetics

The field of human population genetics is continuously evolving, primarily because of the increased amount of data available that has led to the development of many advanced methodologies during the last years.

Population genetics has many aims, including inferring the genetic distance and relatedness, understanding the genetic structure and estimating the ancestral proportions of sampled individuals and populations. Then it is also possible, after assessing the admixtures events, to date those events. Finally, it allows exploring the demographic history of admixed populations.

Each individual has inherited different fragments of chromosomes from each parent (haplotypes) that received fragments from their ancestors (Figure 6), therefore considering an entire population, there are fragments inherited by multiple individuals that came from the same common ancestor. Recombination reduced these fragments by breaking some of them and maintaining others. The study of these blocks of Linkage Disequilibrium (LD) and their common inherited history has led to the development of haplotype-based methods that improved the resolution of population structure and minimized the biases of the SNP analysis. The term ‘linkage disequilibrium' is broadly used to refer to the non-random sharing of combinations of variants; it can also be denominated haplotype structure.

It is necessary to make a clear distinction between two main classes of methods in population genomics: (i) allele-frequency-based and (ii) haplotype-based methods. The former considers the variants, or SNPs, in genomic data as a single entity and rely on allele frequency differences. The latter takes into consideration segments or chunks of DNA that are built from a combination of SNPs. The main distinction between the two resides in how the information included in dense SNP datasets is employed because the allele frequencies approaches are not able to deal with the haplotype structure, creating the non-random association of alleles at different loci; in fact, to avoid biases due to LD variance, the linked markers are usually excluded or clumped. On the other hand, the haplotype-based methods benefit from LD by modelling recombination.

In this section, I will focus mainly on the haplotype-based approaches because they are the ones that I applied the most during my PhD studies.

2.2.4.1. Estimating genetic distance and relatedness

The evaluation of genetic distance and/or relatedness of populations and individuals could be achieved through different methods illustrated below.

At the population level, the genetic distance could be estimated by the F_{ST}, a statistical method developed by Sewall Wright (Wright 1931). It belongs to the family of fixation indexes, and it is possibly the most broadly used statistic in population genetics to measure genetic distances. F_{ST} value ranges from 0 to 1, and it is calculated from genetic diversity data, for example, for the comparison of two populations (pairwise F_{ST}):

$$Fst = Vp/p(1-p)$$

Where p is the mean and Vp is the variance of the allele frequencies between the two populations. F_{ST} equal to 0 means that there is no variance among the populations' frequencies; therefore, the populations are similar or very close genetically. On the contrary, high values of F_{ST} indicate populations genetically differentiated.

Reich and colleagues have applied the same rationale in 2009 (Reich et al. 2009), introducing the f-statistics, later implemented in the ADMIXTOOLS package (Patterson et al. 2012). The f-statistics could be used to assess admixture, population structure and genetic closeness between two or more populations. In this package, three f-statistics algorithms are available, labelled *f*2, *f*3, and *f*4; they are employed to show relationships between two, three, and four taxa/populations, respectively. Moreover, a complex scenario could be assessed joining the results of all the *f*-statistics (*f*2, *f*3 and *f*4) in a group of populations, as implemented in the qpWave, qpADM and qpGRAPH algorithms that can be used to evaluate respectively the number of ancestries that compose a population (given a set of outgroups), the proportions of those ancestries and a model of evolutionary relationships (split and/or admixture and/or drift) in a defined set of populations (Reich et al. 2009; Patterson et al. 2012).

Approaches have been developed based on Identity-by-descent (IBD) measurements to estimate the genetic relatedness among individuals, which have been substantially improved by the availability of dense SNP datasets (Weir et al. 2006; Thompson 2013).

As already specified, the allele-frequency methods do not address haplotype diversity, and therefore, haplotype-based approaches have been developed and implemented.

In 2003 Li and Stephens (Li and Stephens 2003) developed a Hidden Markov Model (HMM) for bypassing the issue with LD by interpreting and analysing patterns of LD across multiple loci, considering all markers at the same time and capturing significant aspects of the coalescence. They invented a crucial concept called chromosome painting that was subsequently employed in future advancements. Briefly, with "chromosome painting", each haplotype of a single individual is described as a mosaic of the haplotypes present in the reference panel, showing the ancestral relationships shared by every two samples (the one in the reference set and the one being evaluated).

To perform the painting, the chromosomes have to be arranged into haplotypes because humans are diploid individuals, and as such, it is necessary to restore the haplotypes inherited from each parent in genotype data or short-read sequencing data. This process is called phasing or haplotype estimation, in which the maternal and paternal gametic phase is reconstructed. Different software, such as Eagle, Beagle, IMPUTE2, Minimac and SHAPEIT, have been developed to perform this step (Browning and Browning 2007; Howie et al. 2009; Howie et al. 2012; Delaneau et al. 2013; Loh et al. 2016). One of the most used phasing software is SHAPEIT (now arrived at version 4 (Delaneau et al. 2019)) that

implements a version of the HMM developed by Li and Stephens (Li and Stephens 2003), modelling local haplotype sharing between individuals taking into account mutation and recombination.

However, the process of phasing could be critical. An improvement in the reconstruction of haplotypes comes from the third-generation sequencing taking advantage of linkage information provided by long reads (from a few tens of kb to some Mb) (Amarasinghe et al. 2020). Two types of long reads are available: true long-reads, such as those generated by the Single-Molecule Real-Time (SMRT) sequencing by Pacific Biosciences (PacBio) (Wenger et al. 2019) and by the Oxford Nanopore Technologies (ONT) sequencing (Jain et al. 2018); the second type is the synthetic long reads, which uses systems of linked reads, proximity ligation strategies, and optical mapping (Amarasinghe et al. 2020).

While it is easy to understand the advantage of sequencing haplotypes directly in long reads, it should be noted that the level of accuracy per single base for these technologies remains lower than the second-generation sequencing (99% Hifi SMRT and 95% Nanopore). Although there are tools able to use the long read information to reconstruct the haplotypes (Amarasinghe et al. 2020), with the ongoing improvement of precision in sequencing, these technologies will be essential for an accurate reconstruction of haplotypes.

Once the chromosomes are phased, to infer haplotype similarity or diversity pattern, the most used algorithm is implemented in CHROMOPAINTER (CP, Figure 7, (Lawson et al. 2012)). As SHAPEIT, it harnesses a modified version of the HMM proposed by Li and Stephens, but Lawson and colleagues implemented it in 2012. This software is based on the knowledge that haplotype segments shared among individuals become shorter over time since they shared a common ancestor because recombination fragments chromosomes progressively in each transmission of genetic material from parents to offspring. Therefore, CP allows reconstructing the chromosomes of a set of individuals (called recipients) as a series of genomic chunks inherited from a set of individuals (called donors). The sets of donors and recipients can be the same or different because donors can also be set as recipients, based on the purpose of the analysis, allowing us to focus on more distant or proximate individuals' relationships.

The information retrieved from painted chromosomes, in which, for a given individual, for each SNP, the most recent common ancestor is assigned, can be effectively summarised into a "co-ancestry matrix". In this matrix, for each recipient (usually the rows of the matrix), the total genome length in cM (centi-morgan) or the number of haplotype chunks copied or inherited from every donor is reported. Conveniently, both the donors and the recipient individuals can be pooled into meaningful entities, often using an approach based on chromosome painting itself, as documented in the next paragraph.

2.2.4.2. Inferring population structure

In the last decades, many methods for the investigation of population structure were developed. Among the others, the most commonly used are, together with the F_{ST} method described above, (i) Principal Component Analysis (PCA), (ii) STRUCTURE-like algorithms and (iii) fineSTRUCTURE. The first two are based on allele frequencies, while the third on haplotypes. PCA was probably the first method to be employed for population genomic studies pioneered by Cavalli-Sforza et al. (1964); it is a multivariate statistical method used to summarise the molecular information considering multiple recombining loci. Although PCA is a valuable tool to describe the genetic variation among individuals and has the advantage to be non-parametric, making it suitable in many different situations, its interpretation can be challenging to relate to geographic patterns and specific migration events, together with the high sensitivity to the number of markers included in the analysis (Novembre and Peter 2016).

Additional methods to assess population structure and identify clusters have been based on Bayesian clustering algorithms and are represented by STRUCTURE-like methods (Pritchard et al. 2000). They were subsequently implemented, with different flavours, in many software, such as ADMIXTURE (Alexander et al. 2009), which uses maximum likelihood methods rather than a Markov Chain Monte Carlo (MCMC) algorithm to analyse more extensive genomic datasets efficiently. In brief, these approaches use genotype data to probabilistically assign individuals to K number of clusters representing genetically distinct groups, each of which is characterised by a set of allele frequencies at each locus, and to group individuals that share underlying common allele frequencies and identify admixture proportions from each cluster K at the individual level (Pritchard et al. 2000; Liu et al. 2013). ADMIXTURE uses a STRUCTURE implemented likelihood model but is more efficient and can handle more SNPs with less computational difficulty (Alexander et al. 2009). STRUCTURE-like methods are widely used because they provide a good representation of the genetic structure when used mindfully.

As for the haplotype-based methods, the authors of CHROMOPAINTER also developed fineSTRUCTURE (FS, (Lawson et al. 2012), an MCMC clustering model based on haplotype similarity patterns. Using the coancestry matrix (chunkcounts.out) generated by CP, the model aims to partition the dataset into K groups with indistinguishable haplotype similarity profiles. These K groups are also called clusters and can be used as units for different analyses instead of grouping samples based on population or geographical origin. After identifying the clusters, it is possible to build a hierarchical tree; this tree can contain many leaves that are difficult to refine or interpret. For this reason, sometimes, the tree is cut at different heights and different parts after visual examining the cluster arrangement. This model's main limitation is that it does not directly detect admixture; therefore, it is unsuitable for straightforwardly describing recently admixed populations.

2.2.4.3. Estimating ancestral sources in admixture events

Some of the strategies for inferring population structure reported above have been tentatively applied to identify admixture (i.e. PCA), despite not being designed for estimating ancestry proportions explicitly. For these reasons, other methods have been developed and applied.

ADMIXTURE has been widely used for ancestry estimation (Alexander et al 2009). Two main strategies can be applied: unsupervised and supervised. With the former, the software groups subsets of individuals' genomes into K partitions, and the optimal number of groups is inferred exploiting a cross-validation approach based on SNPs resampling. The supervised assignment is based on pre-established potential sources of admixture, enabling to obtain the percentages of ancestry associated with the specific sources.

However, results from this method could be subject to over-interpretation, as thoroughly described by Lawson and colleagues in a study published in 2018 (Lawson et al. 2018). This work showed evidence that STRUCTURE/ADMIXTURE results can be misleading. For example, reconstruct genetic histories, suggesting admixture events, for groups that did not undergo admixture in their recent history but were subjects of recent genetic drift or diverged in other ways from the underlying inference model. They are also greatly affected by sample size. Specifically, populations that include fewer individuals or have experienced little population-specific drift of their own will probably fit as mixes of multiple drifted groups rather than assigned to their own ancestral population. Moreover, suppose an ancient sample is included in a modern individual dataset. In that case, it is typically depicted as an admixture of modern populations, which can occur even if the individual sample is older than the modern populations' split date and therefore cannot be admixed (Lawson et al. 2018).

Another crucial analysis is the Local Ancestry (LA) inference to estimate the ancestry proportions of admixed populations. Local ancestry is described as an individual's genetic ancestry at a specific chromosomal position, where an individual can have 0, 1 or 2 copies of an allele derived from each ancestral population. Numerous LA methods have been designed harnessing various statistical algorithms such as Principal Component Analysis (PCAdmix (Brisbin et al. 2012)), Hidden Markov Models (HMM) (HapMix (Price et al. 2009), LAMP-LD (Baran et al. 2012), ELAI (Guan 2014), MOSAIC (Salter-Townshend and Myers 2019)) and machine learning group tools (RFMix (Maples et al. 2013)).

In detail, RFMix, one of the most used LA methods, is a discriminative method that matches segments in admixed individuals to those from the source populations. It works by splitting each chromosome into windows and inferring local ancestry within each window by applying a conditional random field (CRF) that has been parameterized by random forests trained on the reference panels. The ancestries assigned to the windows within the admixed chromosomes are used to refine the haplotype patterns in the ancestral populations and improve the accuracy with an expectation-maximization (EM) step. RFMix was tested on simulated African Americans with mean Native American ancestry of 0.56% to examine if

this approach could identify a low-occurrence ancestry. When using a 99.9% confidence threshold, a mean Native American ancestry slightly over 0.44% of the total ancestry of African Americans was inferred, thus validating the initial estimate (Maples et al. 2013).

However, to infer the ancestral proportions at a sub-continental level, other haplotype-based methods are required. For example, the co-ancestry matrix obtained with the chromosome painting approach can be harnessed to deconvolute the haplotypic profile of individuals or populations into their putative sources. The first approach developed is based on Non-Negative Least Square (NNLS), and it was proposed by Hellenthal et al. (Hellenthal et al. 2014) and applied to infer ancestry proportions in various populations, including the United Kingdom (Leslie et al. 2015) and admixed American populations (Montinaro et al. 2015). This method allows to break down the ancestry of a specific group or individual as a mixture of each potential donor group, usually clusters inferred by fineSTRUCTURE, by comparing the copying vectors of donor and recipients.

In 2018 Chacon-Duque and colleagues presented a new methodological approach based on a Bayesian algorithm, called SOURCEFIND, that allows increasing the resolution to detect real contribution from background noise compared to the NNLS (Chacón-Duque et al. 2018). Through simulations and real data, they tested the accuracy and robustness of both methods' ancestry estimations. Before running SOURCEFIND, the individual similarity profiles are summarized in terms of the donor individuals, preferably grouped according to the clustering provided by fineSTRUCTURE (Figure 7). The individual donor values are summed according to these groups, and the new value is defined as a "copying vector". The donor clusters employed are called surrogates because they are surrogates for the unknown populations that historically contributed ancestry to the recipient populations under study. Then, each recipient individual is modelled as a weighted mixture of the surrogates' copying vectors (Hellenthal et al. 2014; Leslie et al. 2015). The authors recommended performing at least 10 independent runs and combining them for each recipient individual; then, extract and average the estimates with the highest posterior probability and weight them by their posterior probability.

2.2.4.4. Dating admixture events

Admixed populations should preserve DNA segments from all contributing source groups whose dimensions' decrease over time and generations due to recombination. In detail, the process of recombination tends to break haplotype chunks into smaller fragments as time and generations pass, in this way reducing LD. This process is used to date admixture events because these LD patterns will reflect haplotype chunks tracing their ancestry back to the populations involved in the admixture (Hellenthal et al. 2014).

Nowadays, the main available methods to date admixture are ROLLOFF, ALDER, MALDER, GLOBETROTTER, DATES and TRACTS.

In 2011 Moorjani and collaborators developed ROLLOFF that computes the time since mixture using the rate of exponential decline of admixture LD (Moorjani et al. 2011; Patterson et al. 2012). It computes and fits the correlation between a (signed) statistic for LD between a pair of markers and a weight that reflects their allele frequency differentiation in the ancestral populations. An extension of this general methodology was released a few years later (ALDER (Loh et al. 2013)), with the most notable improvements being a new design of the weighted LD statistic that is more robust and makes the weighted LD curve interpretable, a modification of the statistic that calculates unbiased weighted LD using the test population itself as one reference, and a statistical test for admixture. MALDER (Pickrell et al. 2014) also allows the inference of a complex admixture fitting a curve involving multiple genetic exchange dates.

However, a limitation of these three methods is the definition of the source populations that initially contributed to the admixture processes. The reference populations may have considerably diverged concerning the sources or could not descend from the same population. To bypass this limitation, Hellenthal and colleagues developed a new methodology implemented in GLOBETROTTER (GT) which relies on similar theoretical reasoning of weighted LD methods but gets power by using shared haplotype between populations instead of allele frequencies and modelling reference populations as a mixture of several possible "donors" (Hellenthal et al. 2014). GT also accounts for two admixture events by fitting a model with a mixture of LD curves with different decay rates and a diverse set of source populations, estimating the time frame for both admixtures. These improvements provide a more realistic approach to human population history.

In detail, starting from the haplotype similarity patterns obtained with CP, it is possible to use the NNLS to model them as weighted mixtures of the donor populations (Leslie et al. 2015). This modelling enables us to represent the source populations as mixtures of the sampled reference groups, inferring the source rather than fixing it. The target population or individual is then represented as a mixture of the surrogate sources' profiles estimated by the software. The size and the distribution of the segments matching to every source are estimated using another CP output (samples.out) that contains the haplotype matching of every SNP for 10 different samples of the hidden state (i.e. which donor is copied at each SNP) taken from the HMM. This information is then used to produce coancestry curves for each pair of donor populations, plotting genetic distance on the X-axis against a relative probability that measures how often a pair of haplotype segments separated by a given amount of genetic distance correspond to different donors. These probabilities are calculated using information for every pair of SNPs located from 1cM to 50cM from each other. In a single admixture event over a limited period, the decay is expected to be exponential. In the case of multiple admixture events, the decay is expected to be equal to a sum of exponential distributions, one curve per admixture event. GT determines whether the LD decay curves among all surrogates' pairings can be fitted using a single

exponential distribution or whether they are significantly better fitted using the sum of two exponential distributions (Hellenthal et al. 2014).

This approach is valuable for studying admixed American populations as the ancestry proportions at the continental level vary enormously, considering the difficulty in defining a homogeneous population made up of recently admixed individuals.

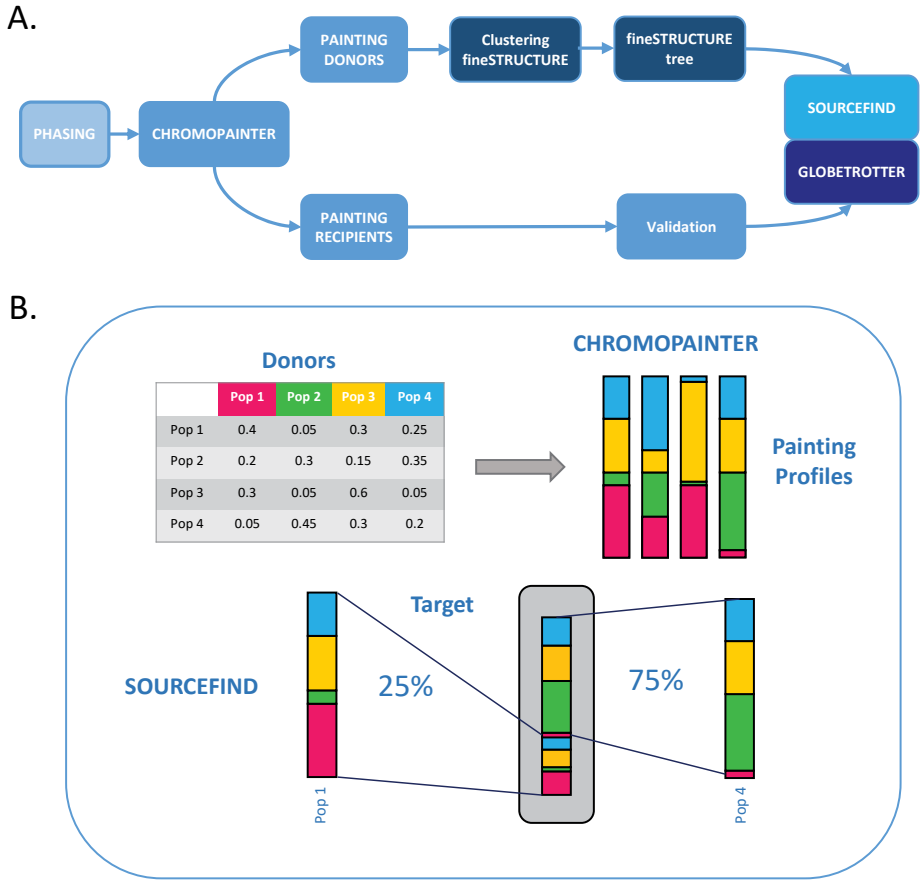

Figure 7. Example of a pipeline involving haplotype-based methods. **A)** Workflow. **B)** Schematic representation of chromosome painting and ancestral proportion estimation.

2.2.4.5. Investigating ancestry-specific demography

Besides detecting and dating admixture events, it is also essential to consider the demographic implications of admixture. The populations' sizes changed through time because of different factors, such as population growth, migration, bottlenecks, epidemics and natural catastrophes. If we examine admixed populations, we must consider the effective population sizes of the ancestral population involved in the admixture in addition to the overall effective population size.

The term “effective population size” (Ne) refers to the number of individuals an idealised population (population with random mating, with constant population size over discrete generations, and with no selection, no migration and no mutation) should have to produce a number of descendants equal to the real size of the population of interest. It represents a genetics-based measure of population size and is a critical factor in evolutionary genetic processes, such as drift and natural selection.

In 2018 Browning and Browning published a study investigating the American populations’ ancestry-specific effective population size (Browning et al. 2018). They proposed an extension to the IBDNe method that exploits two relevant methodologies: Local Ancestry (LA) and IBD segment detection. They used inbreeding effective population size described in terms of coalescence probability, which is the probability that a given pair of haplotypes is descending from a single one in the past.

As already discussed above, the LA inference can be performed using many different methods, such as RFMix (Maples et al. 2013) that, exploiting a random forest algorithm, allows to model ancestry on an admixed chromosome given observed haplotype sequences of known or inferred ancestry. As previously stated, IBD sharing can be used to estimate genetic relatedness in population samples, but it could also estimate recent effective population size. When analysing genotype data, it is possible to detect segments of IBD that are inherited, within the past generations, from recent common ancestors. The numbers and lengths of IBD segments contain information about coalescence probabilities. If the number of IBD chunks is high, a higher number of coalescence events have happened, suggesting that the coalescence probability is high, and therefore the effective size is low. Likewise, if the number of IBD segments is low, the effective size is high. These relationships can be quantified mathematically to estimate effective size, including estimating effective population size changes over time.

To estimate the Ne, it is required to consider both the length of detected ancestry-specific IBD segments that gives information about coalescence probabilities and the overall IBD segment’s length, which provides information about the coalescence time. The IBDNe software assigns each IBD segment fractionally to various coalescence times, measured in discrete numbers of generations, depending on the IBD segment length and the current estimates of effective population size history. It then uses the sum of fractional counts across IBD segments for each generation to re-estimate the effective population size of that generation. For ancestry specific effective size, it is necessary to consider pairs of haplotypes of the particular ancestry, those pieces of IBD chunks that belong to that ancestry, but the full IBD segment lengths to estimate the coalescence times are used. One way to achieve this would be to weight the IBD segments by their proportion of the given ancestry. However, this would add complexity to the IBDNe software, so alternatively, it randomly assigns the IBD chunk to an ancestry based on its ancestry proportions (Browning et al. 2018). IBDNe was tested on simulated and real data, particularly on admixed American populations, and this is what I have done in this thesis.

In addition, it is also possible to investigate the demographic changes related to uniparental systems, such as mtDNA and Y chromosome. One of the main software used for this purpose is BEAST, a cross-platform program for Bayesian analysis of molecular sequences, using an MCMC algorithm (Drummond et al. 2012). A significant output of BEAST is the demographic trend, which is achieved by the Bayesian skyline plot (BSP). The BSP model uses standard MCMC sampling procedures to estimate a posterior distribution of effective population size (Ne) through time, directly from a sample of nucleotide sequences, given any specific nucleotide-substitution model. The demographic history reconstruction includes estimating the genealogy and inferring the Ne at different points along with the genealogical timescale (Ho and Shapiro 2011).

2.2.5. Modern American populations ancestral portrait

Given their historical and epidemiological implications, a substantial number of research studies investigated the genetic diversity of modern admixed American populations (Bryc et al. 2010; Moreno-Estrada et al. 2013; Moreno-Estrada et al. 2014; Bryc et al. 2015; Homburger et al. 2015; Montinaro et al. 2015; Han et al. 2017; Chacón-Duque et al. 2018; Gouveia et al. 2020; Micheletti et al. 2020).

In the beginning, the preponderance of them focused on a single country or small areas, and they were mainly interested in reconstructing the variation in continental ancestries (African, European and Native American). For instance, Bryc et al. (Bryc et al. 2010) analysed individuals from Ecuador, Colombia, Puerto Rico, Dominican Republic and Mexico using allele frequency methods, such as PCA and STRUCTURE, to investigate genome-wide patterns of African, European, and Native American population structure within and among Hispanic and Latino populations. Moreno-Estrada et al. (Moreno-Estrada et al. 2013) investigated the population genetic history of the Caribbean, developing a new method called Ancestry-Specific Principal Component Analysis (AS-PCA), which allowed an increase in resolution using phased data to infer local continental ancestries to mask specific ancestries. This approach allows the use of PCA to examine differences at the sub-continental level. When this method was applied in the Native American component of populations from Central and South America, it consistently shows that these populations' ancestry is most closely related to natives sampled in the same areas (Moreno-Estrada et al. 2013; Homburger et al. 2015; Conley et al. 2017). In 2014 Moreno-Estrada and colleagues detected this structure even within a single country, Mexico, suggesting that the degree of population structure present in Native American populations is more significant than initially believed (Moreno-Estrada et al. 2014).

As for European ancestry, recent studies started to unravel its sub-continental contribution to American populations. As expected, in central and south America, most European ancestry derives from the Iberian Peninsula (Bryc et al. 2015; Montinaro et al. 2015; Browning et al. 2016; Conley et al. 2017). However, in Brazil, they found a contribution from Italy, mainly in the South and South-East

of the country, together with a contribution from North-western European ancestry, especially Germany in the South (Kehdy et al. 2015; Chacón-Duque et al. 2018). The Italian ancestry was also observed in Argentina, and this is consistent with the historical records about the recent migrations from Italy to Argentina starting from the end of the XIX century (Homburger et al. 2015).

A recent survey (Chacón-Duque et al. 2018) analysed more than 6000 Latin American individuals and presented a novel haplotype-based method already described above, SOURCEFIND. Besides other findings, they unveil a notable and widespread East/South Mediterranean, mainly related to Sephardic, ancestry across the region, probably connected with the persecution of non-Christians in Spain during the colonial period.

Moreover, several studies (Moreno-Estrada et al. 2013; Homburger et al. 2015; Montinaro et al. 2015; Gouveia et al. 2020; Micheletti et al. 2020) present a continental-wide analysis of the origin and dynamics of the African diaspora into the Americas. Most of these studies have found a predominant contribution of North-west and West-central Africa populations as primary ancestry sources, with smaller contributions from East and South African areas. Gouveia et al. (Gouveia et al. 2020) observed that West-Central Africa and Western Africa contributions are more prevalent in the Americas' northern latitudes. In contrast, the South/East Africa-associated ancestry cluster is more prevalent in the Americas' southern latitudes. Furthermore, some regional variation has been reported, with a higher amount of South and East African ancestry in Brazil, and more precisely in the southern part of the country (Kehdy et al. 2015).

In conclusion, these research studies highlight the complexity and high population structure of admixed American populations.

2.2.6. The Istmo-Colombian area ancestral portrait

Nowadays, the Panama population comprises Native Americans and mixed individuals. According to the last census in 2010, the ethnic groups make around 12% of the total population. The main ethnic groups are the Ngäbe (also known as Ngöbe), Kuna (or Guna) and Embera', while smaller groups are the Wounaa'n, Bribri, Naso (also known as Teribe) (Perego et al. 2012).

The Panamanian population's genetic composition has been investigated employing the two uniparental markers, the chrY and the mtDNA; mainly, the Native American component has been investigated.

The Y chromosome Native American component, the haplogroup Q, is found with a frequency greater than 50% only in three provinces of Panama: Bocas del Toro, Colon and the comarca of Kuna Yala (Grugni et al. 2015). In the rest of Panama (especially in the general admixed population), the main component is represented by western Eurasian haplogroups, reflecting the solid male genetic impact of European colonization; it is also represented by Sub-saharan African lineages a consequence of the Trans-Atlantic slave trade and more recent migrations. This dominance of the main founding haplogroup Q and its haplotype

differentiation shows that the Paleo-Indians have rapidly inhabited Panama. Then there was a local evolution of the male Native ancestral gene pool (Grugni et al. 2015) due in particular to the dramatic event of the European colonization.

Unlike the Y chromosome, the mitochondrial Native American legacy in Panama comprises more than 80% of the analyzed lineages (Perego et al. 2012). Other minor haplogroups are from Sub-Saharan Africa and Western Eurasia (mainly Europe), confirming the impact and the relevance of the European colonization on this country, but with a minor effect on maternal inheritance. This Native American percentage is composed of all the typical "Pan-American" haplogroups (A2, B2, C1, D1), but none of the rare Native American lineages (D4h3a, X2a, C4c) is present. The main haplogroup is A2 (more than 50% of the Panamanians belong to it), which is also the most common native lineage in Central America (Perego et al. 2010). The age estimate of A2 is at 15–19 kya (Fagundes et al. 2008; Perego et al. 2009; Kumar et al. 2011), suggesting Panama's initial settlement occurred rapidly after the initial peopling of the Americas. This hypothesis is also supported because the most common A2 sub-clade (A2af) has been dated more than 10 kya, testifying for an ancient colonization and settlement of this region and a maternal legacy between modern Panamanians and Paleo-Indians (Perego et al. 2012). Furthermore, these data are in concordance with the Pacific coastal route as the primary path, followed by the American continent's first settlers.

2.2.7. Sex-biased admixture

As already highlighted in the previous section by the analysis of uniparental markers in Panama, the present-day human genetic diversity of the Americas has been affected by a general imbalance in the proportion of disembarked males and females documented during the Slave Trade and the Colonial Era and, although less striking, in more recent times. Therefore, this imbalance may have left a signature on modern-day populations' genomes characterised by relatively high admixture levels. The strong male bias in migration at the beginning of the colonial settlements made the mating between European men and Native women a common feature.

In the last decades, the analysis of uniparental systems and the evaluation of continental proportion ratio of autosomal and X chromosome revealed that several American populations trace most of their paternal ancestry to Europeans and their maternal ancestry to Native Americans (Alves-Silva et al. 2000; Carvajal-Carmona et al. 2000; Carvajal-Carmona et al. 2003; Ruiz-Linares et al. 2014; Grugni et al. 2015). The general observation across many populations is that the proportion of European ancestry using Y chromosome markers is consistently more prominent than the proportion assessed with mtDNA. Conversely, Native American and African ancestries are larger when estimated with mtDNA markers (Adhikari et al. 2016). However, this imbalance's consistency and degree are

variable, suggesting that other variables, such as cultural and social practices, may have played a significant role in shaping this bias.

Autosomal data have allowed comparative analysis between autosomal markers and X chromosome, displaying lower ratings of European ancestry in the X chromosome compared to the autosomal ones because women contribute two X chromosomes to the offspring while men only contribute one (Moreno-Estrada et al. 2013; Bryc et al. 2015; Homburger et al. 2015; Kehdy et al. 2015).

However, no investigation has evaluated sex imbalance using haplotype data so far, which contains substantially more information than genotypes and may help shed light on this complex process.

3. AIMS OF THE STUDY

The three studies included in this dissertation shared a primary topic, the genomic history of American populations. The first two publications' main goal is to shed light on the migrations that lead to the first peopling of the American continents and subsequent population movements. The first one provides an in-depth phylogenetic and phylogeographic analysis of the Pan-American Y chromosome haplogroup Q. The second employed both ancient and modern DNA to reconstruct the genomic history of the Isthmo-Colombian area. The third publication aims to examine how more recent migrations shaped the genomic background of admixed American populations. The specific goals established in the three publications are outlined below.

Aims of the first reference (Ref I):
- Investigate from a male perspective the genetic history of South America through a fine dissection of the Pan-American haplogroup Q.
- Reconstruct a comprehensive and detailed haplogroup Q phylogeography and that of its sub-lineages.

Aims of the second reference (Ref. II):
- Define the genomic background of indigenous populations of Panama in order to evaluate the intra-Isthmus variability.
- Shed light on the genomic history of pre-Columbian Americans assessing the connection between the Isthmo-Colombian area and the rest of the Americas.

Aims of the third reference (Ref. III):
- Evaluate the demographic and genomic impact of the European colonization, the African slave trade and the more recent migration in shaping modern American populations genetic background.
- In detail, this study aims to reconstruct the fine-scale ancestry composition, estimate the time of admixture, examine the demographic evolution of different continental ancestries after the admixture and assess the extent and magnitude of sex-biased gene-flow dynamics.

4. MATERIAL AND METHODS

The theoretical aspect of applied methods is described in the Literature Overview section of this thesis. At the same time, the origin of individuals' DNA analysed, together with the experimental and computational methods applied, are described in detail in the respective publications and/or their supplementary materials.

The newly published DNA samples were obtained from the individuals after receiving informed consent following the guidelines of the ethical committees of the institutions involved.

In the first study, 34 unrelated males belonging to Native American haplogroup Q were selected based on their geographic origin and Y-STR haplotypes from Battaglia et al. (2013). For all of them, high-depth re-sequencing of a large MSY portion (3.7 Mb) was carried out by Beijing Genomics Institute (BGI) in Shenzhen (Guangdong, China); of these, 1.5 Mb were considered for the analysis. These samples were compared with 115 modern (Balanovsky et al., 2017; Hallast et al., 2014; Karmin et al., 2015; Mallick et al., 2016; Raghavan et al., 2014; The 1000 Genomes Project Consortium, 2015; Zhou et al., 2013) and three ancient Y chromosomes (Rasmussen et al, 2010; 2014; 2015) gathered from the literature. Two sequences belonging to Hg R1b, the sister clade of haplogroup Q (Scozzari et al., 2014; Underhill et al., 2015), were also included in our dataset. The distribution of the main sub-haplogroups identified with the phylogenetic analysis was investigated by combining literature data with the classification of 409 samples of our dataset, Native Americans (320) and Eurasians (89), obtained by hierarchical genotyping of the main haplogroup-defining-markers (Ref. I).

In the second study, we obtained the first 12 low-coverage ancient genomes from the Isthmo-Colombian area, specifically excavated in two archaeological sites, Coco del Mar and Panama Viejo, both located in Panama City. Ten of these are pre-contact radiocarbon dated from 600 to 1430 CE. We also analysed for the first time genomic data from five indigenous (Bribrì, Emberá, Guna, Naso Djërdi and Ngäbe) and two admixed populations (self-identified as "Mestizo" or "Hispano-Indigenous" and others who identified themselves as "Morenos" or Afro-descendants), for a total of 84 from the Isthmus of Panama (Figure 1 in Ref. II). The new data were compared with a worldwide dataset of 1486 modern individuals (rWD1560) and 333 ancient Siberian and American individuals (Ref. II).

In the third study, we assembled a genome-wide dataset from the literature of 17,722 individuals genotyped with different Illumina platforms (Data S1 of Ref. III). Of these, 11,607 individuals belong to 22 admixed American populations, and they were employed as 'recipients'; the remaining 6,115 individuals come from 239 worldwide source populations, and they were considered 'donors'. Detailed filtering steps and analyses are described in the respective section of Ref. III.

5. RESULTS AND DISCUSSION

This section gives an overview of the three scientific publications that constitute the original part of this dissertation. The following is a compressed summary of the main results and discussions. More detailed information can be found in the publications and their respective supplementary information.

5.1. The dissection of Y chromosome haplogroup Q shed light on the peopling of the Americas (Ref. I)

5.1.1. Refined phylogeny of haplogroup Q

To refine haplogroup Q phylogeny, we compared 154 (152 belonging to Hg Q and 2 to Hg R) Y chromosome sequences, of which 34 presented in this work for the first time, with the A00 sequence. From this analysis, a total of 1,550 variant positions were identified. Also, we considered informative SNPs outside the studied regions, which have become available from literature (Karmin et al., 2015, Poznik et al., 2016) and/or from genealogy websites (ISOGG tree; YFull tree). The relations between variants are shown in the maximum parsimony phylogenetic tree in Figure 8 obtained using Network and MEGA software (Bandelt et al. 1999; Tamura et al. 2013). Mutations outside the studied regions are indicated in Italics in the tree.

The phylogeny shows four main bifurcations identifying five main branches characterising haplogroup Q: Q-L275, Q-F1096, Q-Y2659, Q-L330 and Q-M1107. The first four lineages are almost entirely carried by Eurasian Y chromosomes, whereas all branches marked by M1107 mutation harbour virtually only Native Americans. The only exception is represented by Q-L804, characterised by an English Y chromosome in the tree (Hallast et al., 2014). The divergence dating was obtained for each haplogroup Q sub-lineage using the Bayesian approach developed in BEAST (Drummond et al., 2012). The first split, distinguishing the Eurasian Q-L275 branch, occurred before 26 kya. Despite the early phylogenetic separation of this Eurasian branch, the age of its MRCA has been evaluated around 15 kya (14.0 ± 2.2 kya, this study and 15.1 ± 1.2 kya, Balanovsky et al., 2017). In particular, in our dataset, Q-L275 includes two groups of chromosomes with relatively young MRCA (Q-M378: 6.6 ± 1.3 kya; Q-Y1150: 6.1 ± 1.2 kya). The second subdivision generates Q-F1096, dated 19.3 ± 2.6 kya. It splits into two main branches distributed across Eurasia and the Middle East: Q-M25 (12.4 ± 2.2 kya) and Q-F746 (14.9 ± 2.2 kya). Both branches are further subdivided. Q-M25 splits into Q-YP4385 (6.9 ± 1.5 kya) present in samples from India and Pakistan and Q-L713 (2.2 ± 0.8 kya) by samples from Uzbekistan and Iran; Q-F746 is subdivided into Q-M120 (4.6 ± 0.9 kya), mainly represented by individuals from East Asia, and into Q-YP1500 (8.4 ± 1.3 kya), comprising Siberian Y chromosomes and the Greenland ancient DNA of Saqqaq (4 kya, C14 dated).

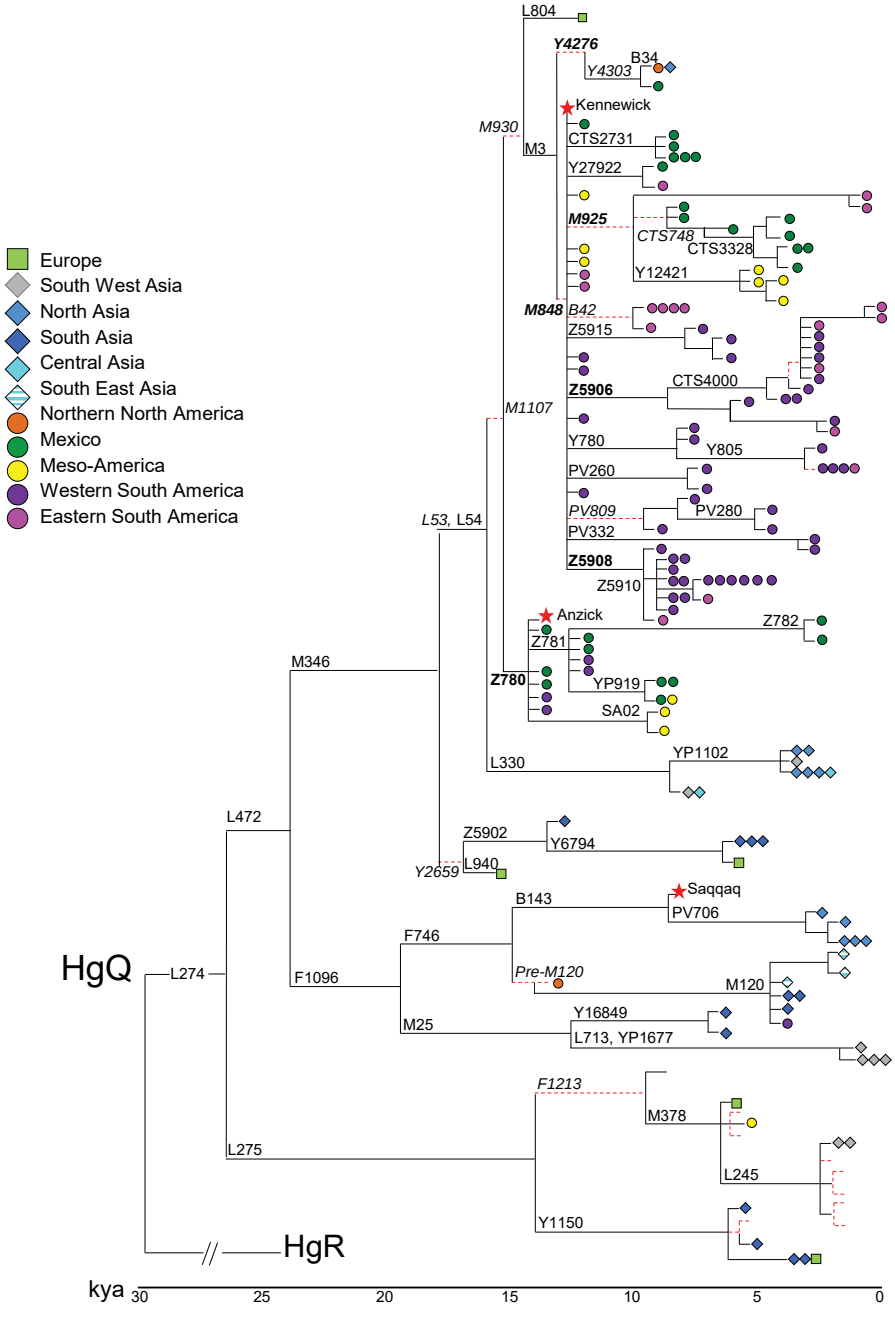

Figure 8. Condensed version of the most parsimonious (MP) tree of Y-chromosome haplogroup Q. This tree was obtained by considering 1550 variable positions in 149 modern and 3 ancient Y chromosomes. The length of each branch is proportional to its age estimate. The name of the marker(s) defining the branches is shown above them. Markers reported in italics are outside the sequenced fragments, and the relative branches are reported as dashed red lines. Stars indicate ancient Y chromosomes, squares Europeans, rhombi Asians and circles Native Americas; colours of the symbols are according to the macro-areas defined in Table S2 of Ref. I. Modified from Figure 1 of Ref. I (Grugni et al. 2019) under the terms of the Creative Commons Attribution 4.0 International License (https://creativecommons.org/licenses/by/4.0/).

The third bifurcation, which occurred around 17 kya, distinguishes the Q-Y2659 branch. This clade has been dated 16.8 ± 2.1 kya. It includes the lineage Q-L940 observed in one Ukrainian individual (Ref. I) and in one Dutch of Karmin et al. (2015) (not included in this study) and the clade Q-Z5902 (13.4 ± 1.9 kya), which, except for one Croatian Q-YP1600, harbours mainly chromosomes of Asian origin. The fourth branching, which risen about 16 kya, splits Q-L54 into Q-L330 (MRCA 8.3±1.5 kya), diffused in Asia, and Q-M1107 (MRCA 15.2±1.7 kya), observed in North Eurasia and predominant in the Americas. The latter split into the Native American branch Q-Z780, characterising the Anzick-1 aDNA (12.6 kya, C14 dated) and Q-M930. This branch is divided into Q-L804, spread in North Europe, and Q-M3, one of the two main founding lineages of Native Americans, together with Q-Z780. Finally, Q-M3, which characterises the Kennewick aDNA (9 kya, C14 dated), differentiated sometimes after 14 kya into Q-Y4303 and Q-M848. While the first branch harbours both Siberian and Native American Y chromosomes with an MRCA dated 9.3±1.2 kya, Q-M848 is Native American specific. It is dated 12.5±1.6 kya and includes eleven main sub-branches and ten singletons.

5.1.2. Reconstructing the first peopling of the Americas from the male perspective

Through the refinement of haplogroup Q phylogeny, the only Pan-American chrY haplogroup, and the hierarchical genotyping of the principal haplogroup-defining markers in 409 individuals, we confirmed the Asian origin of Native Americans. Moreover, we provided additional information about the main Asian-American migrations, together with suggestions of the occurrence of "unsuccessful" migration attempts and back migrations from a male perspective. Q-M1107 provides crucial information about the Americas' first peopling encompassing most Native American Y chromosomes. Its sublineage Q-Z780 is observed only in the Americas, whereas Q-M930 includes both the Pan-American Q-M3 and the Northwest European Q-L804. Therefore, it is possible that Q-Z780 and Q-M930 were present in the ancestral Asian/Beringian source population that gave rise to Native Americans. Q-Z780, whose age estimate appears to overlap (14.3 ± 1.6 kya), the melting of ice sheets (14.5–15.5 kya) was most likely brought by the first settlers of the American continents, rapidly expanding southward following the Pacific coastal route. Furthermore, it has been observed in the Y chromosome of the ancient individuals Anzick-1 (12.6 kya) in North America and four individuals dated 8.3–3.3 kya in South America (Posth et al. 2018).

Q-M930 separated into Q-L804 and Q-M3 during the Beringian standstill. Carriers of the first branch moved to the west from Beringia, and they reached Northern Europe, while Q-M3 entered the Americas, differentiating into Q-M848 and Q-Y4276. Q-M848 likely also moved southward along the Pacific coast and is probably associated with the "Southern Native American" (SNA) component (Scheib et al. 2018; Moreno-Mayar, Vinner, et al. 2018; Posth et al. 2018). In

Central America, Q-Z780 and Q-M848 present clades older than 10 kya, confirming the region's rapid and nearly simultaneous arrival. Only Q-M848 is well represented in South America, presenting different potentially "area-specific" groups with coalescent times around 8.0 kya. This haplogroup characterizes nearly the totality of the ancient samples (Figure S10 of Ref. I) from North America (Rasmussen et al. 2010; Moreno-Mayar, Potter, et al. 2018), the Californian Islands (Scheib et al. 2018) and Patagonia (de la Fuente et al. 2018).

Q-Y4276 likely arose in Beringia, quickly evolving into Q-Y4303 in northern North America as suggested by its estimated age (9.3 ± 1.2 ky) (Figure 3A.2 of Ref. I). This marker, Q-Y4303, was spread in the southern part of California and Mexico, while its sub-branch Q-Y4300 arrived in the eastern part of northern North America, characterizing Algonquian groups. An early differentiation in this area of Q-Y4303 would also explain the distribution of its sub-branch Q-B34 (5.4 ± 1.2 kya) in the northern area of North America, including its presence in two ancient samples from Quebec and Alaska (Scheib et al. 2018). However, the presence of Q-B34 also in the Koryak of Siberia can be easily attributed to a back migration (Figure 3A.2 of Ref. I), as postulated for mtDNA haplogroup A2a (Achilli 2013). Q-M25, the sub-lineage of Q-F1096, is frequent in modern Western Eurasians (Grugni et al. 2012; Di Cristofaro et al. 2013), and it is found in ancient samples from the Beringian area (Flegontov et al. 2019). This could indicate that, during the warmer mid-Holocene period, populations carrying different haplogroup Q lineages reached the former Beringian area but gave a minor contribution to the modern Y chromosome gene pool. Differently, Q-F746 is common in Southeast Asia as Q-M120 and encompasses the Q-B143 lineage (8.4 ± 1.3 ky), which characterizes the Saqqaq Paleo-Eskimo ancient individual (4 kya), and a new branch Q-PV706 (2.8 ± 0.9 ky) found in a few Koryaks of Northeastern Siberia. Q-B143 would trace the Paleo-Eskimo migration at around 4 kya. In this scenario, the Q-F746 Y chromosomes observed in the North American Arctic (Dulik, Owings, et al. 2012) and not yet assessed for B143 could include Paleo and Neo-Eskimo contributions to the Arctic people. In turn, the lineage Q-PV706 was observed in the Koryaks, and it could describe either an East Asian evolution of Q-F746 or a back migration from North America as hypothesized for Q-B34.

Q-Z780 and Q-M3 experienced additional differentiation and genetic drift before entering the Americas with the first settlers. Instead, Q-F746 does not seem to have participated in America's first peopling. Thus, concerning the first peopling, the split of Q-M3 into Q-M848 and Q-Y4276 could correspond to the separation of the two main ancestral population groups (NNA and SNA). In this scenario, Q-M848 and Q-Z780 would have been carried along the Pacific coast by the population group that gave rise to the SNA ancestral component. Then, Q-Y4276 could have followed the internal route as Q-B34 and Q-Y4300 contributing, together with the mtDNA haplogroups X2a and C4c, to the NNA component that mostly appears to characterize northern Native Americans. In such a scenario, considering the back migration of Q-B34, the split of the populations ancestral to SNA and NNA would be best placed in eastern Beringia before

entering into America. On the other hand, new data (Lesnek et al. 2018) indicate that the ice-free corridor was practicable much earlier than previously considered (15.6–14.8 kya), restoring the possibility of different migration pathways of the ancestral Native American populations. The observation that the Kennewick genome, which belongs to the NNA component and carries mtDNA haplogroup X2a, is characterised by the Y chromosome haplogroup Q-M848 also suggests that the following admixture events (Scheib et al. 2018; Posth et al. 2018; Moreno-Mayar et al. 2018) must have started very early in North America and suggest a more complex admixture scenario.

5.1.3. Bayesian analysis and demography

Through the generation of Bayesian Skyline plots (BSPs), we estimated the posterior distribution of the effective population size through time for the entire sample of Native Americans (all Q-M1107 samples) and the individuals belonging to the most significant sub-haplogroups (Figure 2 of Ref. I). The analysis of Q-M1107 shows a major phase of population growth later than 15 kya, after the first peopling of the Americas, well described by the Z780 curve, followed by a period of constant population size from 8 to 3 kya and a subtle, but a not significant, sign of population growth from 3 kya. The second period of growth could be marked by M925 and Z5906, while the rapid growth of Z5908 at 8 kya seems not to have influenced the global trend of Native American populations. Interestingly, archaeological data from South America reported by Goldberg and collaborators show a similar trend (Goldberg et al. 2016): the first sign of growth associated with a resource-limited (megafauna extinction) growth over time, then 9 kya domestication slowly started in NW South America until 3 kya when there was a shift to a predominantly sedentary and agricultural subsistence. The second period of growth was probably not linked to a climatic change but rather to a cultural and technological change. However, South America's diverse environments, geographic barriers to gene flow and low population density did not lead to the diffusion of a single cohesive culture. As evident from our analyses, it seems that the cultural and technological revolution reached at that time was isolated and different for every population and, for this reason, not able to cause an evident expansion of the South American populations.

5.2. The peopling of the Americas from an archaeogenomic point of view (Ref. II)

5.2.1. Identifying "unadmixed" Indigenous American individuals

One of the main limitations of the study of pre-Columbian American history using modern data is the presence of genetically non-American components present in variable proportions, even in indigenous individuals.

Therefore, at first, we identified and removed the non-American genetic components in indigenous people. This phase was divided into several steps. The first step implied identifying the indigenous samples that held a negligible amount of indigenous components. For this reason, we carried out three different analyses: ADMIXTURE, *f*4, and finally, once we selected a group of individuals to be used as a source, RFMix for the Local Ancestry inference. From the superimposition of these three analyses' results, we have identified two datasets, the first called uIA217 with individuals containing less than 5% of non-indigenous American DNA and the second uIA89 with individuals containing less than 1% of Africa and 2% of Europe. Moreover, for the remaining samples, we masked the non-American components from the Local Ancestry output (mIA417).

5.2.2. Assessing the Intra-Isthmo genetic variability

We investigate the comparison among Isthmian individuals, both with allele frequencies and haplotype-based methods. The individuals separated into clusters that mirror the indigenous communities in the Isthmus. In particular, looking at the fineSTRUCTURE tree (Figure 9A), it is possible to distinguish five clusters. Four of them include only individuals that belong to the same indigenous group (Emberá, Guna, Naso and Ngäbe) while the Bribrì cluster together with the individuals from the Southern Atlantic Costa-Rica, and belong to the Cabecar group (Western Isthmus cluster). The ancient Panamanian individuals, included in the allele frequencies analyses, such as PCAs and the outgroup *f*3 hierarchical tree (Figure 2 and Figure S3 of Ref. II), are assembled in a different group that could be associated with the Panamanian population who lived in the Pacific Panamanian region and disappeared with the arrival of the European colonizers. Looking at the variability inside the Isthmian clusters, the Guna show the highest level of similarity in intra- and inter-cluster comparisons (Figure 9B and Figure S4B of Ref. II), and the other isthmian clusters show a high level of similarity. This result could be explained by the drift generated by the bottleneck due to the colonization, but probably occurred at different times, considering the blocks of IBD sharing inside the clusters, with the east population that show peaks shorter (7–13 cM) than the groups on the west (13–27 and 27–53 cM). Analysing the interactions between the clusters inside the Isthmus, we observed the geographic distribution of the variability. The northeast populations of the region seem to have had more interactions with each other, forming a macro cluster in the tree.

This macro-cluster largely overlaps with the geographic region of the pre-colonial Greater Chiriquı' cultural area. In the east, in the populations that occupy what was once the geographical region of the great Darien cultural area, the interactions are less extensive. The Guna show peculiar behaviour, acting almost as an outgroup to the rest of the Isthmian populations. This also reflects an unexpected demographic pattern, estimated with IBDne, that backdates the Isthmian population's decline before the colonizers' arrival. These peculiarities could derive from three not mutually exclusive reasons: i) a decline of the population in the Greater Darien region that occurs before the arrival of the colonizers; ii) a minor impact of the colonizers, and iii) a fast recovery of the population size soon after the arrival of the Europeans.

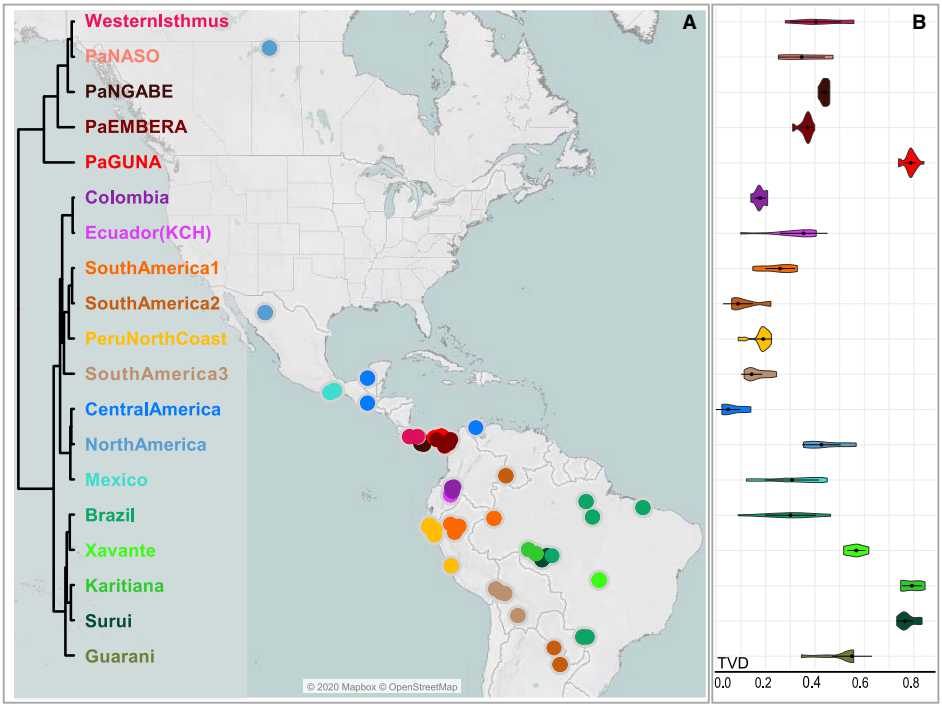

Figure 9. Population genetic structure as revealed by haplotype analysis of modern Panamanian and Indigenous American populations. **A)** fineSTRUCTURE unrooted dendrogram showing the 19 identified Indigenous clusters and the geographic distributions of the individuals in the nearly unadmixed IA (uIA217) dataset. **B)** Violin plot showing cluster self-copy lengths (fragments copied from members of their own cluster) in the uIA217 dataset; higher values are for more isolated groups. Modified from Figure 3 of Ref. II (Capodiferro et al. 2021) under the terms of the Creative Commons Attribution 4.0 International License (https://creativecommons.org/licenses/by/4.0/).

5.2.3. Reconstructing the first peopling of the Americas from a genomic perspective

After outlining the genomic variability of Isthmus, we looked at it within the Americas' genomic scenario. Since the first exploratory analysis, the PCA of the American individuals and the worldwide ADMIXTURE, a specificity of the Isthmo-Colombian region's populations emerged, characterized by a specific cline shown in the PCA and by a specific component in the ADMIXTURE analysis (Figure 2 of Ref. II). This peculiarity is also visible when PCA and ADMIXTURE are applied to a different dataset, based on Illumina data, which allows the analysis of more indigenous peoples of Costa Rica and North-East Colombia (Figure S2C of Ref. II). Even when haplotype-based approaches were applied, particularly fineSTRUCTURE, it was possible to observe a particular attitude. The Isthmus' genetic clusters are placed in an outlier position compared to the rest of the American variability. Moreover, the CHROMOPAINTER-based Total Variation Distance (TVD) heatmap shows that excluding Emberà, which shows a low TVD value compared to other American genetic clusters, shows a slight affinity with the rest of the dataset evident in the Guna (Figure S4B of Ref. II). When differences in interactions are investigated as an excess of allele sharing with f4, more significant interactions of the Emberà with other South American populations are displayed, and different contacts of the populations west of Panama than the Ancient Panamanian individuals and the Guna (Figure S5B of Ref. II). The peculiarity of the Isthmo-Colombian population, which we have called distinctiveness, seems to have a pre-Hispanic origin, mainly because these populations seem to have a shared long history of drift and separation from other indigenous peoples. Indeed, when through *f*4 we formally checked for an excess of allele sharing in a broad context, considering both modern and ancient individuals from the Americas, it was possible to observe a particular trend for the Panamanian populations and the Costa Rican Cabecar, with the exclusion of the Emberà. While, as expected, it has been possible to observe higher proximity of modern Panamanian populations with the other modern populations of Central and South America, when compared to Anzick, this signal is absent when using Spirit Cave and other ancient individuals, except for the oldest individuals of 7 kya on the Pacific Coast in place of Anzick. These results suggest that the Isthmus populations have equal distance with Spirit Cave and other modern Central and South American populations (Figure 6A and Table S4 of Ref. II). Besides, the Gunas always have a shared history with the ancient individuals representing the main American ancestry, very low compared to the average of other modern indigenous peoples (mostly <25th percentile) (Figure S3D of Ref. II). Using qpWave, we tested the minimum number of sources needed to describe pairs of populations compared to outgroups representing the main Indigenous American Ancestry. This analysis has allowed us to demonstrate that, especially for Gunas, two sources are needed to explain the variability when coupled to populations, not of the Isthmus, and one compared with an Isthmus population (p-value >0.01, Figure 6b of Ref. II). Starting from the results of previous analyses, we then tested a model

with qpGraph that provides for the presence of an ancient unsampled population that characterizes Isthmus populations (UPopI: Unsampled Population from the Isthmus). This model involves the first migration of a population, represented by Lagoa Santa in Brazil, which has spread to both American continents and to both coasts of South America (which we have called SNA1). Later, between the late Pleistocene and the Early Holocene, a new population (SNA2) represented by Spirit Cave in North America migrated through the continents, admixed with the previous population already present, reaching only the coasts of Pacific South America. These happened until the subsequent repopulation, with the improvement of the climatic conditions during the Holocene of all Central and South America, as described by (Moreno-Mayar et al 2018 and Posth et al., 2018). In this scenario, UpopI, which is part of SNA2, was separated before Spirit Cave (so probably in North America), and then probably also reached South America, but leaving traces to the present day, especially in the Isthmo-Colombian region.

5.3. Recent migrations contributed largely to the genomic pool of modern American populations (Ref. III)

5.3.1. Macro and micro geographic patterns of European, African and Native American ancestry in American populations

At first, we grouped the worldwide donor individuals into 89 genetically homogeneous clusters (Figure S2A, Data S2A of Ref. III), based on haplotype similarities using CHROMOPAINTER (CP) and fineSTRUCTURE (FS). Hence, we could reduce the donors' genetic heterogeneity in the ancestry characterisation process. These results revealed the worldwide genetic heterogeneity pattern at the continental scale also observed in previous surveys.

To reconstruct the American population's ancestral mosaic, we used SOURCEFIND to fit each of them as a mixture of the identified FS clusters. We found that 21 clusters contribute no less than 2% in at least one recipient population; the results are shown in Figure 10A. In detail, we found that the African ancestries distribution reflects the complexity of the Slave Trade dynamics. Our analysis revealed that West-Central Africa ancestry is the most prevalent in the American continents, as previously stated (Montinaro et al. 2015; Gouveia et al. 2020). Moreover, we identified a high contribution from Senegal and Gambia in the Caribbean, Mexico, and Colombia according to African slave arrivals predominantly to Spanish-speaking America until the 1620s (Klein 1999). Then, we estimated a high contribution from Benin and Nigeria in all the Caribbean populations and populations from the US; this is in agreement with the dynamics of the slave trade because about half of all West African slaves were deported to Dutch, French and British sugar plantations in the Caribbean. Among all the analysed populations, ACB (Barbados) is characterised by the highest sub-Saharan ancestry proportion (~88%), likely due to the presence of the sugar cane industry together with the relatively low European immigration in the XVIII century (Curtin 1972). At a

microgeographic scale, ACB individuals derive their African ancestry from "BeninNigeria" (~50%) and "BeninIvoryCoast" (~21%) (Figure 10A), two of the primary source areas reported for the British-mediated slave trade.

In contrast, Brazil showed a different African ancestral composition, distinguished by a high ancestral contribution linked to modern-day Angola and Namibia, consistent with the Portuguese settlement in Angola from the beginning of the XVII century. In Argentina, a related African component is also observed, probably due to the slaves' arrival primarily to Brazil via the Portuguese slave trade from Angola (Edwards 2014; Eltis & Richardson 2010).

Towards the end of the slave trade period, Mozambique was the third-largest supplier of slaves (Klein 1999). Therefore, we observed Southern East African ancestries in a non-negligible proportion of individuals from Bambui and Pelotas.

As for the European ancestries, we confirmed the notable impact of Great Britain, France and the Iberian Peninsula for all the examined populations, with a distribution reflecting the Americas' geographic occupation in the Colonial Era. Furthermore, our approach revealed several European secondary sources contributing to a substantial proportion of American populations. For example, Italian ancestry was also found at a considerable proportion, higher than 5%, in individuals of multiple populations, especially in Argentina and Brazil. Italy has been reported as one of the primary sources of migrants to South America during the XIX century, giving rise to the "Italian diaspora", second only to the Spanish and Portuguese influences. Genetic signals of these migrations were found in all the three Brazilian populations analysed, principally related to North Italy, with the highest proportion in Pelotas, followed by Bambui and Salvador. In Argentina, the identified Italian contribution is related both to the Northern and Southern part of the peninsula, which is in accordance with movements of millions of individuals from Northern (earlier) and Southern (later) Italy registered from the second half of 1800 throughout the 1950s (Baily & Míguez 2003; Meade 2011). Therefore, the Italian components' distribution is heterogeneous at a Pan-American level and closely reflects the one reported by historical records. Differently from the other Brazilian groups analysed, Pelotas is also characterised by contributions from additional sources, such as Central and North-Europe ("GreatBritain1", "France", "CentralEurope1-2" and "Scandinavia") mirroring historical records.

Looking at the Native American ancestry, we found that, except for Mayan individuals, this ancestry is high in South American populations and Mexico, with the highest rates in Peru, Ecuador and Argentina (Figure 10 and Data S2B of Ref. III). Interestingly, we identified a non-negligible proportion of individuals with Native American-related ancestry in both the analysed African-American population samples.

Finally, with our analysis of the individual ancestry distribution, we identified Jewish and Levantine ancestries' presence in virtually all the examined populations, including those from the Caribbean (Figure S5 of Ref. III). These results confirmed what Chacon-Duque and colleagues reported in their study (Chacón-Duque et al. 2018). They identified South and East Mediterranean ancestries in the Americas, interpreting it as a contribution from Converso Jews.

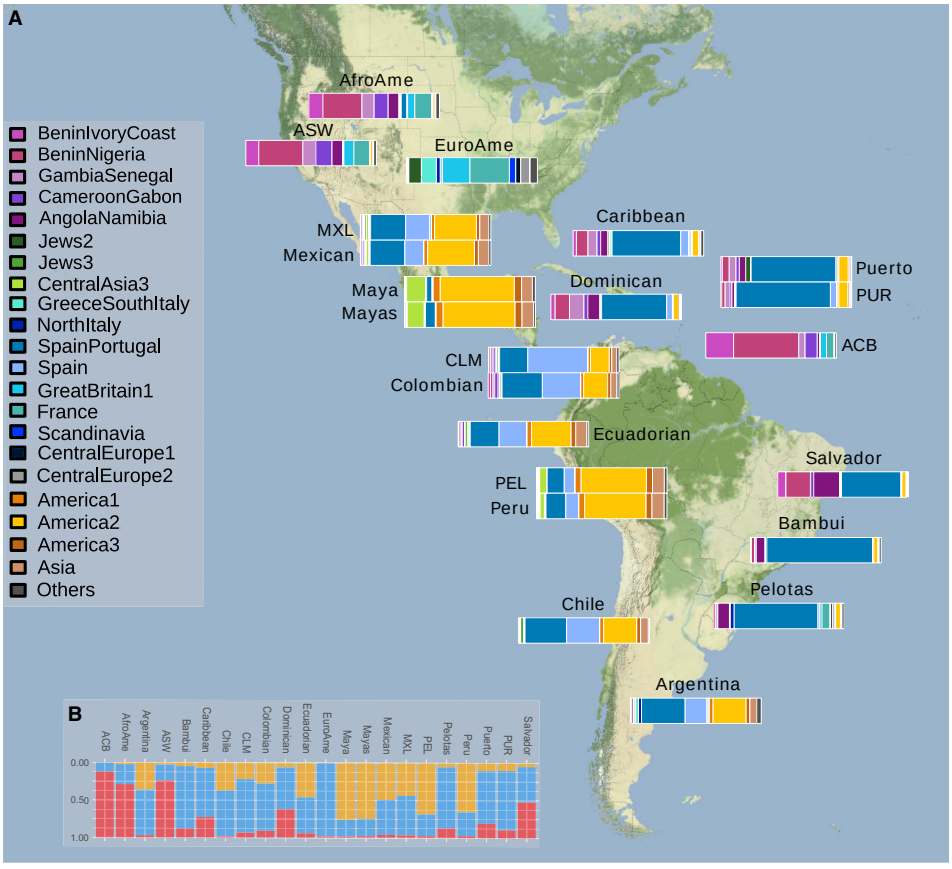

Figure 10. The ancestral mosaic of American populations reveals a highly complex ancestral composition. **A)** Barplots representing ancestral genetic proportions based on SOURCEFIND results for North and South American populations. Only the contribution for the 21 most representative fineSTRUCTURE clusters (contributing ≥2% in at least one recipient population) is reported. **B)** Proportion of continental ancestries for all target populations. Ancestries are represented in red for Africa, blue for Europe, and yellow for America/Asia. Reprinted from Figure 1 of Ref. III (Ongaro et al. 2019); permission conveyed through Copyright Clearance Center.

5.3.2. Evaluating admixture times and sources characterising modern American populations

To identify the admixture sources and events and to date these events, we applied GLOBETROTTER (GT) at two different levels, "Population" and "Individual"; we concentrated on admixture events inferred to have happened in the last 30 generations.

At the population level, we found that all American populations under study are the consequence of at least one admixture event including Native American, African, and European sources that happened within the last 6–12 generations

(considering a generation time of 28 years), corresponding to 1644 Common Era (CE) and 1812 CE (Figure 2A of Ref III). In 14 populations, we identified one admixture event with inferred times between ~6 and ~11 generations ago. The results are reported in Figure 2 and Table S2C of Ref. III. In two Caribbean populations, PUR (Puertorican) and Dominican, we found a single admixture event involving more than two sources from Africa, Europe and America, dated ~9–11 generations ago. However, the six remaining populations (Mexican, EuroAme, Pelotas, Caribbean, AfroAme and Bambui) displayed a sign of at least two admixture events mainly concerning American, European and African sources, with the most recent occurring ~6–8 generations ago.

Moreover, our per-individual analysis provided several insights into the complexity of admixture in the Americas. Interestingly, the inferred admixture dates that involve Senegal and Gambia populations are older than those involving all the others, mirroring historical data; this area remained the main slave trade site for the Spanish possessions until 1640 (Barry 1998). Likewise, all the dates concerning clusters related to Angolan and Namibian individuals are characterized by younger recent admixture times. Furthermore, times involving "BeninIvoryCoast" are significantly older than the ones involving "BeninNigeria" and "CameroonGabon". Lastly, dates regarding "CameroonGabon" are older than "BeninNigeria" (Figure 2D of Ref. III).

For European sources, the estimated admixture dates characterized by gene flow from the Iberian Peninsula are older than the dates with France/Great Britain sources, which, in turn, are older than admixture events including Italian sources, which, according to historical records, became substantial only in the second half of the XIX century (Figure 2C of Ref. III).

5.3.3. Reconstructing the ancestry-specific demography of American populations

To investigate continental ancestries' demographic history, we performed IBDNe analysis, as in Browning et al. (Browning et al. 2018).

This analysis demonstrated that, despite their composition, most of the continental ancestries underwent a general decrease until approximately ten generations ago, after which a general population size recovery was inferred (Figure 11, Figure S6G of Ref. III). According to previous surveys reporting their low heterogeneity (Kehdy et al. 2015), this pattern is not observed in all the American populations: Bambui (Brazil) showed a general decline in population size for African and European ancestry. Also, the European ancestry of European Americans (EuroAme) does not display signs of demographic decline, perhaps reflecting multiple European migrations contributing to this population.

Moreover, the African population component's recovery postdates those of the European one, possibly revealing the various conditions encountered by African slaves and European settlers.

On the other hand, the effective population size of the Native American component in Peruvians and Mexicans does not show evidence of a decrease, in contrast with historical reports describing a general dramatic decline of the Native American population after European colonization. This result matches Browning et al. (Browning et al. 2018), which found a minor decline in Mexicans effective population size for Native American ancestry compared to other populations; and also with our GLOBETROTTER results, where we saw a sign of admixture between two Native American related sources around 15 generations ago. It may be possible that the vast population decrease did not massively affect the genetic variability of survivor populations; or that individuals from distinctive isolated native groups have been placed in contact as a result of the European colonization and deportation, also recently suggested for Peru (Harris et al. 2018). This would result in an inflated effective population size estimate, as we see in our IBDNe analysis.

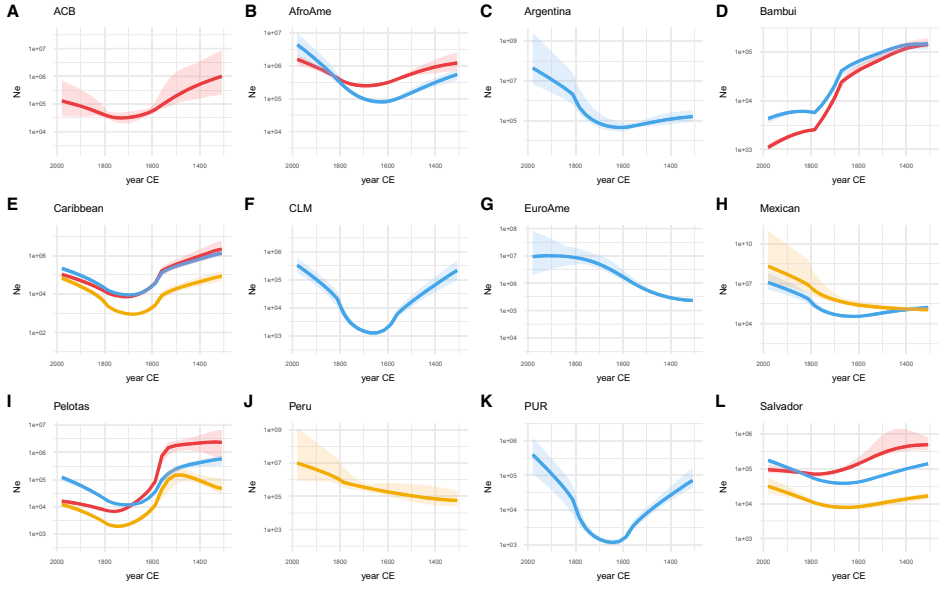

Figure 11. Ancestry-specific effective population size of American populations obtained with IBDNe. The x-axes show time expressed in years of CE. The y-axes show ancestry-specific effective population size (Ne), plotted on a log scale. Solid lines show estimated ancestry-specific effective population sizes (red, African ancestry; blue, European ancestry; yellow, Native American ancestry), with ribbons indicating the 95% CIs. Only the population ancestries, in which $\alpha(\text{continent}) \times n > 50$ (where α is the proportion of a specific ancestry and n is the total number of chromosomes in the analyzed population), are represented. Reprinted from Figure 4 of Ref. III (Ongaro et al. 2019); permission conveyed through Copyright Clearance Center.

5.3.4. Assessing the impact of sex-biased admixture in the American continents

To estimate the impact of sex-biased admixture in the American populations under study, we performed ADMIXTURE on the autosomal and the X chromosomes, comparing the related continental proportions. Concerning European ancestry, the proportions of autosomes were significantly higher in all comparisons, suggesting a higher contribution of European males than females in American populations' gene pool (Figure 3 and Data S2D of Ref. III), in agreement with previous continental-scale reports based on more limited data (Bryc et al. 2010; Kehdy et al. 2015; Fortes-Lima et al. 2017). This result is supported by the fact that the Native American ancestry proportion of autosomal data is always lower (except for Dominican) than that of the X chromosome. In contrast, when examining African ancestry, many populations do not display any indications of sex imbalance. In detail, in only eight out of 19 comparisons (ACB, AfroAme, Bambui, Caribbean, EuroAme, Pelotas, PUR and Salvador), the autosomal proportion was significantly lower than that of the X chromosome. Except for ACB, all these significant differences were associated with sample sizes greater than 100. These results contrast with historical records documenting a higher number of disembarked male slaves (Klein 1999) and might reflect complex admixture dynamics. In fact, gender imbalance in the slaves' treatment could have led to diverse chances to have descendants and consequently explain, at least in part, these results. Alternatively, they could reflect limitations in the approach employed here, as previously suggested (Goldberg and Rosenberg 2015). These results confirm that the European and African components play a fundamental role in shaping the genetic differentiation of different American groups, although their demographic evolution after the arrival in the "new world" is still unknown.

6. CONCLUSIONS

Ref. I:

- A high-resolution haplogroup Q phylogeny that presents new insights into its Eurasian and American branches' geographic distribution in modern and ancient samples was ascertained and dated.
- For the first time, two distinct Y chromosome lineages reflecting the two main ancestral components (SNA and NNA) earlier described by recent genomic studies were observed. The differentiation of these lineages probably occurred in eastern Beringia before entering the Americas through two routes: the coastal (SNA, Q-Z780/Q-M848) and the internal route (NNA, Q-Y4276). Once there, these two ancestral components probably admixed very early in North America, as suggested by the ancient Kennewick nuclear genome belonging to SNA (Q-M848) yet carrying an NNA mtDNA haplogroup (X2a).
- Two significant expansions of the SNA lineages in Meso- and South America, one around 15 kya, early after the first peopling, and another at 3 kya, following climatic changes and local cultural shifts, were revealed.

Ref. II:

- A remarkable genomic structure within Panama was identified, mainly overlapping with past and present Indigenous groups analyzed in this study. These groups also show relatedness, especially in the Caribbean region on the border between western Panama and southeastern Costa Rica over thousands of years. Fewer genetic similarities were identified between the Indigenous groups located in eastern Panama and between the Emberá and the pre-Hispanic Panamanians who lived in the area around Old Panama before European contact.
- A previously undescribed ancestry among ancient Indigenous peoples of the Americas was revealed. This ancestry is unique to the region and detectable in the ancient pre-Hispanic individuals and the self-identified descendants of current Indigenous, African and Hispano-Indigenous groups. It reached the Panama land bridge over 10 thousand years ago, expanded locally during the early Holocene, and left genomic traces up to the present day, especially among the Guna.

Ref. III:

- The European genetic contribution in American populations mirrors the geopolitical situation during colonization. Several European secondary sources contributing to a substantial proportion of American populations were revealed, e.g. Italy in Brazil and Argentina, Central Europe in Brazil. A differential contribution of African sources among American populations was inferred.
- Times of admixture are concordant with migration waves from Europe and reflect differences in African areas exploited through time.

- The investigation of the demographic impact of admixture reveals a general decline and recovery pattern in several populations under study corresponding to the beginning and the end of the Colonial Era. However, Peru and Mexico are characterized by different demographic trajectories.
- The analysis of sex-biased admixture dynamics suggests that a higher number of American females than males have contributed to the modern populations, while European males had a more significant contribution than females from the same continent. In contrast, some populations, but not all, showed evidence for a higher female contribution, partially conflicting with historical records for African ancestry.

SUMMARY IN ESTONIAN

Ameerika populatsioonide genoomne portree

Ameerika populatsioonide evolutsiooni on käsitlenud mitmed multidistsiplinaarsed uuringud. Meie teadmised Ameerika maailmajao geneetilise mitmekesisuse kujunemisest on endiselt ebatäielikud, ehkki geneetilised uuringud lisavad sel teemal pidevalt uusi detaile. Uute tehnoloogiate nagu järgmise põlvkonna sekveneerimine (NGS) väljaarendamine koos teiste tehniliste edasiminekutega avavad võimaluse eraldada ja analüüsida DNA-d iidsetest proovidest, tehes "iidsest genoomikast" (aDNA) ühe paljudest põhilistest tööriistadest meie esivanemate mineviku mõistmiseks. Veelgi enam, need tehnoloogiad on tohutult suurendanud genoomsete andmete hulka kogu maailmast, sealhulgas Ameerika mandritelt.

Ehkki Ameerika maailmajagu oli viimane, milleni meie *sapiens*'i esivanemad jõudsid, on selle geneetilise varieeruvuse protsessid olnud väga keerukad. Nende uuringud on rohkem kui kolme kümnendi jooksul olnud paljude geneetikaalaste teadustööde teemaks. Algul domineerisid Ameerika populatsioonide populatsioonigeneetilistes uuringutes uniparentaalsed geneetilised süsteemid, alustades mitokondriaalse DNA-ga (mtDNA) ja peagi kaasates Y-kromosoomi (chrY) analüüsi. Viimasest selgus, et põlisameeriklaste kaks chrY asutajahaplogruppi olid tõenäoliselt hg C ja hg Q, mida leiti vastavalt umbes 5% ja 75% põlisameerika meestest. Kuid nende haplogruppide resolutsioon ei paranenud oluliselt enne kui mõne aasta eest.

Selle doktoritöö esimese publikatsiooni (Ref I) eesmärgiks on uurida Ameerika maailmajao geneetilist ajalugu meeste perspektiivist, lahates suure täpsusastmega üleameerikalist haplogruppi Q, ning koostada kõikehõlmav ja detailne haplogrupp Q ja selle alamliinide fülogeograafia.

Uniparentaalseid geneetilisi süsteeme võib pidada kaheks lookuseks, mida kasutatakse inimese ajaloo nais- ja meesperspektiivi mõistmiseks. Nad saavad kirjeldada ainult kaht esivanemat neist tuhandetest, kes on seotud tänapäeva populatsioonide geneetilise pärandi kujundamisega. Olulisem arv esivanemaid on genoomis esindatud autosomaalsetes markerites. Seega on autosomaalsed markerid hädavajalikud Ameerika maailmajao populatsioonide liikumiste ajastuse ja dünaamika mõistmiseks. Tänu arheoloogilistele ja geneetilistele tõenditele tunnistatakse nüüdseks, et esimesed Põhja-Ameerikasse jõudnud inimesed tulid Siberist, ületades pärast hilist jääaega Beringi maakitsuse. Algsetele asulakohtadele järgnesid ulatuslikud inimeste ränded, mis jõudsid Lõuna-Ameerika lõunaossa suhteliselt kiiresti, juba ~15 000 aastat tagasi. Mitu hiljutist uuringut on selle teema kohta uut informatsiooni andnud, rekonstrueerides Ameerika maailmajao erinevate piirkondade põliselanike rühmade genoomset ajalugu, kuid Isthmo-Colombia piirkond on seni puudu.

Seega rakendab selle doktoritöö teine publikatsioon (Ref II) nii iidse kui ka tänapäeva DNA andmete analüüsi, et rekonstrueerida Isthmo-Colombia piir-

konna genoomset ajalugu. Selle eesmärgiks on teha kindlaks Panama põlispopulatsioonide genoomne taust, et hinnata maakitsuse sisest varieeruvust ja selgitada Kolumbuse-eelsete ameeriklaste genoomset ajalugu, hinnates Isthmo-Colombia piirkonna sidemeid ülejäänud Ameerika maailmajaoga.

Lisaks esialgsetele rännetele pärinevad Ameerika populatsioonid mitmest segunemisest, alates koloniseerimisest ja Atlandi orjakaubandusest. Peale selle toimus viimase kahe sajandi jooksul palju rändelaineid, millele järgnes kohalik segunemine, ning nende mõju on suuresti uurimata.

Selle doktoritöö kolmas publikatsioon (Ref III) uurib, kuidas hilisemad ränded kujundasid segunenud Ameerika populatsioonide genoomset tausta. Täpsemalt on selle uuringu eesmärgiks rekonstrueerida kõrgel lahutusastmel põlvnemise komponendid, anda hinnang segunemise ajale, uurida erinevate mandrite põlvnemise demograafilist evolutsiooni pärast segunemist ning hinnata soost sõltuva geenivoolu dünaamika ulatust ja tugevust segunenud Ameerika populatsioonides.

Käesoleva doktoritöö peamised tulemused ja järeldused on järgmised:

- Tehti kindlaks ja dateeriti kõrge resolutsiooniga haplogrupp Q fülogeneesipuu, mis annab uut informatsiooni oma Euraasia ja Ameerika harude geograafilise jaotuse kohta tänapäeva ja iidsetes proovides.
- Esimest korda tuvastati kaks eristuvat Y-kromosoomi liini, mis peegeldavad hiljutistes genoomsetes uuringutes varem kirjeldatud kaht peamist põlvnemiskomponenti (SNA ja NNA). Nende liinide lahknemine toimus tõenäoliselt Beringi maakitsuse idaosas enne Ameerika maailmajakku sisenemist, milleks kasutati kaht teed: ranniku (SNA, Q-Z780/Q-M848) ja sisemaa teed (NNA, Q-Y4276). Sinna jõudnuna segunesid need kaks põlvnemiskomponenti Põhja-Ameerikas tõenäoliselt väga vara, millele viitab iidne Kennewicki mees, kelle tuumagenoom kuulub SNA komponenti (Q-M848), kuid mtDNA haplogrupp on NNA-st (X2a).
- Avastati SNA liinide kaks märkimisväärset ekspansiooni Meso- ja Lõuna-Ameerikas, üks umbes 15 000 aastat tagasi, kohe pärast esmaasustamist, ja teine 3000 aastat tagasi pärast klimaatilisi muutusi ja kohalikke kultuurilisi nihkeid.
- Panama sees tuvastati märkimisväärne geneetiline struktuur, mis kattus üldjoontes käesolevas uuringus analüüsitud mineviku ja praeguste põliselanike rühmadega. Need rühmad on ka tuhandeid aastaid suguluses olnud, eriti Kariibi mere piirkonnas Panama lääne- ja Costa Rica kaguosa piiril. Ida-Panama põliselanike rühmade vahel ning Emberá ja hispaanlaste-eelsete panamalaste vahel, kes elasid Vana Panamat ümbritsevas piirkonnas enne kontakti eurooplastega, leiti vähem geneetilisi sarnasusi.
- Ameerika maailmajao iidsete põliselanike seas avastati varem kirjeldamata põlvnemiskomponent. See komponent esineb ainult selles piirkonnas ning on tuvastatav iidsetes hispaanlaste-eelsetes indiviidides ja inimestes, kes ise identifitseerivad end tänapäeva põliselanike, Aafrika ja latiino-põliselanike rühmade järglastena. See jõudis Panama maakitsusele rohkem kui 10 000 aastat

tagasi, levis varases Holotseenis lokaalselt ning jättis tänapäevani püsivaid genoomseid jälgi, eriti Guna rahva hulgas.

- Euroopa geneetiline panus Ameerika populatsioonidesse peegeldab kolonisatsiooni aegset geopoliitilist olukorda. Avastati mitu sekundaarset Euroopa allikat, mis panustasid arvestatavasse osassse Ameerika populatsioonidest, nt Itaalia Brasiilias ja Argentiinas, Kesk-Euroopa Brasiilias. Tuletati Aafrika allikate eristuv panus Ameerika populatsioonidesse.
- Segunemise ajad langevad kokku rändelainetega Euroopast ja peegeldavad ekspluateeritud Aafrika piirkondade muutumist ajas.
- Segunemise demograafilise mõju analüüsist selgub üldine languse ja taastumise muster mitmes uuritavas populatsioonis, mis vastab koloniaalajastu algusele ja lõpule. Kuid Peruud ja Mehhikot iseloomustavad erinevad demograafilised trajektoorid.
- Soost sõltuva segunemise dünaamika analüüs viitab sellele, et tänapäeva populatsioonidesse on panustanud rohkem Ameerika naisi kui mehi. Vastupidiselt oli Euroopa meeste panus olulisem kui samalt mandrilt pärinevate naiste oma. Sellele vastandlikult ilmnes mõnes populatsioonis, kuid mitte kõigis, tõendeid suuremast naiste panusest, mis on osaliselt vastuolus ajalooliste andmetega Aafrika päritolust.

REFERENCES

1000 Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL, McCarthy S, McVean GA, et al. 2015. A global reference for human genetic variation. *Nature* 526:68–74.

Achilli A, Perego UA, Bravi CM, Coble MD, Kong Q-P, Woodward SR, Salas A, Torroni A, Bandelt H-J. 2008. The phylogeny of the four pan-American MtDNA haplogroups: implications for evolutionary and disease studies. *PLoS One* 3:e1764.

Achilli A, Perego UA, Lancioni H, Olivieri A, Gandini F, Hooshiar Kashani B, Battaglia V, Grugni V, Angerhofer N, Rogers MP, et al. 2013. Reconciling migration models to the Americas with the variation of North American native mitogenomes. *Proc. Natl. Acad. Sci. U. S. A.* 110:14308–14313.

Acosta-Ochoa G, Pérez-Martínex P, Ulloa-Montemayor X. 2019. The Clovis-like and Fishtail occupations of southern Mexico and Central America. *People & Culture in Ice Age Americas*:93–107.

Adhikari K, Chacón-Duque JC, Mendoza-Revilla J, Fuentes-Guajardo M, Ruiz-Linares A. 2017. The Genetic Diversity of the Americas. *Annu. Rev. Genomics Hum. Genet.* 18:277–296.

Adhikari K, Mendoza-Revilla J, Chacón-Duque JC, Fuentes-Guajardo M, Ruiz-Linares A. 2016. Admixture in Latin America. *Curr. Opin. Genet. Dev.* 41:106–114.

Alexander DH, Novembre J, Lange K. 2009. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 19:1655–1664.

Alves-Silva J, da Silva Santos M, Guimarães PE, Ferreira AC, Bandelt HJ, Pena SD, Prado VF. 2000. The ancestry of Brazilian mtDNA lineages. *Am. J. Hum. Genet.* 67:444–461.

Amarasinghe SL, Su S, Dong X, Zappia L, Ritchie ME, Gouil Q. 2020. Opportunities and challenges in long-read sequencing data analysis. *Genome Biol.* 21:30.

Anda LAAD, De anda LAA, Maldonado-Koerdell M. 1953. Association of Artifacts with Mammoth in the Valley of Mexico. *American Antiquity*. 18:332–340.

Anderson RP, Handley CO Jr. 2002. Dwarfism in insular sloths: biogeography, selection, and evolutionary rate. *Evolution* 56:1045–1058.

Ardelean CF, Becerra-Valdivia L, Pedersen MW, Schwenninger J-L, Oviatt CG, Macías-Quintero JI, Arroyo-Cabrales J, Sikora M, Ocampo-Díaz YZE, Rubio-Cisneros II, et al. 2020. Evidence of human occupation in Mexico around the Last Glacial Maximum. *Nature* 584:87–92.

Baily SL, Míguez EJ. 2003. Mass Migration to Modern Latin America. Rowman & Littlefield.

Balanovsky O, Gurianov V, Zaporozhchenko V, Balaganskaya O, Urasin V, Zhabagin M, Grugni V, Canada R, Al-Zahery N, Raveane A, et al. 2017. Phylogeography of human Y-chromosome haplogroup Q3-L275 from an academic/citizen science collaboration. *BMC Evol. Biol.* 17:18.

Bandelt HJ, Forster P, Röhl A. 1999. Median-joining networks for inferring intraspecific phylogenies. *Mol. Biol. Evol.* 16:37–48.

Baran Y, Pasaniuc B, Sankararaman S, Torgerson DG, Gignoux C, Eng C, Rodriguez-Cintron W, Chapela R, Ford JG, Avila PC, et al. 2012. Fast and accurate inference of local ancestry in Latino populations. *Bioinformatics* 28:1359–1367.

Barbieri C, Hübner A, Macholdt E, Ni S, Lippold S, Schröder R, Mpoloka SW, Purps J, Roewer L, Stoneking M, et al. 2016. Refining the Y chromosome phylogeny with southern African sequences. *Hum. Genet.* 135:541–553.

Barrantes R, Smouse PE, Mohrenweiser HW, Gershowitz H, Azofeifa J, Arias TD, Neel JV. 1990. Microevolution in lower Central America: genetic characterization of the Chibcha-speaking groups of Costa Rica and Panama, and a consensus taxonomy based on genetic and linguistic affinity. *Am. J. Hum. Genet.* 46:63–84.

Barry B. 1998. Senegambia and the Atlantic Slave Trade. Cambridge University Press

Batini C, Ferri G, Destro-Bisol G, Brisighelli F, Luiselli D, Sánchez-Diz P, Rocha J, Simonson T, Brehm A, Montano V, et al. 2011. Signatures of the preagricultural peopling processes in sub-Saharan Africa as revealed by the phylogeography of early Y chromosome lineages. *Mol. Biol. Evol.* 28:2603–2613.

Battaglia V, Fornarino S, Al-Zahery N, Olivieri A, Pala M, Myres NM, King RJ, Rootsi S, Marjanovic D, Primorac D, et al. 2009. Y-chromosomal evidence of the cultural diffusion of agriculture in Southeast Europe. *Eur. J. Hum. Genet.* 17:820–830.

Battaglia V, Grugni V, Perego UA, Angerhofer N, Gomez-Palmieri JE, Woodward SR, Achilli A, Myres N, Torroni A, Semino O. 2013. The first peopling of South America: new evidence from Y-chromosome haplogroup Q. *PLoS One* 8:e71390.

Bayor RH. 2016. The Oxford Handbook of American Immigration and Ethnicity. Oxford University Press

Becerra-Valdivia L, Higham T. 2020. The timing and effect of the earliest human arrivals in North America. *Nature* 584:93–97.

Beck C, Jones GT. 2010. Clovis and Western Stemmed: Population Migration and the Meeting of Two Technologies in the Intermountain West. *Am. Antiq.* 75:81–116.

Behar DM, van Oven M, Rosset S, Metspalu M, Loogväli E-L, Silva NM, Kivisild T, Torroni A, Villems R. 2012. A "Copernican" Reassessment of the Human Mitochondrial DNA Tree from its Root. *Am. J. Hum. Genet.* 90:675–684.

Bethell L. 1984. The Cambridge History of Latin America. Cambridge University Press

Bonaffini L, Perricone J. 2014. Poets of the Italian Diaspora: A Bilingual Anthology. Fordham University Press

Bonch-Bruevich X. 2008. Ideologies of the Spanish Reconquest and Isidore's Political Thought. *Mediterr. Stud.* 17:27–45.

Bortolini M-C, Salzano FM, Thomas MG, Stuart S, Nasanen SPK, Bau CHD, Hutz MH, Layrisse Z, Petzl-Erler ML, Tsuneto LT, et al. 2003. Y-chromosome evidence for differing ancient demographic histories in the Americas. *Am. J. Hum. Genet.* 73:524–539.

Bosch E, Calafell F, Rosser ZH, Nørby S, Lynnerup N, Hurles ME, Jobling MA. 2003. High level of male-biased Scandinavian admixture in Greenlandic Inuit shown by Y-chromosomal analysis. *Hum. Genet.* 112:353–363.

Boyd-Bowman P. 1976. Patterns of Spanish Emigration to the Indies until 1600. *Hisp. Am. Hist. Rev.* 56:580.

Brandini S, Bergamaschi P, Cerna MF, Gandini F, Bastaroli F, Bertolini E, Cereda C, Ferretti L, Gómez-Carballa A, Battaglia V, et al. 2018. The Paleo-Indian Entry into South America According to Mitogenomes. *Mol. Biol. Evol.* 35:299–311.

Brisbin A, Bryc K, Byrnes J, Zakharia F, Omberg L, Degenhardt J, Reynolds A, Ostrer H, Mezey JG, Bustamante CD. 2012. PCAdmix: principal components-based assignment of ancestry along each chromosome in individuals with admixed ancestry from two or more populations. *Hum. Biol.* 84:343–364.

Browning SR, Browning BL. 2007. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am. J. Hum. Genet.* 81:1084–1097.

Browning SR, Browning BL, Daviglus ML, Durazo-Arvizu RA, Schneiderman N, Kaplan RC, Laurie CC. 2018. Ancestry-specific recent effective population size in the Americas. *PLoS Genet.* 14:e1007385.

Browning SR, Grinde K, Plantinga A, Gogarten SM, Stilp AM, Kaplan RC, Larissa Avilés-Santa M, Browning BL, Laurie CC. 2016. Local Ancestry Inference in a Large US-Based Hispanic/Latino Study: Hispanic Community Health Study/Study of Latinos (HCHS/SOL). *G3: Genes|Genomes|Genetics*. 6:1525–1534.

Brown JC. 2010. A Brief History of Argentina. Facts On File

Bryc K, Auton A, Nelson MR, Oksenberg JR, Hauser SL, Williams S, Froment A, Bodo J-M, Wambebe C, Tishkoff SA, et al. 2010. Genome-wide patterns of population structure and admixture in West Africans and African Americans. *Proc. Natl. Acad. Sci. U. S. A.* 107:786–791.

Bryc K, Durand EY, Macpherson JM, Reich D, Mountain JL. 2015. The genetic ancestry of African Americans, Latinos, and European Americans across the United States. *Am. J. Hum. Genet.* 96:37–53.

Capodiferro MR, Aram B, Raveane A, Rambaldi Migliore N, Colombo G, Ongaro L, Rivera J, Mendizábal T, Hernández-Mora I, Tribaldos M, et al. 2021. Archaeogenomic distinctiveness of the Isthmo-Colombian area. *Cell* 184:1706–1723.e24.

Carvajal-Carmona LG, Ophoff R, Service S, Hartiala J, Molina J, Leon P, Ospina J, Bedoya G, Freimer N, Ruiz-Linares A. 2003. Genetic demography of Antioquia (Colombia) and the Central Valley of Costa Rica. *Hum. Genet.* 112:534–541.

Carvajal-Carmona LG, Soto ID, Pineda N, Ortíz-Barrientos D, Duque C, Ospina-Duque J, McCarthy M, Montoya P, Alvarez VM, Bedoya G, et al. 2000. Strong Amerind/white sex bias and a possible Sephardic contribution among the founders of a population in northwest Colombia. *Am. J. Hum. Genet.* 67:1287–1295.

Cavalli-Sforza LL, Barrai I, Edwards AW. 1964. Analysis of Human Evolution Under Random Genetic Drift. *Cold Spring Harb. Symp. Quant. Biol.* 29:9–20.

Chacón-Duque J-C, Adhikari K, Fuentes-Guajardo M, Mendoza-Revilla J, Acuña-Alonzo V, Barquera R, Quinto-Sánchez M, Gómez-Valdés J, Everardo Martínez P, Villamil-Ramírez H, et al. 2018. Latin Americans show wide-spread Converso ancestry and imprint of local Native ancestry on physical appearance. *Nat. Commun.* 9:5388.

Chiaroni J, Underhill PA, Cavalli-Sforza LL. 2009. Y chromosome diversity, human expansion, drift, and cultural evolution. *Proc. Natl. Acad. Sci. U. S. A.* 106:20174–20179.

Conley AB, Rishishwar L, Norris ET, Valderrama-Aguirre A, Mariño-Ramírez L, Medina-Rivas MA, Jordan IK. 2017. A comparative analysis of genetic ancestry and admixture in the Colombian populations of Choco and Medellin. *G3: Genes, Genomes, Genetics* 7:3435–3447.

Cooke R. 2005. Prehistory of Native Americans on the Central American Land Bridge: Colonization, Dispersal, and Divergence. *Journal of Archaeological Research* 13:129–187.

Cooke R, Ranere A, Pearson G, Dickau R. 2013. Radiocarbon chronology of early human settlement on the Isthmus of Panama (13,000–7000 BP) with comments on cultural affinities, environments, subsistence, and technological change. *Quaternary International.* 301:3–22.

Cropp S, Boinski S. 2000. The Central American squirrel monkey (Saimiri oerstedii): introduced hybrid or endemic species? *Mol. Phylogenet. Evol.* 16:350–365.

Cruciani F, Trombetta B, Massaia A, Destro-Bisol G, Sellitto D, Scozzari R. 2011. A revised root for the human Y chromosomal phylogenetic tree: the origin of patrilineal diversity in Africa. *Am. J. Hum. Genet.* 88:814–818.

Curtin PD. 1972. The Atlantic Slave Trade: A Census. Univ of Wisconsin Press

Davis LG, Nyers AJ, Willis SC. 2014. Context, Provenance and Technology of a Western Stemmed Tradition Artifact Cache from the Cooper's Ferry Site, Idaho. *Am. Antiq.* 79:596–615.

Delaneau O, Zagury J-F, Marchini J. 2013. Improved whole-chromosome phasing for disease and population genetic studies. *Nat. Methods* 10:5–6.

Delaneau O, Zagury J-F, Robinson MR, Marchini JL, Dermitzakis ET. 2019. Accurate, scalable and integrative haplotype estimation. *Nature Communications*. 10.

Denevan WM. 1992. The Native Population of the Americas in 1492. Univ of Wisconsin Press

Deng W, Shi B, He X, Zhang Z, Xu J, Li B, Yang J, Ling L, Dai C, Qiang B, et al. 2004. Evolution and migration history of the Chinese population inferred from Chinese Y-chromosome evidence. *J. Hum. Genet.* 49:339–348.

Di Cristofaro J, Pennarun E, Mazières S, Myres NM, Lin AA, Temori SA, Metspalu M, Metspalu E, Witzel M, King RJ, et al. 2013. Afghan Hindu Kush: where Eurasian sub-continent gene flows converge. *PLoS One* 8:e76748.

Di Gaetano C, Cerutti N, Crobu F, Robino C, Inturri S, Gino S, Guarrera S, Underhill PA, King RJ, Romano V, et al. 2009. Differential Greek and northern African migrations to Sicily are supported by genetic evidence from the Y chromosome. *Eur. J. Hum. Genet.* 17:91–99.

Dillehay TD, Goodbred S, Pino M, Vásquez Sánchez VF, Tham TR, Adovasio J, Collins MB, Netherly PJ, Hastorf CA, Chiou KL, et al. 2017. Simple technologies and diverse food strategies of the Late Pleistocene and Early Holocene at Huaca Prieta, Coastal Peru. *Sci Adv* 3:e1602778.

Driver JC, Handly M, Fladmark KR, Nelson DE, Sullivan GM, Preston R. 1996. Stratigraphy, Radiocarbon Dating, and Culture History of Charlie Lake Cave, British Columbia. *Arctic* 49:265–277.

Drummond AJ, Suchard MA, Xie D. 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol. Biol*.

Dulik MC, Owings AC, Gaieski JB, Vilar MG, Andre A, Lennie C, Mackenzie MA, Kritsch I, Snowshoe S, Wright R, et al. 2012. Y-chromosome analysis reveals genetic divergence and new founding native lineages in Athapaskan- and Eskimoan-speaking populations. *Proc. Natl. Acad. Sci. U. S. A.* 109:8471–8476.

Dulik MC, Zhadanov SI, Osipova LP, Askapuli A, Gau L, Gokcumen O, Rubinstein S, Schurr TG. 2012. Mitochondrial DNA and Y chromosome variation provides evidence for a recent common ancestry between Native Americans and Indigenous Altaians. *Am. J. Hum. Genet.* 90:229–246.

Edwards E. 2014. Slavery in Argentina. *Oxford Bibliographies Online Datasets*.

Elliott JH. 1984. The Spanish conquest and settlement of America. *The Cambridge History of Latin America* 1:149–206.

Eltis D, Richardson D. 2010. Atlas of the transatlantic slave trade. The Lewis Walpole series in eighteenth-century culture and history. *New Haven [ua]*.

Fagundes NJR, Kanitz R, Eckert R, Valls ACS, Bogo MR, Salzano FM, Smith DG, Silva WA Jr, Zago MA, Ribeiro-dos-Santos AK, et al. 2008. Mitochondrial population genomics supports a single pre-Clovis origin with a coastal route for the peopling of the Americas. *Am. J. Hum. Genet.* 82:583–592.

Fernandes DM, Sirak KA, Ringbauer H, Sedig J, Rohland N, Cheronet O, Mah M, Mallick S, Olalde I, Culleton BJ, et al. 2021. A genetic history of the pre-contact Caribbean. *Nature* 590:103–110.

Flegontov P, Altınışık NE, Changmai P, Rohland N, Mallick S, Adamski N, Bolnick DA, Broomandkhoshbacht N, Candilio F, Culleton BJ, et al. 2019. Palaeo-Eskimo genetic ancestry and the peopling of Chukotka and North America. *Nature* 570:236–240.

Fornarino S, Pala M, Battaglia V, Maranta R, Achilli A, Modiano G, Torroni A, Semino O, Santachiara-Benerecetti SA. 2009. Mitochondrial and Y-chromosome diversity of the Tharus (Nepal): a reservoir of genetic variation. *BMC Evol. Biol.* 9:154.

Fortes-Lima C, Gessain A, Ruiz-Linares A, Bortolini M-C, Migot-Nabias F, Bellis G, Moreno-Mayar JV, Restrepo BN, Rojas W, Avendaño-Tamayo E, et al. 2017. Genome-wide Ancestry and Demographic History of African-Descendant Maroon Communities from French Guiana and Suriname. *Am. J. Hum. Genet.* 101:725–736.

Fortes-Lima C, Verdu P. 2021. Anthropological genetics perspectives on the transatlantic slave trade. *Hum. Mol. Genet.* 30:R79–R87.

Francalacci P, Morelli L, Angius A, Berutti R, Reinier F, Atzeni R, Pilu R, Busonero F, Maschio A, Zara I, et al. 2013. Low-pass DNA sequencing of 1200 Sardinians reconstructs European Y-chromosome phylogeny. *Science* 341:565–569.

de la Fuente C, Ávila-Arcos MC, Galimany J, Carpenter ML, Homburger JR, Blanco A, Contreras P, Cruz Dávalos D, Reyes O, San Roman M, et al. 2018. Genomic insights into the origin and diversification of late maritime hunter-gatherers from the Chilean Patagonia. *Proc. Natl. Acad. Sci. U. S. A.* 115:E4006–E4012.

Fu Q, Hajdinjak M, Moldovan OT, Constantin S, Mallick S, Skoglund P, Patterson N, Rohland N, Lazaridis I, Nickel B, et al. 2015. An early modern human from Romania with a recent Neanderthal ancestor. *Nature* 524:216–219.

Fu Q, Meyer M, Gao X, Stenzel U, Burbano HA, Kelso J, Pääbo S. 2013. DNA analysis of an early modern human from Tianyuan Cave, China. *Proc. Natl. Acad. Sci. U. S. A.* 110:2223–2227.

Gansauge M-T, Meyer M. 2013. Single-stranded DNA library preparation for the sequencing of ancient or damaged DNA. *Nat. Protoc.* 8:737–748.

Gardner LI Jr, Stern MP, Haffner SM, Gaskill SP, Hazuda HP, Relethford JH, Eifler CW. 1984. Prevalence of diabetes in Mexican Americans. Relationship to percent of gene pool derived from native American sources. *Diabetes* 33:86–92.

Geppert M, Ayub Q, Xue Y, Santos S, Ribeiro-dos-Santos Â, Baeta M, Núñez C, Martínez-Jarreta B, Tyler-Smith C, Roewer L. 2015. Identification of new SNPs in native South American populations by resequencing the Y chromosome. *Forensic Sci. Int. Genet.* 15:111–114.

Giles RE, Blanc H, Cann HM, Wallace DC. 1980. Maternal inheritance of human mitochondrial DNA. *Proc. Natl. Acad. Sci. U. S. A.* 77:6715–6719.

Gittleman JL. 2014. Phylogeny. *Encyclopædia Britannica Online*.

Giunta E, Sciorra J. 2014. Embroidered Stories: Interpreting Women's Domestic Needlework from the Italian Diaspora. Univ. Press of Mississippi

Goebel T, Keene JL. 2014. Are Great Basin stemmed points as old as Clovis in the Intermountain West? *Archaeology in the Great Basin and Southwest*

Goebel T, Waters MR, O'Rourke DH. 2008. The late Pleistocene dispersal of modern humans in the Americas. *Science* 319:1497–1502.

Goetz LH, Uribe-Bruce L, Quarless D, Libiger O, Schork NJ. 2014. Admixture and clinical phenotypic variation. *Hum. Hered.* 77:73–86.

Goldberg A, Mychajliw AM, Hadly EA. 2016. Post-invasion demography of prehistoric humans in South America. Nature 532:232–235.
Goldberg A, Rosenberg NA. 2015. Beyond 2/3 and 1/3: The Complex Signatures of Sex-Biased Admixture on the X Chromosome. *Genetics* 201:263–279.
Gouveia MH, Borda V, Leal TP, Moreira RG, Bergen AW, Aquino MM, Araujo GS, Araujo NM, Kehdy FSG, Liboredo R, et al. 2020. Origins, admixture dynamics and homogenization of the African gene pool in the Americas. *Molecular Biology and Evolution.* 37(6):1647–1656.
Greenberg JH, Turner CG, Zegura SL, Campbell L, Fox JA, Laughlin WS, Weiss KM, Woolford E. 1986. The Settlement of the Americas: A Comparison of the Linguistic, Dental, and Genetic Evidence [and Comments and Reply]. *Curr. Anthropol.* 27:477–497.
Grugni V, Battaglia V, Hooshiar Kashani B, Parolo S, Al-Zahery N, Others. 2012. Ancient Migratory Events in the Middle East: New Clues from the Y. 7(7): e41252 Grugni V, Battaglia V, Perego UA, Raveane A, Lancioni H, Olivieri A, Ferretti L, Woodward SR, Pascale JM, Cooke R, et al. 2015. Exploring the Y Chromosomal Ancestry of Modern Panamanians. *PLoS One* 10:e0144223.
Grugni V, Raveane A, Ongaro L, Battaglia V, Trombetta B, Colombo G, Capodiferro MR, Olivieri A, Achilli A, Perego UA, et al. 2019. Analysis of the human Y-chromosome haplogroup Q characterizes ancient population movements in Eurasia and the Americas. *BMC Biol.* 17:3.
Guan Y. 2014. Detecting structure of haplotypes and local ancestry. *Genetics* 196:625–642.
Hallast P, Batini C, Zadik D, Maisano Delser P, Wetton JH, Arroyo-Pardo E, Cavalleri GL, de Knijff P, Destro Bisol G, Dupuy BM, et al. 2015. The Y-chromosome tree bursts into leaf: 13,000 high-confidence SNPs covering the majority of known clades. *Mol. Biol. Evol.* 32:661–673.
Hammer MF, Karafet T, Rasanayagam A, Wood ET, Altheide TK, Jenkins T, Griffiths RC, Templeton AR, Zegura SL. 1998. Out of Africa and back again: nested cladistic analysis of human Y chromosome variation. *Mol. Biol. Evol.* 15:427–441.
Han E, Carbonetto P, Curtis RE, Wang Y, Granka JM, Byrnes J, Noto K, Kermany AR, Myres NM, Barber MJ, et al. 2017. Clustering of 770,000 genomes reveals post-colonial population structure of North America. *Nat. Commun.* 8:14238.
Harris DN, Song W, Shetty AC, Levano KS, Cáceres O, Padilla C, Borda V, Tarazona D, Trujillo O, Sanchez C, et al. 2018. Evolutionary genomic dynamics of Peruvians before, during, and after the Inca Empire. *Proc. Natl. Acad. Sci. U. S. A.* 115:E6526–E6535.
Hellenthal G, Busby GBJ, Band G, Wilson JF, Capelli C, Falush D, Myers S. 2014. A genetic atlas of human admixture history. *Science* 343:747–751.
Herlihy PH. 2003. Participatory research mapping of indigenous lands in Darién, Panama. *Hum. Organ.* 62:315–331.
Homburger JR, Moreno-Estrada A, Gignoux CR, Nelson D, Sanchez E, Ortiz-Tello P, Pons-Estel BA, Acevedo-Vasquez E, Miranda P, Langefeld CD, et al. 2015. Genomic Insights into the Ancestry and Demographic History of South America. *PLoS Genet.* 11:e1005602.
Ho SYW, Shapiro B. 2011. Skyline-plot methods for estimating demographic history from nucleotide sequences. *Mol. Ecol. Resour.* 11:423–434.

Howie B, Fuchsberger C, Stephens M, Marchini J, Abecasis GR. 2012. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat. Genet.* 44:955–959.

Howie BN, Donnelly P, Marchini J. 2009. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet.* 5:e1000529.

Hughes JF, Page DC. 2015. The Biology and Evolution of Mammalian Y Chromosomes. *Annu. Rev. Genet.* 49:507–527.

Jackson D, Méndez C, Seguel R, Maldonado A, Vargas G. 2007. Initial Occupation of the Pacific Coast of Chile during Late Pleistocene Times. *Curr. Anthropol.* 48:725–731.

Jain M, Koren S, Miga KH, Quick J, Rand AC, Sasani TA, Tyson JR, Beggs AD, Dilthey AT, Fiddes IT, et al. 2018. Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nat. Biotechnol.* 36:338–345.

Jenkins DL, Davis LG, Stafford TW Jr, Campos PF, Hockett B, Jones GT, Cummings LS, Yost C, Connolly TJ, Yohe RM 2nd, et al. 2012. Clovis age Western Stemmed projectile points and human coprolites at the Paisley Caves. *Science* 337:223–228.

Jobling MA, Pandya A, Tyler-Smith C. 1997. The Y chromosome in forensic analysis and paternity testing. *Int. J. Legal Med.* 110:118–124.

Jobling MA, Tyler-Smith C. 2003. The human Y chromosome: an evolutionary marker comes of age. *Nat. Rev. Genet.* 4:598–612.

Jobling MA, Tyler-Smith C. 2017. Human Y-chromosome variation in the genome-sequencing era. *Nat. Rev. Genet.* 18:485–497.

Jobling M, Hurles M, Tyler-Smith C. 2013. Human Evolutionary Genetics: Origins, Peoples & Disease. Garland Science

Jota MS, Lacerda DR, Sandoval JR, Vieira PPR, Ohasi D, Santos-Júnior JE, Acosta O, Cuellar C, Revollo S, Paz-Y-Miño C, et al. 2016. New native South American Y chromosome lineages. *J. Hum. Genet.* 61:593–603.

Joyce DJ. 2013. Pre-Clovis megafauna butchery sites in the western Great Lakes region, USA. *Paleoamerican Odyssey*.

Kamen H. 2003. Spain's road to empire: The making of a world power, 1492–1763. Penguin UK

Karafet TM, Mendez FL, Meilerman MB, Underhill PA, Zegura SL, Hammer MF. 2008. New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. *Genome Res.* 18:830–838.

Karafet TM, Zegura SL, Posukh O, Osipova L, Bergen A, Long J, Goldman D, Klitz W, Harihara S, de Knijff P, et al. 1999. Ancestral Asian Source(s) of New World Y-Chromosome Founder Haplotypes. *The American Journal of Human Genetics*. 64:817–831.

Karafet T, Zegura SL, Vuturo-Brady J, Posukh O, Osipova L, Wiebe V, Romero F, Long JC, Harihara S, Jin F, et al. 1997. Y chromosome markers and trans-Bering Strait dispersals. *American Journal of Physical Anthropology*. 102:301–314.

Karmin M, Saag L, Vicente M, Wilson Sayres MA, Järve M, Talas UG, Rootsi S, Ilumäe A-M, Mägi R, Mitt M, et al. 2015. A recent bottleneck of Y chromosome diversity coincides with a global change in culture. *Genome Res.* 25:459–466.

Kehdy FSG, Gouveia MH, Machado M, Magalhães WCS, Horimoto AR, Horta BL, Moreira RG, Leal TP, Scliar MO, Soares-Souza GB, et al. 2015. Origin and dynamics of admixture in Brazilians and its effect on the pattern of deleterious mutations. *Proc. Natl. Acad. Sci. U. S. A.* 112:8696–8701.

Kemp BM, Malhi RS, McDonough J, Bolnick DA, Eshleman JA, Rickards O, Martinez-Labarga C, Johnson JR, Lorenz JG, Dixon EJ, et al. 2007. Genetic analysis of early holocene skeletal remains from Alaska and its implications for the settlement of the Americas. *Am. J. Phys. Anthropol.* 132:605–621.

Kivisild T. 2017. The study of human Y chromosome variation through ancient DNA. *Hum. Genet.* 136:529–546.

Klein HS. 1999. The Atlantic Slave Trade. Cambridge University Press

Knapp M, Clarke AC, Horsburgh KA, Matisoo-Smith EA. 2012. Setting the stage – building and working in an ancient DNA laboratory. *Ann. Anat.* 194:3–6.

Korunes KL, Goldberg A. 2021. Human genetic admixture. *PLoS Genet.* 17:e1009374.

Kumar S, Bellis C, Zlojutro M, Melton PE, Blangero J, Curran JE. 2011. Large scale mitochondrial sequencing in Mexican Americans suggests a reappraisal of Native American origins. *BMC Evol. Biol.* 11:293.

Lavrin A. 1989. Sexuality and Marriage in Colonial Latin America. U of Nebraska Press

Lawson DJ, van Dorp L, Falush D. 2018. A tutorial on how not to over-interpret STRUCTURE and ADMIXTURE bar plots. *Nat. Commun.* 9:3258.

Lawson DJ, Hellenthal G, Myers S, Falush D. 2012. Inference of population structure using dense haplotype data. *PLoS Genet.* 8:e1002453.

Leslie S, Winney B, Hellenthal G, Davison D, Boumertit A, Day T, Hutnik K, Royrvik EC, Cunliffe B, Wellcome Trust Case Control Consortium 2, et al. 2015. The fine-scale genetic structure of the British population. *Nature* 519:309–314.

Lesnek AJ, Briner JP, Lindqvist C, Baichtal JF, Heaton TH. 2018. Deglaciation of the Pacific coastal corridor directly preceded the human colonization of the Americas. *Sci Adv* 4:eaar5040.

Lindo J, Achilli A, Perego UA, Archer D, Valdiosera C, Petzelt B, Mitchell J, Worl R, Dixon EJ, Fifield TE, et al. 2017. Ancient individuals from the North American Northwest Coast reveal 10,000 years of regional genetic continuity. *Proc. Natl. Acad. Sci. U. S. A.* 114:4093–4098.

Li N, Stephens M. 2003. Modeling linkage disequilibrium and identifying recombination hotspots using single-nucleotide polymorphism data. *Genetics* 165:2213–2233.

Lippold S, Xu H, Ko A, Li M, Renaud G, Butthof A, Schröder R, Stoneking M. 2014. Human paternal and maternal demographic histories: insights from high-resolution Y chromosome and mtDNA sequences. *Investig. Genet.* 5:13.

Liu JJ, Norman RMG, Manchanda R, De Luca V. 2013. Admixture analysis of age at onset in schizophrenia: evidence of three subgroups in a first-episode sample. *Gen. Hosp. Psychiatry* 35:664–667.

Loh P-R, Danecek P, Palamara PF, Fuchsberger C, A Reshef Y, K Finucane H, Schoenherr S, Forer L, McCarthy S, Abecasis GR, et al. 2016. Reference-based phasing using the Haplotype Reference Consortium panel. *Nat. Genet.* 48:1443–1448.

Loh P-R, Lipson M, Patterson N, Moorjani P, Pickrell JK, Reich D, Berger B. 2013. Inferring admixture histories of human populations using linkage disequilibrium. *Genetics* 193:1233–1254.

Lothrop SK. 1942. Cocle: An Archaeological Study of Central Panama. Pottery of the Sitia Conte..

Mallick S, Li H, Lipson M, Mathieson I, Gymrek M, Racimo F, Zhao M, Chennagiri N, Nordenfelt S, Tandon A, et al. 2016. The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. *Nature* 538:201–206.

Maples BK, Gravel S, Kenny EE, Bustamante CD. 2013. RFMix: a discriminative modeling approach for rapid and robust local-ancestry inference. *Am. J. Hum. Genet.* 93:278–288.

Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen Y-J, Chen Z, et al. 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437:376–380.

Maricic T, Whitten M, Pääbo S. 2010. Multiplexed DNA sequence capture of mitochondrial genomes using PCR products. *PLoS One* 5:e14004.

Marjanovic D, Fornarino S, Montagna S, Primorac D, Hadziselimovic R, Vidovic S, Pojskic N, Battaglia V, Achilli A, Drobnic K, et al. 2005. The peopling of modern Bosnia-Herzegovina: Y-chromosome haplogroups in the three main ethnic groups. *Ann. Hum. Genet.* 69:757–763.

Marnetto D, Pärna K, Läll K, Molinaro L, Montinaro F, Haller T, Metspalu M, Mägi R, Fischer K, Pagani L. 2020. Ancestry deconvolution and partial polygenic score can improve susceptibility predictions in recently admixed individuals. *Nat. Commun.* 11:1628.

Martín JG, Cooke RG, Bustamante F, Holst I, Lara A, Redwood S. 2016. Ocupaciones prehispánicas en Isla Pedro González, Archipielago de Las Perlas, Panamá: Aproximación a una cronología con comentarios sobre las conexiones externas. *Latin American Antiquity*:378–396.

Mason JA, Johnson F. 1940. The native languages of Middle America.

Meade TA. 2011. A History of Modern Latin America: 1800 to the Present. John Wiley & Sons

Mendez FL, Krahn T, Schrack B, Krahn A-M, Veeramah KR, Woerner AE, Fomine FLM, Bradman N, Thomas MG, Karafet TM, et al. 2013. An African American paternal lineage adds an extremely ancient root to the human Y chromosome phylogenetic tree. *Am. J. Hum. Genet.* 92:454–459.

Micheletti SJ, Bryc K, Ancona Esselmann SG, Freyman WA, Moreno ME, Poznik GD, Shastri AJ, 23andMe Research Team, Beleza S, Mountain JL. 2020. Genetic Consequences of the Transatlantic Slave Trade in the Americas. *Am. J. Hum. Genet.* 107:265–277.

Montinaro F, Busby GBJ, Pascali VL, Myers S, Hellenthal G, Capelli C. 2015. Unravelling the hidden ancestry of American admixed populations. *Nat. Commun.* 6:6596.

Moorjani P, Patterson N, Hirschhorn JN, Keinan A, Hao L, Atzmon G, Burns E, Ostrer H, Price AL, Reich D. 2011. The history of African gene flow into Southern Europeans, Levantines, and Jews. *PLoS Genet.* 7:e1001373.

Moreno-Estrada A, Gignoux CR, Fernández-López JC, Zakharia F, Sikora M, Contreras AV, Acuña-Alonzo V, Sandoval K, Eng C, Romero-Hidalgo S, et al. 2014. Human genetics. The genetics of Mexico recapitulates Native American substructure and affects biomedical traits. *Science* 344:1280–1285.

Moreno-Estrada A, Gravel S, Zakharia F, McCauley JL, Byrnes JK, Gignoux CR, Ortiz-Tello PA, Martínez RJ, Hedges DJ, Morris RW, et al. 2013. Reconstructing the population genetic history of the Caribbean. *PLoS Genet.* 9:e1003925.

Moreno-Mayar JV, Vinner L, de Barros Damgaard P, de la Fuente C, Chan J, Spence JP, Allentoft ME, Vimala T, Racimo F, Pinotti T, et al. 2018. Early human dispersals within the Americas. *Science*. 362.

Mörner M. 1967. Race mixture in the history of Latin America.

Nielsen R, Akey JM, Jakobsson M, Pritchard JK, Tishkoff S, Willerslev E. 2017. Tracing the peopling of the world through genomics. *Nature* 541:302–310.
Novembre J, Peter BM. 2016. Recent advances in the study of fine-scale population structure in humans. *Curr. Opin. Genet. Dev.* 41:98–105.
Ongaro L, Scliar MO, Flores R, Raveane A, Marnetto D, Sarno S, Gnecchi-Ruscone GA, Alarcón-Riquelme ME, Patin E, Wangkumhang P, et al. 2019. The Genomic Impact of European Colonization of the Americas. *Curr. Biol.* 29:3974–3986.e4.
Orlando L, Gilbert MTP, Willerslev E. 2015. Reconstructing ancient genomes and epigenomes. *Nat. Rev. Genet.* 16:395–408.
O'Rourke DH, Raff JA. 2010. The human genetic history of the Americas: the final frontier. *Curr. Biol.* 20:R202–R207.
van Oven M, Van Geystelen A, Kayser M, Decorte R, Larmuseau MHD. 2014. Seeing the wood for the trees: a minimal reference phylogeny for the human Y chromosome. *Hum. Mutat.* 35:187–191.
Overstreet DF. 2005. Late-glacial ice-marginal adaptation in southeastern Wisconsin. *Paleoamerican Origins: Beyond Clovis*:183–195.
Patterson N, Moorjani P, Luo Y, Mallick S, Rohland N, Zhan Y, Genschoreck T, Webster T, Reich D. 2012. Ancient admixture in human history. *Genetics* 192:1065–1093.
Pearson GA. 2017. Bridging the Gap: An Updated Overview of Clovis across Middle America and its Techno-Cultural Relation with Fluted Point Assemblages from South America. *PaleoAmerica* 3:203–230.
Perego UA, Achilli A, Angerhofer N, Accetturo M, Pala M, Olivieri A, Hooshiar Kashani B, Ritchie KH, Scozzari R, Kong Q-P, et al. 2009. Distinctive Paleo-Indian migration routes from Beringia marked by two rare mtDNA haplogroups. *Curr. Biol.* 19:1–8.
Perego UA, Angerhofer N, Pala M, Olivieri A, Lancioni H, Hooshiar Kashani B, Carossa V, Ekins JE, Gómez-Carballa A, Huber G, et al. 2010. The initial peopling of the Americas: a growing number of founding mitochondrial genomes from Beringia. *Genome Res.* 20:1174–1179.
Perego UA, Lancioni H, Tribaldos M, Angerhofer N, Ekins JE, Olivieri A, Woodward SR, Pascale JM, Cooke R, Motta J, et al. 2012. Decrypting the mitochondrial gene pool of modern Panamanians. *PLoS One* 7:e38337.
Peterson RE, Kuchenbaecker K, Walters RK, Chen C-Y, Popejoy AB, Periyasamy S, Lam M, Iyegbe C, Strawbridge RJ, Brick L, et al. 2019. Genome-wide Association Studies in Ancestrally Diverse Populations: Opportunities, Methods, Pitfalls, and Recommendations. *Cell* 179:589–603.
Pickrell JK, Patterson N, Loh P-R, Lipson M, Berger B, Stoneking M, Pakendorf B, Reich D. 2014. Ancient west Eurasian ancestry in southern and eastern Africa. *Proc. Natl. Acad. Sci. U. S. A.* 111:2632–2637.
Piperno DR. 2011. The Origins of Plant Cultivation and Domestication in the New World Tropics. *Current Anthropology.* 52:S453–S470.
Popejoy AB, Fullerton SM. 2016. Genomics is failing on diversity.
Posth C, Nakatsuka N, Lazaridis I, Skoglund P, Mallick S, Lamnidis TC, Rohland N, Nägele K, Adamski N, Bertolini E, et al. 2018. Reconstructing the Deep Population History of Central and South America. *Cell* 175:1185–1197.e22.
Potter BA, Baichtal JF, Beaudoin AB, Fehren-Schmitz L, Haynes CV, Holliday VT, Holmes CE, Ives JW, Kelly RL, Llamas B, et al. 2018. Current evidence allows multiple models for the peopling of the Americas. *Sci Adv* 4:eaat5473.

Poznik GD, Henn BM, Yee M-C, Sliwerska E, Euskirchen GM, Lin AA, Snyder M, Quintana-Murci L, Kidd JM, Underhill PA, et al. 2013. Sequencing Y chromosomes resolves discrepancy in time to common ancestor of males versus females. *Science* 341:562–565.

Poznik GD, Xue Y, Mendez FL, Willems TF, Massaia A, Wilson Sayres MA, Ayub Q, McCarthy SA, Narechania A, Kashin S, et al. 2016. Punctuated bursts in human male demography inferred from 1,244 worldwide Y-chromosome sequences. *Nat. Genet.* 48:593–599.

Prates L, Politis G, Steele J. 2013. Radiocarbon chronology of the early human occupation of Argentina. *Quat. Int.* 301:104–122.

Price AL, Tandon A, Patterson N, Barnes KC, Rafaels N, Ruczinski I, Beaty TH, Mathias R, Reich D, Myers S. 2009. Sensitive detection of chromosomal segments of distinct ancestry in admixed populations. *PLoS Genet.* 5:e1000519.

Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155:945–959.

Quilter J, Hoopes JW. 2003. Gold and Power in Ancient Costa Rica, Panama, and Colombia: A Symposium at Dumbarton Oaks, 9 and 10 October 1999. Dumbarton Oaks

Raghavan M, DeGiorgio M, Albrechtsen A, Moltke I, Skoglund P, Korneliussen TS, Grønnow B, Appelt M, Gulløv HC, Friesen TM, et al. 2014. The genetic prehistory of the New World Arctic. *Science* 345:1255832.

Rasmussen M, Anzick SL, Waters MR, Skoglund P, DeGiorgio M, Stafford TW Jr, Rasmussen S, Moltke I, Albrechtsen A, Doyle SM, et al. 2014. The genome of a Late Pleistocene human from a Clovis burial site in western Montana. *Nature* 506:225–229.

Rasmussen M, Li Y, Lindgreen S, Pedersen JS, Albrechtsen A, Moltke I, Metspalu M, Metspalu E, Kivisild T, Gupta R, et al. 2010. Ancient human genome sequence of an extinct Palaeo-Eskimo. *Nature* 463:757–762.

Rasmussen M, Sikora M, Albrechtsen A, Korneliussen TS, Moreno-Mayar JV, Poznik GD, Zollikofer CPE, de León MP, Allentoft ME, Moltke I, et al. 2015. The ancestry and affiliations of Kennewick Man. *Nature* 523:455–458.

Redwood SD. 2020. Late Pleistocene to Holocene sea level rise in the Gulf of Panama, Panama, and its influence on early human migration through the Isthmus. *Caribb. J. Earth Sci.* 51:15–31.

Regueiro M, Cadenas AM, Gayden T, Underhill PA, Herrera RJ. 2006. Iran: tri-continental nexus for Y-chromosome driven migration. *Hum. Hered.* 61:132–143.

Reich D, Patterson N, Campbell D, Tandon A, Mazieres S, Ray N, Parra MV, Rojas W, Duque C, Mesa N, et al. 2012. Reconstructing Native American population history. *Nature* 488:370–374.

Reich D, Thangaraj K, Patterson N, Price AL, Singh L. 2009. Reconstructing Indian population history. *Nature* 461:489–494.

Renaud G, Slon V, Duggan AT, Kelso J. 2015. Schmutzi: estimation of contamination and endogenous mitochondrial consensus calling for ancient DNA. *Genome Biol.* 16:224.

Rocca RA, Magoon G, Reynolds DF, Krahn T, Tilroe VO, Op den Velde Boots PM, Grierson AJ. 2012. Discovery of Western European R1b1a2 Y chromosome variants in 1000 genomes project data: an online community approach. *PLoS One* 7:e41634.

Rohland N, Harney E, Mallick S, Nordenfelt S, Reich D. 2015. Partial uracil–DNA–glycosylase treatment for screening of ancient DNA. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 370:20130624.

Rosenberg NA, Huang L, Jewett EM, Szpiech ZA, Jankovic I, Boehnke M. 2010. Genome-wide association studies in diverse populations. *Nat. Rev. Genet.* 11:356–366.

Ruiz-Linares A, Adhikari K, Acuña-Alonzo V, Quinto-Sanchez M, Jaramillo C, Arias W, Fuentes M, Pizarro M, Everardo P, de Avila F, et al. 2014. Admixture in Latin America: geographic structure, phenotypic diversity and self-perception of ancestry based on 7,342 individuals. *PLoS Genet.* 10:e1004572.

Salter-Townshend M, Myers S. 2019. Fine-Scale Inference of Ancestry Segments Without Prior Knowledge of Admixing Groups. *Genetics* 212:869–889.

Salzano FM, Bortolini MC. 2005. The Evolution and Genetics of Latin American Populations. Cambridge University Press

Sánchez-Albornoz N. 1994. La fundación Rockefeller y los exiliados españoles. *Historia 16*:116–120.

Sánchez-Albornoz N. 2014. Historia mínima de la población de América Latina, desde los tiempos precolombinos al año 2025. El Colegio de México

Sánchez-Albornoz N, Others. 1974. The Population of Latin America.

Sanger F, Coulson AR. 1975. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *J. Mol. Biol.* 94:441–448.

Scheib CL, Li H, Desai T, Link V, Kendall C, Dewar G, Griffith PW, Mörseburg A, Johnson JR, Potter A, et al. 2018. Ancient human parallel lineages within North America contributed to a coastal expansion. *Science* 360:1024–1027.

Scozzari R, Massaia A, Trombetta B, Bellusci G, Myres NM, Novelletto A, Cruciani F. 2014. An unbiased resource of novel SNP markers provides a new chronology for the human Y chromosome and reveals a deep phylogenetic structure in Africa. *Genome Res.* 24:535–544.

Sengupta S, Zhivotovsky LA, King R, Mehdi SQ, Edmonds CA, Chow C-ET, Lin AA, Mitra M, Sil SK, Ramesh A, et al. 2006. Polarity and temporality of high-resolution y-chromosome distributions in India identify both indigenous and exogenous expansions and reveal minor genetic influence of Central Asian pastoralists. *Am. J. Hum. Genet.* 78:202–221.

Sharma A, Jaloree S, Thakur RS. 2018. Review of Clustering Methods: Toward Phylogenetic Tree Constructions. In: Proceedings of International Conference on Recent Advancement on Computer and Communication. Springer Singapore. p. 475–480.

Sikora M, Pitulko VV, Sousa VC, Allentoft ME, Vinner L, Rasmussen S, Margaryan A, de Barros Damgaard P, de la Fuente C, Renaud G, et al. 2019. The population history of northeastern Siberia since the Pleistocene. *Nature* 570:182–188.

Skoglund P, Mathieson I. 2018. Ancient Genomics of Modern Humans: The First Decade. *Annu. Rev. Genomics Hum. Genet.* 19:381–404.

Solberg CE. 1970. Immigration and nationalism, Argentina and Chile, 1890–1914. Published for the Institute of Latin American Studies by the University of Texas Press

Tamm E, Kivisild T, Reidla M, Metspalu M, Smith DG, Mulligan CJ, Bravi CM, Rickards O, Martinez-Labarga C, Khusnutdinova EK, et al. 2007. Beringian standstill and spread of Native American founders. *PLoS One* 2:e829.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol. Biol. Evol.* 30:2725–2729.

Thompson EA. 2013. Identity by descent: variation in meiosis, across genomes, and in populations. *Genetics* 194:301–326.
Thornton J, Thornton JKT. 1998. Africa and Africans in the Making of the Atlantic World, 1400–1800. Cambridge University Press
Thornton R. 1987. American Indian Holocaust and Survival: A Population History Since 1492. University of Oklahoma Press
Torroni A, Schurr TG, Cabell MF, Brown MD, Neel JV, Larsen M, Smith DG, Vullo CM, Wallace DC. 1993. Asian affinities and continental radiation of the four founding Native American mtDNAs. *Am. J. Hum. Genet.* 53:563–590.
Tosi L. 2002. La tutela internazionale dell'emigrazione. *P. Bevilacqua, A. De Clementi, E. Franzina (a cura di), Storia dell'emigrazione italiana* 2:439–456.
Underhill PA, Jin L, Zemans R, Oefner PJ, Cavalli-Sforza LL. 1996. A pre-Columbian Y chromosome-specific transition and its implications for human evolutionary history. *Proc. Natl. Acad. Sci. U. S. A.* 93:196–200.
Underhill PA, Poznik GD, Rootsi S, Järve M, Lin AA, Wang J, Passarelli B, Kanbar J, Myres NM, King RJ, et al. 2015. The phylogenetic and geographic structure of Y-chromosome haplogroup R1a. *Eur. J. Hum. Genet.* 23:124–131.
Veerappa AM, Padakannaya P, Ramachandra NB. 2013. Copy number variation-based polymorphism in a new pseudoautosomal region 3 (PAR3) of a human X-chromosome-transposed region (XTR) in the Y chromosome. *Funct. Integr. Genomics* 13:285–293.
Waters MR. 2019. Late Pleistocene exploration and settlement of the Americas by modern humans. *Science*. 365.
Waters MR, Forman SL, Jennings TA, Nordt LC, Driese SG, Feinberg JM, Keene JL, Halligan J, Lindquist A, Pierson J, et al. 2011. The Buttermilk Creek complex and the origins of Clovis at the Debra L. Friedkin site, Texas. *Science* 331:1599–1603.
Waters MR, Stafford TW Jr. 2007. Redefining the age of Clovis: implications for the peopling of the Americas. *Science* 315:1122–1126.
Wei L-H, Yan S, Lu Y, Wen S-Q, Huang Y-Z, Wang L-X, Li S-L, Yang Y-J, Wang X-F, Zhang C, et al. 2018. Whole-sequence analysis indicates that the Y chromosome C2*-Star Cluster traces back to ordinary Mongols, rather than Genghis Khan. *Eur. J. Hum. Genet.* 26:230–237.
Weir BS, Anderson AD, Hepler AB. 2006. Genetic relatedness analysis: modern data and new challenges. *Nat. Rev. Genet.* 7:771–780.
Wei W, Ayub Q, Chen Y, McCarthy S, Hou Y, Carbone I, Xue Y, Tyler-Smith C. 2013. A calibrated human Y-chromosomal phylogeny based on resequencing. *Genome Res.* 23:388–395.
Wenger AM, Peluso P, Rowell WJ, Chang P-C, Hall RJ, Concepcion GT, Ebler J, Fungtammasan A, Kolesnikov A, Olson ND, et al. 2019. Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. *Nat. Biotechnol.* 37:1155–1162.
Williams RC, Long JC, Hanson RL, Sievers ML, Knowler WC. 2000. Individual Estimates of European Genetic Admixture Associated with Lower Body-Mass Index, Plasma Glucose, and Prevalence of Type 2 Diabetes in Pima Indians. *The American Journal of Human Genetics*. 66:527–538.
Wittke C. 1952. Refugees of revolution: The German forty-eighters in America. Baltimore, MD: University of Pennsylvania Press.
Wong AS. 2006. Race and the nation in liberal Italy, 1861–1911: Meridionalism, empire, and diaspora. Palgrave Macmillan

Wood ET, Stover DA, Ehret C, Destro-Bisol G, Spedini G, McLeod H, Louie L, Bamshad M, Strassmann BI, Soodyall H, et al. 2005. Contrasting patterns of Y chromosome and mtDNA variation in Africa: evidence for sex-biased demographic processes. *Eur. J. Hum. Genet.* 13:867–876.

Wright S. 1931. Evolution in Mendelian Populations. *Genetics* 16:97–159.

Xue Y, Wang Q, Long Q, Ng BL, Swerdlow H, Burton J, Skuce C, Taylor R, Abdellah Z, Zhao Y, et al. 2009. Human Y chromosome base-substitution mutation rate measured by direct sequencing in a deep-rooting pedigree. *Curr. Biol.* 19:1453–1457.

Y Chromosome Consortium. 2002. A nomenclature system for the tree of human Y-chromosomal binary haplogroups. *Genome Res.* 12:339–348.

Zegura SL, Karafet TM, Zhivotovsky LA, Hammer MF. 2004. High-resolution SNPs and microsatellite haplotypes point to a single, recent entry of Native American Y chromosomes into the Americas. *Mol. Biol. Evol.* 21:164–175.

Zhou D, Udpa N, Ronen R, Stobdan T, Liang J, Appenzeller O, Zhao HW, Yin Y, Du Y, Guo L, et al. 2013. Whole-genome sequencing uncovers the genetic basis of chronic mountain sickness in Andean highlanders. *Am. J. Hum. Genet.* 93:452–462.

Web resources

Slave voyages	https://www.slavevoyages.org/
ISOGG	https://isogg.org/tree/

ACKNOWLEDGEMENTS

When I decided to move to Estonia for my PhD, I was excited and, of course, very nervous. After four years, I am more than glad to have jumped into this adventure, and I am sure that was a great decision. These years were tough, but they were also great mainly because of the great people I have met and shared this adventure with me.

First and foremost, I want to thank my supervisors, starting with Mait, for welcoming me to the EBC group and to Estonia, where I felt very soon at home, and for all the opportunity he granted me. A huge thank you to Luca, who helped me understand how things worked in a new country and always tried to make the time spent here more fun and enjoyable. I want to thank him also for all the precious feedback and suggestions he always offered me. I want to thank Toomas because even though he is no longer my official supervisor, he was always ready to help me when needed with his experience and great ideas. Finally, I am incredibly grateful to Francesco, who taught me everything I know, or at least he tried and had to be so patient with me. I feel so lucky to have met him and have worked with him during these years; he was always able to make the work more fun and less stressful with his incredible optimism.

The time spent in Estonia wouldn't have been so great if I did not have my colleagues. I want to thank Burak, with whom I started this PhD, for his support and help; I am glad that I had someone to share all the bureaucratic issues that we faced, especially in the last months. Starting from the second year, I had a true friend, Ludovica, with whom I shared many memorable moments and experiences. I am truly grateful to her for her friendship and the support she always offered me. Many many thanks to Davide, Rodrigo, Katri and Mayukh for their friendship and all the fun that we had together. Thanks to Helen, Ene and Siiri, which helped me with everything related to the wet lab analysis and were always very friendly and patient. I want to thank the other residents of the pigeon room: Ajai, Hovik, Stefania, Mathilde, Tina and Anne-Mai. It was so lovely to be in the same office with you. Then, of course, thanks to all the great people that I worked with and shared many nice moments (both work and fun related): Vasili, Lena, Monika, Freddi, Meriam, Marcel, Lehti, Bayazit and Erwan. Then, many thanks to Kristiina for her kindness, guidance and availability. Thanks to Professor Villems for his wisdom and knowledge. I would like to thank Jüri for his super entertaining stories, his pictures and for always being so kind to me. A special thank goes to Merilin because she was extremely helpful during these four years and extremely patient every time I asked her something stupid. I would like to thank Mari for her kind help with the translations in Estonian. A special thanks to Professor Maido Remm for kindly agreeing to be my internal reviewer.

I am thankful to the European Regional Developmental Fund and the Archimedes Foundation for funding a part of my PhD studies and for supporting my attendance of international courses and conferences. I am very grateful to Garrett Hellenthal and his group at UCL for hosting me in their lab; it was a great

experience, although shorter than expected. I would like to thank Ornella Semino and all the people at the lab in Pavia, with whom I started my adventure in the Human Population Genetics field, for hosting me when in Italy and for the constant support. Thanks to Alessandro for his friendship and for believing in me. Thanks to Eleonora and Enrico for adopting me during the Estonian lockdown and for sharing the food with me.

I want to thank my parents sincerely because when I told them that I was moving to Estonia, they immediately supported me, and they always encouraged me. Thanks also to all my relatives and friends; I was able to go through this experience only because I knew you were always there for me.

Finally, I don't even know where to start to thank Marco for everything he has done during these four years. Thanks for always believing in me, especially when I could not believe in myself, for the constant support, for putting up with me, making everything fun, helping me and everything else.

PUBLICATIONS

CURRICULUM VITAE

Name: Linda Ongaro
Date of birth: November 12, 1993
Nationality: Italian
Address: University of Tartu, Institute of Genomics, Riia 23b, 51010 Tartu, Estonia
E-mail: linda.ongaro@ut.ee

Education:

2017–2021 Doctoral studies, University of Tartu, Faculty of Science and Technology, Institute of Genomics, Chair of Evolutionary Biology

2015–2017 MSc, *cum laude*, University of Pavia, Department of Biology and Biotechnology, Group of Genomics of Human and Animal Populations, supervisor Ornella Semino "The peopling of South America: new information from the phylogenetic refinement of human Y-chromosome haplogroup Q"

2012–2015 BSc, University of Pavia, Department of Biology and Biotechnology, Group of Genomics of Human and Animal Populations, supervisor Ornella Semino "Characterization of the haplogroup I2a-L160 of human Y-chromosome in Europe"

Professional employment:

2017–… University of Tartu, Institute of Genomics, Junior Researcher

Teaching:

2020 Teaching assistant, University of Pavia, Department of Earth and Environmental Sciences, Bachelor in Natural Sciences and Technology, Genetics and Human Biology course

2019 Teaching assistant, University of Tartu, Bachelor in Science and Technology, Chair of Evolutionary Biology, Evolution and the Natural World course

2018 Teaching assistant, University of Tartu, Bachelor in Science and Technology, Chair of Evolutionary Biology, Evolutionary processes course

2016–2017 Tutor, University of Pavia, Department of Biology and Biotechnology, Bachelor in Biology and in Biotechnology, Genetics course

International Courses and Conferences:

2019 Oral presentation at AAI conference, Padua, Italy

2019 Poster presentation at Centenary of Human Population Genetics, Moscow, Russia

2019 EMBO Practical Course, Population genomics: Background, tools and programming, Procida (NA), Italy

2018 Oral presentation at FISV, Rome, Italy

2018 PhD summer course, Analyses of genotyping and sequencing data in medical and population genetics, Copenhagen, Denmark

Publications:

Vicuña L, Norambuena T, Miranda JP, Pereira A, Mericq V, **Ongaro L**, Montinaro F, Santos JL, Eyheramendy S. 2021. **Novel loci and mapuche genetic ancestry are associated with pubertal growth traits in Chilean boys.** *Hum Genet*, 10.1007/s00439-021-02290-3.

Capodiferro MR, Aram B, Raveane A, Rambaldi Migliore N, Colombo G, **Ongaro L**, Rivera J, Mendizábal T, Hernández-Mora I, Tribaldos M, Perego UA, Li H, Scheib CL, Modi A, Gòmez-Carballa A, Grugni V, Lombardo G, Hellenthal G, Pascale JM, Bertolini F, Grieco GS, Cereda C, Lari M, Caramelli D, Pagani L, Metspalu M, Friedrich R, Knipper C, Olivieri A, Salas A, Cooke R, Montinaro F, Motta J, Torroni A, Martín JG, Semino O, Malhi RS, Achilli A. 2021. **Archaeogenomic distinctiveness of the Isthmo-Colombian area.** *Cell*, 184(7):1706-1723.e24.

Eaaswarkhanth M, Pathak AK, **Ongaro L**, Montinaro F, Hebbar P, Alsmadi O, Metspalu M, Al-Mulla F, Thanaraj TA. **Unraveling a fine-scale high genetic heterogeneity and recent continental connections of an Arabian Peninsula population.** 2021. *Eur J Hum Genet,* 10.1038/s41431-021-00861-6.

Molinaro L, Marnetto D, Mondal M, **Ongaro L**, Yelmen B, Lawson DJ, Montinaro F, Pagani L. 2021. **A Chromosome-Painting-Based Pipeline to Infer Local Ancestry under Limited Source Availability.** *Genome Biol Evol,* 13(4):evab025.

Yelmen B, Decelle A, **Ongaro L,** Marnetto D, Tallec C, Montinaro F, Furtlehner C, Pagani L, Jay F. 2021. Creating artificial human genomes **using generative neural networks.** *PLoS Genet,* 17(2):e1009303.

Grugni V, Raveane A, Colombo G, Nici C, Crobu F, **Ongaro L**, Battaglia V, Sanna D, Al-Zahery N, Fiorani O, Lisa A, Ferretti L, Achilli A, Olivieri A, Francalacci P, Piazza A, Torroni A, Semino O. 2019. **Y-chromosome and Surname Analyses for Reconstructing Past Population Structures: The Sardinian Population as a Test Case.** *Int J Mol Sci*, 20(22):5763.

Ongaro L, Scliar MO, Flores R, Raveane A, Marnetto D, Sarno S, Gnecchi-Ruscone GA, Alarcón-Riquelme ME, Patin E, Wangkumhang P, Hellenthal G, Gonzalez-Santos M, King RJ, Kouvatsi A, Balanovsky O, Balanovska E, Atramentova L, Turdikulova S, Mastana S, Marjanovic D, Mulahasanovic L, Leskovac A, Lima-Costa MF, Pereira AC, Barreto ML, Horta BL, Mabunda N, May CA, Moreno-Estrada A, Achilli A, Olivieri A, Semino O, Tambets K, Kivisild T, Luiselli D, Torroni A, Capelli C, Tarazona-Santos E, Metspalu M, Pagani L, Montinaro F. 2019. **The Genomic Impact of European Colonization of the Americas.** *Curr Biol*, 29(23):3974-3986.e4.

Grugni V, Raveane A, **Ongaro L,** Battaglia V, Trombetta B, Colombo G, Capodiferro MR, Olivieri A, Achilli A, Perego UA, Motta J, Tribaldos M, Woodward SR, Ferretti L, Cruciani F, Torroni A, Semino O. 2019. **Analysis of the human Y-chromosome haplogroup Q characterizes ancient population movements in Eurasia and the Americas.** *BMC Biol*, 17(1):3.

Pathak AK, Kadian A, Kushniarevich A, Montinaro F, Mondal M, **Ongaro L,** Singh M, Kumar P, Rai N, Parik J, Metspalu E, Rootsi S, Pagani L, Kivisild T, Metspalu M, Chaubey G, Villems R. 2018. T**he Genetic Ancestry of Modern Indus Valley Populations from Northwest India.** *Am J Hum Genet,* 103(6):918–929.

ELULOOKIRJELDUS

Nimi: Linda Ongaro
Sünniaeg: 12.11.1993
Kontakt: Tartu Ülikool, Genoomika Instituut Riia 23b, 51010 Tartu, Eesti
E-post: linda.ongaro@ut.ee

Hariduskäik:

2017–… Doktoriõpe, Tartu Ülikool, loodus- ja täppisteaduste valdkond, genoomika instituut, evolutsioonilise bioloogia õppetool.

2015–2017 Magistrikraad, *cum laude*, Pavia Ülikool, bioloogia ja biotehnoloogia osakond, inim- ja loomapopulatsioonide genoomika töörühm, juhendaja Ornella Semino "Lõuna-Ameerika asustamine: uus informatsioon inimese Y-kromosoomi haplogrupi Q fülogeneetilisest täpsustamisest."

2012–2015 Bakalaureusekraad, Pavia Ülikool, bioloogia ja biotehnoloogia osakond, inim- ja loomapopulatsioonide genoomika töörühm, juhendaja Ornella Semino "Inimese Y-kromosoomi haplogrupi I2a-L160 iseloomustus Euroopas."

Töökogemus:

2017–… Tartu Ülikool, genoomika instituut, nooremteadur.

Õpetamiskogemus:

2020 Õppeülesannete täitja, Pavia Ülikool, maa- ja keskkonnateaduste osakond, loodusteaduste ja tehnoloogia bakalaureuseõppekava, kursus Geneetika ja inimese bioloogia.

2019 Õppeülesannete täitja, Tartu Ülikool, loodusteaduste ja tehnoloogia rahvusvaheline bakalaureuseõppekava (*Science & Technology*), evolutsioonilise bioloogia õppetool, kursus Elusloodus ja evolutsioon.

2018 Õppeülesannete täitja, Tartu Ülikool, loodusteaduste ja tehnoloogia rahvusvaheline bakalaureuseõppekava (*Science & Technology*), evolutsioonilise bioloogia õppetool, kursus Evolutsiooniprotsessid.

2016–2017 Juhendaja, Pavia Ülikool, bioloogia ja biotehnoloogia osakond, bioloogia ja biotehnoloogia bakalaureuseõppekava, kursus Geneetika.

Rahvusvahelised kursused ja konverentsid:

2019 Suuline ettekanne, AAI konverents, Padua, Itaalia.

2019 Poster, 100 aastat inimese populatsioonigeneetikat, Moskva, Venemaa.

2019	EMBO praktiline kursus “Populatsioonigenoomika: taust, tööriistad ja programmeerimine”, Procida (NA), Itaalia.
2018	Suuline ettekanne, FISV, Rooma, Itaalia.
2018	Doktorantide suvekursus “Genotüpiseerimis- ja sekveneerimis-andmete analüüsid meditsiinilises ja populatsioonigeneetikas”, Kopenhaagen, Taani.

Publikatsioonid:
Loetletud inglisekeelse CV rubriigis *Publications*

DISSERTATIONES BIOLOGICAE UNIVERSITATIS TARTUENSIS

1. **Toivo Maimets**. Studies of human oncoprotein p53. Tartu, 1991, 96 p.
2. **Enn K. Seppet**. Thyroid state control over energy metabolism, ion transport and contractile functions in rat heart. Tartu, 1991, 135 p.
3. **Kristjan Zobel**. Epifüütsete makrosamblike väärtus õhu saastuse indikaatoritena Hamar-Dobani boreaalsetes mägimetsades. Tartu, 1992, 131 lk.
4. **Andres Mäe**. Conjugal mobilization of catabolic plasmids by transposable elements in helper plasmids. Tartu, 1992, 91 p.
5. **Maia Kivisaar**. Studies on phenol degradation genes of *Pseudomonas* sp. strain EST 1001. Tartu, 1992, 61 p.
6. **Allan Nurk**. Nucleotide sequences of phenol degradative genes from *Pseudomonas sp.* strain EST 1001 and their transcriptional activation in *Pseudomonas putida*. Tartu, 1992, 72 p.
7. **Ülo Tamm**. The genus *Populus* L. in Estonia: variation of the species biology and introduction. Tartu, 1993, 91 p.
8. **Jaanus Remme**. Studies on the peptidyltransferase centre of the *E.coli* ribosome. Tartu, 1993, 68 p.
9. **Ülo Langel**. Galanin and galanin antagonists. Tartu, 1993, 97 p.
10. **Arvo Käärd**. The development of an automatic online dynamic fluorescense-based pH-dependent fiber optic penicillin flowthrought biosensor for the control of the benzylpenicillin hydrolysis. Tartu, 1993, 117 p.
11. **Lilian Järvekülg**. Antigenic analysis and development of sensitive immunoassay for potato viruses. Tartu, 1993, 147 p.
12. **Jaak Palumets**. Analysis of phytomass partition in Norway spruce. Tartu, 1993, 47 p.
13. **Arne Sellin**. Variation in hydraulic architecture of *Picea abies* (L.) Karst. trees grown under different enviromental conditions. Tartu, 1994, 119 p.
13. **Mati Reeben**. Regulation of light neurofilament gene expression. Tartu, 1994, 108 p.
14. **Urmas Tartes**. Respiration rhytms in insects. Tartu, 1995, 109 p.
15. **Ülo Puurand**. The complete nucleotide sequence and infections *in vitro* transcripts from cloned cDNA of a potato A potyvirus. Tartu, 1995, 96 p.
16. **Peeter Hõrak**. Pathways of selection in avian reproduction: a functional framework and its application in the population study of the great tit (*Parus major*). Tartu, 1995, 118 p.
17. **Erkki Truve**. Studies on specific and broad spectrum virus resistance in transgenic plants. Tartu, 1996, 158 p.
18. **Illar Pata**. Cloning and characterization of human and mouse ribosomal protein S6-encoding genes. Tartu, 1996, 60 p.
19. **Ülo Niinemets**. Importance of structural features of leaves and canopy in determining species shade-tolerance in temperature deciduous woody taxa. Tartu, 1996, 150 p.

20. **Ants Kurg**. Bovine leukemia virus: molecular studies on the packaging region and DNA diagnostics in cattle. Tartu, 1996, 104 p.
21. **Ene Ustav**. E2 as the modulator of the BPV1 DNA replication. Tartu, 1996, 100 p.
22. **Aksel Soosaar**. Role of helix-loop-helix and nuclear hormone receptor transcription factors in neurogenesis. Tartu, 1996, 109 p.
23. **Maido Remm**. Human papillomavirus type 18: replication, transformation and gene expression. Tartu, 1997, 117 p.
24. **Tiiu Kull**. Population dynamics in *Cypripedium calceolus* L. Tartu, 1997, 124 p.
25. **Kalle Olli**. Evolutionary life-strategies of autotrophic planktonic microorganisms in the Baltic Sea. Tartu, 1997, 180 p.
26. **Meelis Pärtel**. Species diversity and community dynamics in calcareous grassland communities in Western Estonia. Tartu, 1997, 124 p.
27. **Malle Leht**. The Genus *Potentilla* L. in Estonia, Latvia and Lithuania: distribution, morphology and taxonomy. Tartu, 1997, 186 p.
28. **Tanel Tenson**. Ribosomes, peptides and antibiotic resistance. Tartu, 1997, 80 p.
29. **Arvo Tuvikene**. Assessment of inland water pollution using biomarker responses in fish *in vivo* and *in vitro.* Tartu, 1997, 160 p.
30. **Urmas Saarma**. Tuning ribosomal elongation cycle by mutagenesis of 23S rRNA. Tartu, 1997, 134 p.
31. **Henn Ojaveer**. Composition and dynamics of fish stocks in the gulf of Riga ecosystem. Tartu, 1997, 138 p.
32. **Lembi Lõugas**. Post-glacial development of vertebrate fauna in Estonian water bodies. Tartu, 1997, 138 p.
33. **Margus Pooga**. Cell penetrating peptide, transportan, and its predecessors, galanin-based chimeric peptides. Tartu, 1998, 110 p.
34. **Andres Saag**. Evolutionary relationships in some cetrarioid genera (Lichenized Ascomycota). Tartu, 1998, 196 p.
35. **Aivar Liiv**. Ribosomal large subunit assembly *in vivo*. Tartu, 1998, 158 p.
36. **Tatjana Oja**. Isoenzyme diversity and phylogenetic affinities among the eurasian annual bromes (*Bromus* L., Poaceae). Tartu, 1998, 92 p.
37. **Mari Moora**. The influence of arbuscular mycorrhizal (AM) symbiosis on the competition and coexistence of calcareous grassland plant species. Tartu, 1998, 78 p.
38. **Olavi Kurina**. Fungus gnats in Estonia (*Diptera: Bolitophilidae, Keroplatidae, Macroceridae, Ditomyiidae, Diadocidiidae, Mycetophilidae*). Tartu, 1998, 200 p.
39. **Andrus Tasa**. Biological leaching of shales: black shale and oil shale. Tartu, 1998, 98 p.
40. **Arnold Kristjuhan**. Studies on transcriptional activator properties of tumor suppressor protein p53. Tartu, 1998, 86 p.
41. **Sulev Ingerpuu**. Characterization of some human myeloid cell surface and nuclear differentiation antigens. Tartu, 1998, 163 p.

42. **Veljo Kisand**. Responses of planktonic bacteria to the abiotic and biotic factors in the shallow lake Võrtsjärv. Tartu, 1998, 118 p.
43. **Kadri Põldmaa**. Studies in the systematics of hypomyces and allied genera (Hypocreales, Ascomycota). Tartu, 1998, 178 p.
44. **Markus Vetemaa**. Reproduction parameters of fish as indicators in environmental monitoring. Tartu, 1998, 117 p.
45. **Heli Talvik**. Prepatent periods and species composition of different *Oesophagostomum* spp. populations in Estonia and Denmark. Tartu, 1998, 104 p.
46. **Katrin Heinsoo**. Cuticular and stomatal antechamber conductance to water vapour diffusion in *Picea abies* (L.) karst. Tartu, 1999, 133 p.
47. **Tarmo Annilo**. Studies on mammalian ribosomal protein S7. Tartu, 1998, 77 p.
48. **Indrek Ots**. Health state indicies of reproducing great tits (*Parus major*): sources of variation and connections with life-history traits. Tartu, 1999, 117 p.
49. **Juan Jose Cantero**. Plant community diversity and habitat relationships in central Argentina grasslands. Tartu, 1999, 161 p.
50. **Rein Kalamees**. Seed bank, seed rain and community regeneration in Estonian calcareous grasslands. Tartu, 1999, 107 p.
51. **Sulev Kõks**. Cholecystokinin (CCK) – induced anxiety in rats: influence of environmental stimuli and involvement of endopioid mechanisms and serotonin. Tartu, 1999, 123 p.
52. **Ebe Sild**. Impact of increasing concentrations of O_3 and CO_2 on wheat, clover and pasture. Tartu, 1999, 123 p.
53. **Ljudmilla Timofejeva**. Electron microscopical analysis of the synaptonemal complex formation in cereals. Tartu, 1999, 99 p.
54. **Andres Valkna**. Interactions of galanin receptor with ligands and G-proteins: studies with synthetic peptides. Tartu, 1999, 103 p.
55. **Taavi Virro**. Life cycles of planktonic rotifers in lake Peipsi. Tartu, 1999, 101 p.
56. **Ana Rebane**. Mammalian ribosomal protein S3a genes and intron-encoded small nucleolar RNAs U73 and U82. Tartu, 1999, 85 p.
57. **Tiina Tamm**. Cocksfoot mottle virus: the genome organisation and translational strategies. Tartu, 2000, 101 p.
58. **Reet Kurg**. Structure-function relationship of the bovine papilloma virus E2 protein. Tartu, 2000, 89 p.
59. **Toomas Kivisild**. The origins of Southern and Western Eurasian populations: an mtDNA study. Tartu, 2000, 121 p.
60. **Niilo Kaldalu**. Studies of the TOL plasmid transcription factor XylS. Tartu, 2000, 88 p.
61. **Dina Lepik**. Modulation of viral DNA replication by tumor suppressor protein p53. Tartu, 2000, 106 p.
62. **Kai Vellak**. Influence of different factors on the diversity of the bryophyte vegetation in forest and wooded meadow communities. Tartu, 2000, 122 p.

63. **Jonne Kotta**. Impact of eutrophication and biological invasionas on the structure and functions of benthic macrofauna. Tartu, 2000, 160 p.
64. **Georg Martin**. Phytobenthic communities of the Gulf of Riga and the inner sea the West-Estonian archipelago. Tartu, 2000, 139 p.
65. **Silvia Sepp**. Morphological and genetical variation of *Alchemilla L.* in Estonia. Tartu, 2000. 124 p.
66. **Jaan Liira**. On the determinants of structure and diversity in herbaceous plant communities. Tartu, 2000, 96 p.
67. **Priit Zingel**. The role of planktonic ciliates in lake ecosystems. Tartu, 2001, 111 p.
68. **Tiit Teder**. Direct and indirect effects in Host-parasitoid interactions: ecological and evolutionary consequences. Tartu, 2001, 122 p.
69. **Hannes Kollist**. Leaf apoplastic ascorbate as ozone scavenger and its transport across the plasma membrane. Tartu, 2001, 80 p.
70. **Reet Marits**. Role of two-component regulator system PehR-PehS and extracellular protease PrtW in virulence of *Erwinia Carotovora* subsp. *Carotovora*. Tartu, 2001, 112 p.
71. **Vallo Tilgar**. Effect of calcium supplementation on reproductive performance of the pied flycatcher *Ficedula hypoleuca* and the great tit *Parus major,* breeding in Nothern temperate forests. Tartu, 2002, 126 p.
72. **Rita Hõrak**. Regulation of transposition of transposon Tn*4652* in *Pseudomonas putida*. Tartu, 2002, 108 p.
73. **Liina Eek-Piirsoo**. The effect of fertilization, mowing and additional illumination on the structure of a species-rich grassland community. Tartu, 2002, 74 p.
74. **Krõõt Aasamaa**. Shoot hydraulic conductance and stomatal conductance of six temperate deciduous tree species. Tartu, 2002, 110 p.
75. **Nele Ingerpuu**. Bryophyte diversity and vascular plants. Tartu, 2002, 112 p.
76. **Neeme Tõnisson**. Mutation detection by primer extension on oligonucleotide microarrays. Tartu, 2002, 124 p.
77. **Margus Pensa**. Variation in needle retention of Scots pine in relation to leaf morphology, nitrogen conservation and tree age. Tartu, 2003, 110 p.
78. **Asko Lõhmus**. Habitat preferences and quality for birds of prey: from principles to applications. Tartu, 2003, 168 p.
79. **Viljar Jaks**. p53 – a switch in cellular circuit. Tartu, 2003, 160 p.
80. **Jaana Männik**. Characterization and genetic studies of four ATP-binding cassette (ABC) transporters. Tartu, 2003, 140 p.
81. **Marek Sammul**. Competition and coexistence of clonal plants in relation to productivity. Tartu, 2003, 159 p
82. **Ivar Ilves**. Virus-cell interactions in the replication cycle of bovine papillomavirus type 1. Tartu, 2003, 89 p.
83. **Andres Männik**. Design and characterization of a novel vector system based on the stable replicator of bovine papillomavirus type 1. Tartu, 2003, 109 p.

84. **Ivika Ostonen**. Fine root structure, dynamics and proportion in net primary production of Norway spruce forest ecosystem in relation to site conditions. Tartu, 2003, 158 p.
85. **Gudrun Veldre**. Somatic status of 12–15-year-old Tartu schoolchildren. Tartu, 2003, 199 p.
86. **Ülo Väli**. The greater spotted eagle *Aquila clanga* and the lesser spotted eagle *A. pomarina*: taxonomy, phylogeography and ecology. Tartu, 2004, 159 p.
87. **Aare Abroi**. The determinants for the native activities of the bovine papillomavirus type 1 E2 protein are separable. Tartu, 2004, 135 p.
88. **Tiina Kahre**. Cystic fibrosis in Estonia. Tartu, 2004, 116 p.
89. **Helen Orav-Kotta**. Habitat choice and feeding activity of benthic suspension feeders and mesograzers in the northern Baltic Sea. Tartu, 2004, 117 p.
90. **Maarja Öpik**. Diversity of arbuscular mycorrhizal fungi in the roots of perennial plants and their effect on plant performance. Tartu, 2004, 175 p.
91. **Kadri Tali**. Species structure of *Neotinea ustulata*. Tartu, 2004, 109 p.
92. **Kristiina Tambets**. Towards the understanding of post-glacial spread of human mitochondrial DNA haplogroups in Europe and beyond: a phylogeographic approach. Tartu, 2004, 163 p.
93. **Arvi Jõers**. Regulation of p53-dependent transcription. Tartu, 2004, 103 p.
94. **Lilian Kadaja**. Studies on modulation of the activity of tumor suppressor protein p53. Tartu, 2004, 103 p.
95. **Jaak Truu**. Oil shale industry wastewater: impact on river microbial community and possibilities for bioremediation. Tartu, 2004, 128 p.
96. **Maire Peters**. Natural horizontal transfer of the *pheBA* operon. Tartu, 2004, 105 p.
97. **Ülo Maiväli**. Studies on the structure-function relationship of the bacterial ribosome. Tartu, 2004, 130 p.
98. **Merit Otsus**. Plant community regeneration and species diversity in dry calcareous grasslands. Tartu, 2004, 103 p.
99. **Mikk Heidemaa**. Systematic studies on sawflies of the genera *Dolerus, Empria,* and *Caliroa* (Hymenoptera: Tenthredinidae). Tartu, 2004, 167 p.
100. **Ilmar Tõnno**. The impact of nitrogen and phosphorus concentration and N/P ratio on cyanobacterial dominance and N_2 fixation in some Estonian lakes. Tartu, 2004, 111 p.
101. **Lauri Saks**. Immune function, parasites, and carotenoid-based ornaments in greenfinches. Tartu, 2004, 144 p.
102. **Siiri Rootsi**. Human Y-chromosomal variation in European populations. Tartu, 2004, 142 p.
103. **Eve Vedler**. Structure of the 2,4-dichloro-phenoxyacetic acid-degradative plasmid pEST4011. Tartu, 2005. 106 p.
104. **Andres Tover**. Regulation of transcription of the phenol degradation *pheBA* operon in *Pseudomonas putida*. Tartu, 2005, 126 p.
105. **Helen Udras**. Hexose kinases and glucose transport in the yeast *Hansenula polymorpha*. Tartu, 2005, 100 p.

106. **Ave Suija**. Lichens and lichenicolous fungi in Estonia: diversity, distribution patterns, taxonomy. Tartu, 2005, 162 p.
107. **Piret Lõhmus**. Forest lichens and their substrata in Estonia. Tartu, 2005, 162 p.
108. **Inga Lips**. Abiotic factors controlling the cyanobacterial bloom occurrence in the Gulf of Finland. Tartu, 2005, 156 p.
109. **Krista Kaasik**. Circadian clock genes in mammalian clockwork, metabolism and behaviour. Tartu, 2005, 121 p.
110. **Juhan Javoiš**. The effects of experience on host acceptance in ovipositing moths. Tartu, 2005, 112 p.
111. **Tiina Sedman**. Characterization of the yeast *Saccharomyces cerevisiae* mitochondrial DNA helicase Hmi1. Tartu, 2005, 103 p.
112. **Ruth Aguraiuja**. Hawaiian endemic fern lineage *Diellia* (Aspleniaceae): distribution, population structure and ecology. Tartu, 2005, 112 p.
113. **Riho Teras**. Regulation of transcription from the fusion promoters generated by transposition of Tn*4652* into the upstream region of *pheBA* operon in *Pseudomonas putida*. Tartu, 2005, 106 p.
114. **Mait Metspalu**. Through the course of prehistory in India: tracing the mtDNA trail. Tartu, 2005, 138 p.
115. **Elin Lõhmussaar**. The comparative patterns of linkage disequilibrium in European populations and its implication for genetic association studies. Tartu, 2006, 124 p.
116. **Priit Kupper**. Hydraulic and environmental limitations to leaf water relations in trees with respect to canopy position. Tartu, 2006, 126 p.
117. **Heili Ilves**. Stress-induced transposition of Tn*4652* in *Pseudomonas Putida*. Tartu, 2006, 120 p.
118. **Silja Kuusk**. Biochemical properties of Hmi1p, a DNA helicase from *Saccharomyces cerevisiae* mitochondria. Tartu, 2006, 126 p.
119. **Kersti Püssa**. Forest edges on medium resolution landsat thematic mapper satellite images. Tartu, 2006, 90 p.
120. **Lea Tummeleht**. Physiological condition and immune function in great tits (*Parus major* l.): Sources of variation and trade-offs in relation to growth. Tartu, 2006, 94 p.
121. **Toomas Esperk**. Larval instar as a key element of insect growth schedules. Tartu, 2006, 186 p.
122. **Harri Valdmann**. Lynx (*Lynx lynx*) and wolf (*Canis lupus*) in the Baltic region: Diets, helminth parasites and genetic variation. Tartu, 2006. 102 p.
123. **Priit Jõers**. Studies of the mitochondrial helicase Hmi1p in *Candida albicans* and *Saccharomyces cerevisia*. Tartu, 2006. 113 p.
124. **Kersti Lilleväli**. Gata3 and Gata2 in inner ear development. Tartu, 2007, 123 p.
125. **Kai Rünk**. Comparative ecology of three fern species: *Dryopteris carthusiana* (Vill.) H.P. Fuchs, *D. expansa* (C. Presl) Fraser-Jenkins & Jermy and *D. dilatata* (Hoffm.) A. Gray (Dryopteridaceae). Tartu, 2007, 143 p.

126. **Aveliina Helm**. Formation and persistence of dry grassland diversity: role of human history and landscape structure. Tartu, 2007, 89 p.
127. **Leho Tedersoo**. Ectomycorrhizal fungi: diversity and community structure in Estonia, Seychelles and Australia. Tartu, 2007, 233 p.
128. **Marko Mägi**. The habitat-related variation of reproductive performance of great tits in a deciduous-coniferous forest mosaic: looking for causes and consequences. Tartu, 2007, 135 p.
129. **Valeria Lulla**. Replication strategies and applications of Semliki Forest virus. Tartu, 2007, 109 p.
130. **Ülle Reier**. Estonian threatened vascular plant species: causes of rarity and conservation. Tartu, 2007, 79 p.
131. **Inga Jüriado**. Diversity of lichen species in Estonia: influence of regional and local factors. Tartu, 2007, 171 p.
132. **Tatjana Krama**. Mobbing behaviour in birds: costs and reciprocity based cooperation. Tartu, 2007, 112 p.
133. **Signe Saumaa**. The role of DNA mismatch repair and oxidative DNA damage defense systems in avoidance of stationary phase mutations in *Pseudomonas putida*. Tartu, 2007, 172 p.
134. **Reedik Mägi**. The linkage disequilibrium and the selection of genetic markers for association studies in european populations. Tartu, 2007, 96 p.
135. **Priit Kilgas**. Blood parameters as indicators of physiological condition and skeletal development in great tits (*Parus major*): natural variation and application in the reproductive ecology of birds. Tartu, 2007, 129 p.
136. **Anu Albert**. The role of water salinity in structuring eastern Baltic coastal fish communities. Tartu, 2007, 95 p.
137. **Kärt Padari**. Protein transduction mechanisms of transportans. Tartu, 2008, 128 p.
138. **Siiri-Lii Sandre**. Selective forces on larval colouration in a moth. Tartu, 2008, 125 p.
139. **Ülle Jõgar**. Conservation and restoration of semi-natural floodplain meadows and their rare plant species. Tartu, 2008, 99 p.
140. **Lauri Laanisto**. Macroecological approach in vegetation science: generality of ecological relationships at the global scale. Tartu, 2008, 133 p.
141. **Reidar Andreson**. Methods and software for predicting PCR failure rate in large genomes. Tartu, 2008, 105 p.
142. **Birgot Paavel**. Bio-optical properties of turbid lakes. Tartu, 2008, 175 p.
143. **Kaire Torn**. Distribution and ecology of charophytes in the Baltic Sea. Tartu, 2008, 98 p.
144. **Vladimir Vimberg**. Peptide mediated macrolide resistance. Tartu, 2008, 190 p.
145. **Daima Örd**. Studies on the stress-inducible pseudokinase TRB3, a novel inhibitor of transcription factor ATF4. Tartu, 2008, 108 p.
146. **Lauri Saag**. Taxonomic and ecologic problems in the genus *Lepraria* (*Stereocaulaceae*, lichenised *Ascomycota*). Tartu, 2008, 175 p.

147. **Ulvi Karu**. Antioxidant protection, carotenoids and coccidians in greenfinches – assessment of the costs of immune activation and mechanisms of parasite resistance in a passerine with carotenoid-based ornaments. Tartu, 2008, 124 p.
148. **Jaanus Remm**. Tree-cavities in forests: density, characteristics and occupancy by animals. Tartu, 2008, 128 p.
149. **Epp Moks**. Tapeworm parasites *Echinococcus multilocularis* and *E. granulosus* in Estonia: phylogenetic relationships and occurrence in wild carnivores and ungulates. Tartu, 2008, 82 p.
150. **Eve Eensalu**. Acclimation of stomatal structure and function in tree canopy: effect of light and CO_2 concentration. Tartu, 2008, 108 p.
151. **Janne Pullat**. Design, functionlization and application of an *in situ* synthesized oligonucleotide microarray. Tartu, 2008, 108 p.
152. **Marta Putrinš**. Responses of *Pseudomonas putida* to phenol-induced metabolic and stress signals. Tartu, 2008, 142 p.
153. **Marina Semtšenko**. Plant root behaviour: responses to neighbours and physical obstructions. Tartu, 2008, 106 p.
154. **Marge Starast**. Influence of cultivation techniques on productivity and fruit quality of some *Vaccinium* and *Rubus* taxa. Tartu, 2008, 154 p.
155. **Age Tats**. Sequence motifs influencing the efficiency of translation. Tartu, 2009, 104 p.
156. **Radi Tegova**. The role of specialized DNA polymerases in mutagenesis in *Pseudomonas putida.* Tartu, 2009, 124 p.
157. **Tsipe Aavik**. Plant species richness, composition and functional trait pattern in agricultural landscapes – the role of land use intensity and landscape structure. Tartu, 2009, 112 p.
158. **Kaja Kiiver**. Semliki forest virus based vectors and cell lines for studying the replication and interactions of alphaviruses and hepaciviruses. Tartu, 2009, 104 p.
159. **Meelis Kadaja**. Papillomavirus Replication Machinery Induces Genomic Instability in its Host Cell. Tartu, 2009, 126 p.
160. **Pille Hallast**. Human and chimpanzee Luteinizing hormone/Chorionic Gonadotropin beta (*LHB/CGB*) gene clusters: diversity and divergence of young duplicated genes. Tartu, 2009, 168 p.
161. **Ain Vellak**. Spatial and temporal aspects of plant species conservation. Tartu, 2009, 86 p.
162. **Triinu Remmel**. Body size evolution in insects with different colouration strategies: the role of predation risk. Tartu, 2009, 168 p.
163. **Jaana Salujõe**. Zooplankton as the indicator of ecological quality and fish predation in lake ecosystems. Tartu, 2009, 129 p.
164. **Ele Vahtmäe**. Mapping benthic habitat with remote sensing in optically complex coastal environments. Tartu, 2009, 109 p.
165. **Liisa Metsamaa**. Model-based assessment to improve the use of remote sensing in recognition and quantitative mapping of cyanobacteria. Tartu, 2009, 114 p.

166. **Pille Säälik**. The role of endocytosis in the protein transduction by cell-penetrating peptides. Tartu, 2009, 155 p.
167. **Lauri Peil**. Ribosome assembly factors in *Escherichia coli.* Tartu, 2009, 147 p.
168. **Lea Hallik**. Generality and specificity in light harvesting, carbon gain capacity and shade tolerance among plant functional groups. Tartu, 2009, 99 p.
169. **Mariliis Tark**. Mutagenic potential of DNA damage repair and tolerance mechanisms under starvation stress. Tartu, 2009, 191 p.
170. **Riinu Rannap**. Impacts of habitat loss and restoration on amphibian populations. Tartu, 2009, 117 p.
171. **Maarja Adojaan**. Molecular variation of HIV-1 and the use of this knowledge in vaccine development. Tartu, 2009, 95 p.
172. **Signe Altmäe**. Genomics and transcriptomics of human induced ovarian folliculogenesis. Tartu, 2010, 179 p.
173. **Triin Suvi**. Mycorrhizal fungi of native and introduced trees in the Seychelles Islands. Tartu, 2010, 107 p.
174. **Velda Lauringson**. Role of suspension feeding in a brackish-water coastal sea. Tartu, 2010, 123 p.
175. **Eero Talts**. Photosynthetic cyclic electron transport – measurement and variably proton-coupled mechanism. Tartu, 2010, 121 p.
176. **Mari Nelis**. Genetic structure of the Estonian population and genetic distance from other populations of European descent. Tartu, 2010, 97 p.
177. **Kaarel Krjutškov**. Arrayed Primer Extension-2 as a multiplex PCR-based method for nucleic acid variation analysis: method and applications. Tartu, 2010, 129 p.
178. **Egle Köster**. Morphological and genetical variation within species complexes: *Anthyllis vulneraria* s. l. and *Alchemilla vulgaris* (coll.). Tartu, 2010, 101 p.
179. **Erki Õunap**. Systematic studies on the subfamily Sterrhinae (Lepidoptera: Geometridae). Tartu, 2010, 111 p.
180. **Merike Jõesaar**. Diversity of key catabolic genes at degradation of phenol and *p*-cresol in pseudomonads. Tartu, 2010, 125 p.
181. **Kristjan Herkül**. Effects of physical disturbance and habitat-modifying species on sediment properties and benthic communities in the northern Baltic Sea. Tartu, 2010, 123 p.
182. **Arto Pulk**. Studies on bacterial ribosomes by chemical modification approaches. Tartu, 2010, 161 p.
183. **Maria Põllupüü**. Ecological relations of cladocerans in a brackish-water ecosystem. Tartu, 2010, 126 p.
184. **Toomas Silla**. Study of the segregation mechanism of the Bovine Papillomavirus Type 1. Tartu, 2010, 188 p.
185. **Gyaneshwer Chaubey**. The demographic history of India: A perspective based on genetic evidence. Tartu, 2010, 184 p.
186. **Katrin Kepp**. Genes involved in cardiovascular traits: detection of genetic variation in Estonian and Czech populations. Tartu, 2010, 164 p.

187. **Virve Sõber**. The role of biotic interactions in plant reproductive performance. Tartu, 2010, 92 p.
188. **Kersti Kangro**. The response of phytoplankton community to the changes in nutrient loading. Tartu, 2010, 144 p.
189. **Joachim M. Gerhold**. Replication and Recombination of mitochondrial DNA in Yeast. Tartu, 2010, 120 p.
190. **Helen Tammert**. Ecological role of physiological and phylogenetic diversity in aquatic bacterial communities. Tartu, 2010, 140 p.
191. **Elle Rajandu**. Factors determining plant and lichen species diversity and composition in Estonian *Calamagrostis* and *Hepatica* site type forests. Tartu, 2010, 123 p.
192. **Paula Ann Kivistik**. ColR-ColS signalling system and transposition of Tn*4652* in the adaptation of *Pseudomonas putida.* Tartu, 2010, 118 p.
193. **Siim Sõber**. Blood pressure genetics: from candidate genes to genome-wide association studies. Tartu, 2011, 120 p.
194. **Kalle Kipper**. Studies on the role of helix 69 of 23S rRNA in the factor-dependent stages of translation initiation, elongation, and termination. Tartu, 2011, 178 p.
195. **Triinu Siibak**. Effect of antibiotics on ribosome assembly is indirect. Tartu, 2011, 134 p.
196. **Tambet Tõnissoo**. Identification and molecular analysis of the role of guanine nucleotide exchange factor RIC-8 in mouse development and neural function. Tartu, 2011, 110 p.
197. **Helin Räägel**. Multiple faces of cell-penetrating peptides – their intracellular trafficking, stability and endosomal escape during protein transduction. Tartu, 2011, 161 p.
198. **Andres Jaanus**. Phytoplankton in Estonian coastal waters – variability, trends and response to environmental pressures. Tartu, 2011, 157 p.
199. **Tiit Nikopensius**. Genetic predisposition to nonsyndromic orofacial clefts. Tartu, 2011, 152 p.
200. **Signe Värv**. Studies on the mechanisms of RNA polymerase II-dependent transcription elongation. Tartu, 2011, 108 p.
201. **Kristjan Välk**. Gene expression profiling and genome-wide association studies of non-small cell lung cancer. Tartu, 2011, 98 p.
202. **Arno Põllumäe**. Spatio-temporal patterns of native and invasive zooplankton species under changing climate and eutrophication conditions. Tartu, 2011, 153 p.
203. **Egle Tammeleht**. Brown bear (*Ursus arctos*) population structure, demographic processes and variations in diet in northern Eurasia. Tartu, 2011, 143 p.
205. **Teele Jairus**. Species composition and host preference among ectomycorrhizal fungi in Australian and African ecosystems. Tartu, 2011, 106 p.
206. **Kessy Abarenkov**. PlutoF – cloud database and computing services supporting biological research. Tartu, 2011, 125 p.

207. **Marina Grigorova**. Fine-scale genetic variation of follicle-stimulating hormone beta-subunit coding gene (*FSHB*) and its association with reproductive health. Tartu, 2011, 184 p.
208. **Anu Tiitsaar**. The effects of predation risk and habitat history on butterfly communities. Tartu, 2011, 97 p.
209. **Elin Sild**. Oxidative defences in immunoecological context: validation and application of assays for nitric oxide production and oxidative burst in a wild passerine. Tartu, 2011, 105 p.
210. **Irja Saar**. The taxonomy and phylogeny of the genera *Cystoderma* and *Cystodermella* (Agaricales, Fungi). Tartu, 2012, 167 p.
211. **Pauli Saag**. Natural variation in plumage bacterial assemblages in two wild breeding passerines. Tartu, 2012, 113 p.
212. **Aleksei Lulla**. Alphaviral nonstructural protease and its polyprotein substrate: arrangements for the perfect marriage. Tartu, 2012, 143 p.
213. **Mari Järve**. Different genetic perspectives on human history in Europe and the Caucasus: the stories told by uniparental and autosomal markers. Tartu, 2012, 119 p.
214. **Ott Scheler**. The application of tmRNA as a marker molecule in bacterial diagnostics using microarray and biosensor technology. Tartu, 2012, 93 p.
215. **Anna Balikova**. Studies on the functions of tumor-associated mucin-like leukosialin (CD43) in human cancer cells. Tartu, 2012, 129 p.
216. **Triinu Kõressaar**. Improvement of PCR primer design for detection of prokaryotic species. Tartu, 2012, 83 p.
217. **Tuul Sepp**. Hematological health state indices of greenfinches: sources of individual variation and responses to immune system manipulation. Tartu, 2012, 117 p.
218. **Rya Ero**. Modifier view of the bacterial ribosome. Tartu, 2012, 146 p.
219. **Mohammad Bahram**. Biogeography of ectomycorrhizal fungi across different spatial scales. Tartu, 2012, 165 p.
220. **Annely Lorents**. Overcoming the plasma membrane barrier: uptake of amphipathic cell-penetrating peptides induces influx of calcium ions and downstream responses. Tartu, 2012, 113 p.
221. **Katrin Männik**. Exploring the genomics of cognitive impairment: whole-genome SNP genotyping experience in Estonian patients and general population. Tartu, 2012, 171 p.
222. **Marko Prous**. Taxonomy and phylogeny of the sawfly genus *Empria* (Hymenoptera, Tenthredinidae). Tartu, 2012, 192 p.
223. **Triinu Visnapuu**. Levansucrases encoded in the genome of *Pseudomonas syringae* pv. tomato DC3000: heterologous expression, biochemical characterization, mutational analysis and spectrum of polymerization products. Tartu, 2012, 160 p.
224. **Nele Tamberg**. Studies on Semliki Forest virus replication and pathogenesis. Tartu, 2012, 109 p.

225. **Tõnu Esko**. Novel applications of SNP array data in the analysis of the genetic structure of Europeans and in genetic association studies. Tartu, 2012, 149 p.
226. **Timo Arula**. Ecology of early life-history stages of herring *Clupea harengus membras* in the northeastern Baltic Sea. Tartu, 2012, 143 p.
227. **Inga Hiiesalu**. Belowground plant diversity and coexistence patterns in grassland ecosystems. Tartu, 2012, 130 p.
228. **Kadri Koorem**. The influence of abiotic and biotic factors on small-scale plant community patterns and regeneration in boreonemoral forest. Tartu, 2012, 114 p.
229. **Liis Andresen**. Regulation of virulence in plant-pathogenic pectobacteria. Tartu, 2012, 122 p.
230. **Kaupo Kohv**. The direct and indirect effects of management on boreal forest structure and field layer vegetation. Tartu, 2012, 124 p.
231. **Mart Jüssi**. Living on an edge: landlocked seals in changing climate. Tartu, 2012, 114 p.
232. **Riina Klais**. Phytoplankton trends in the Baltic Sea. Tartu, 2012, 136 p.
233. **Rauno Veeroja**. Effects of winter weather, population density and timing of reproduction on life-history traits and population dynamics of moose (*Alces alces*) in Estonia. Tartu, 2012, 92 p.
234. **Marju Keis**. Brown bear (*Ursus arctos*) phylogeography in northern Eurasia. Tartu, 2013, 142 p.
235. **Sergei Põlme**. Biogeography and ecology of *alnus*- associated ectomycorrhizal fungi – from regional to global scale. Tartu, 2013, 90 p.
236. **Liis Uusküla**. Placental gene expression in normal and complicated pregnancy. Tartu, 2013, 173 p.
237. **Marko Lõoke**. Studies on DNA replication initiation in *Saccharomyces cerevisiae.* Tartu, 2013, 112 p.
238. **Anne Aan**. Light- and nitrogen-use and biomass allocation along productivity gradients in multilayer plant communities. Tartu, 2013, 127 p.
239. **Heidi Tamm**. Comprehending phylogenetic diversity – case studies in three groups of ascomycetes. Tartu, 2013, 136 p.
240. **Liina Kangur**. High-Pressure Spectroscopy Study of Chromophore-Binding Hydrogen Bonds in Light-Harvesting Complexes of Photosynthetic Bacteria. Tartu, 2013, 150 p.
241. **Margus Leppik**. Substrate specificity of the multisite specific pseudouridine synthase RluD. Tartu, 2013, 111 p.
242. **Lauris Kaplinski**. The application of oligonucleotide hybridization model for PCR and microarray optimization. Tartu, 2013, 103 p.
243. **Merli Pärnoja**. Patterns of macrophyte distribution and productivity in coastal ecosystems: effect of abiotic and biotic forcing. Tartu, 2013, 155 p.
244. **Tõnu Margus**. Distribution and phylogeny of the bacterial translational GTPases and the Mqsr/YgiT regulatory system. Tartu, 2013, 126 p.
245. **Pille Mänd**. Light use capacity and carbon and nitrogen budget of plants: remote assessment and physiological determinants. Tartu, 2013, 128 p.

246. **Mario Plaas**. Animal model of Wolfram Syndrome in mice: behavioural, biochemical and psychopharmacological characterization. Tartu, 2013, 144 p.
247. **Georgi Hudjašov**. Maps of mitochondrial DNA, Y-chromosome and tyrosinase variation in Eurasian and Oceanian populations. Tartu, 2013, 115 p.
248. **Mari Lepik**. Plasticity to light in herbaceous plants and its importance for community structure and diversity. Tartu, 2013, 102 p.
249. **Ede Leppik**. Diversity of lichens in semi-natural habitats of Estonia. Tartu, 2013, 151 p.
250. **Ülle Saks**. Arbuscular mycorrhizal fungal diversity patterns in boreonemoral forest ecosystems. Tartu, 2013, 151 p.
251. **Eneli Oitmaa**. Development of arrayed primer extension microarray assays for molecular diagnostic applications. Tartu, 2013, 147 p.
252. **Jekaterina Jutkina**. The horizontal gene pool for aromatics degradation: bacterial catabolic plasmids of the Baltic Sea aquatic system. Tartu, 2013, 121 p.
253. **Helen Vellau**. Reaction norms for size and age at maturity in insects: rules and exceptions. Tartu, 2014, 132 p.
254. **Randel Kreitsberg**. Using biomarkers in assessment of environmental contamination in fish – new perspectives. Tartu, 2014, 107 p.
255. **Krista Takkis**. Changes in plant species richness and population performance in response to habitat loss and fragmentation.Tartu, 2014, 141 p.
256. **Liina Nagirnaja**. Global and fine-scale genetic determinants of recurrent pregnancy loss. Tartu, 2014, 211 p.
257. **Triin Triisberg**. Factors influencing the re-vegetation of abandoned extracted peatlands in Estonia. Tartu, 2014, 133 p.
258. **Villu Soon**. A phylogenetic revision of the *Chrysis ignita* species group (Hymenoptera: Chrysididae) with emphasis on the northern European fauna. Tartu, 2014, 211 p.
259. **Andrei Nikonov**. RNA-Dependent RNA Polymerase Activity as a Basis for the Detection of Positive-Strand RNA Viruses by Vertebrate Host Cells. Tartu, 2014, 207 p.
260. **Eele Õunapuu-Pikas**. Spatio-temporal variability of leaf hydraulic conductance in woody plants: ecophysiological consequences. Tartu, 2014, 135 p.
261. **Marju Männiste**. Physiological ecology of greenfinches: information content of feathers in relation to immune function and behavior. Tartu, 2014, 121 p.
262. **Katre Kets**. Effects of elevated concentrations of CO_2 and O_3 on leaf photosynthetic parameters in *Populus tremuloides*: diurnal, seasonal and interannual patterns. Tartu, 2014, 115 p.
263. **Külli Lokko**. Seasonal and spatial variability of zoopsammon communities in relation to environmental parameters. Tartu, 2014, 129 p.
264. **Olga Žilina**. Chromosomal microarray analysis as diagnostic tool: Estonian experience. Tartu, 2014, 152 p.

265. **Kertu Lõhmus**. Colonisation ecology of forest-dwelling vascular plants and the conservation value of rural manor parks. Tartu, 2014, 111 p.
266. **Anu Aun**. Mitochondria as integral modulators of cellular signaling. Tartu, 2014, 167 p.
267. **Chandana Basu Mallick**. Genetics of adaptive traits and gender-specific demographic processes in South Asian populations. Tartu, 2014, 160 p.
268. **Riin Tamme**. The relationship between small-scale environmental heterogeneity and plant species diversity. Tartu, 2014, 130 p.
269. **Liina Remm**. Impacts of forest drainage on biodiversity and habitat quality: implications for sustainable management and conservation. Tartu, 2015, 126 p.
270. **Tiina Talve**. Genetic diversity and taxonomy within the genus *Rhinanthus.* Tartu, 2015, 106 p.
271. **Mehis Rohtla**. Otolith sclerochronological studies on migrations, spawning habitat preferences and age of freshwater fishes inhabiting the Baltic Sea. Tartu, 2015, 137 p.
272. **Alexey Reshchikov**. The world fauna of the genus *Lathrolestes* (Hymenoptera, Ichneumonidae). Tartu, 2015, 247 p.
273. **Martin Pook**. Studies on artificial and extracellular matrix protein-rich surfaces as regulators of cell growth and differentiation. Tartu, 2015, 142 p.
274. **Mai Kukumägi**. Factors affecting soil respiration and its components in silver birch and Norway spruce stands. Tartu, 2015, 155 p.
275. **Helen Karu**. Development of ecosystems under human activity in the North-East Estonian industrial region: forests on post-mining sites and bogs. Tartu, 2015, 152 p.
276. **Hedi Peterson**. Exploiting high-throughput data for establishing relationships between genes. Tartu, 2015, 186 p.
277. **Priit Adler**. Analysis and visualisation of large scale microarray data, Tartu, 2015, 126 p.
278. **Aigar Niglas**. Effects of environmental factors on gas exchange in deciduous trees: focus on photosynthetic water-use efficiency. Tartu, 2015, 152 p.
279. **Silja Laht**. Classification and identification of conopeptides using profile hidden Markov models and position-specific scoring matrices. Tartu, 2015, 100 p.
280. **Martin Kesler**. Biological characteristics and restoration of Atlantic salmon *Salmo salar* populations in the Rivers of Northern Estonia. Tartu, 2015, 97 p.
281. **Pratyush Kumar Das**. Biochemical perspective on alphaviral nonstructural protein 2: a tale from multiple domains to enzymatic profiling. Tartu, 2015, 205 p
282. **Priit Palta**. Computational methods for DNA copy number detection. Tartu, 2015, 130 p.
283. **Julia Sidorenko**. Combating DNA damage and maintenance of genome integrity in pseudomonads. Tartu, 2015, 174 p.
284. **Anastasiia Kovtun-Kante**. Charophytes of Estonian inland and coastal waters: distribution and environmental preferences. Tartu, 2015, 97 p.

285. **Ly Lindman**. The ecology of protected butterfly species in Estonia. Tartu, 2015, 171 p.
286. **Jaanis Lodjak**. Association of Insulin-like Growth Factor I and Corticosterone with Nestling Growth and Fledging Success in Wild Passerines. Tartu, 2016, 113 p.
287. **Ann Kraut**. Conservation of Wood-Inhabiting Biodiversity – Semi-Natural Forests as an Opportunity. Tartu, 2016, 141 p.
288. **Tiit Örd**. Functions and regulation of the mammalian pseudokinase TRIB3. Tartu, 2016, 182. p.
289. **Kairi Käiro**. Biological Quality According to Macroinvertebrates in Streams of Estonia (Baltic Ecoregion of Europe): Effects of Human-induced Hydromorphological Changes. Tartu, 2016, 126 p.
290. **Leidi Laurimaa**. *Echinococcus multilocularis* and other zoonotic parasites in Estonian canids. Tartu, 2016, 144 p.
291. **Helerin Margus**. Characterization of cell-penetrating peptide/nucleic acid nanocomplexes and their cell-entry mechanisms. Tartu, 2016, 173 p.
292. **Kadri Runnel**. Fungal targets and tools for forest conservation. Tartu, 2016, 157 p.
293. **Urmo Võsa**. MicroRNAs in disease and health: aberrant regulation in lung cancer and association with genomic variation. Tartu, 2016, 163 p.
294. **Kristina Mäemets-Allas**. Studies on cell growth promoting AKT signaling pathway – a promising anti-cancer drug target. Tartu, 2016, 146 p.
295. **Janeli Viil**. Studies on cellular and molecular mechanisms that drive normal and regenerative processes in the liver and pathological processes in Dupuytren's contracture. Tartu, 2016, 175 p.
296. **Ene Kook**. Genetic diversity and evolution of *Pulmonaria angustifolia* L. and *Myosotis laxa sensu lato* (Boraginaceae). Tartu, 2016, 106 p.
297. **Kadri Peil**. RNA polymerase II-dependent transcription elongation in *Saccharomyces cerevisiae*. Tartu, 2016, 113 p.
298. **Katrin Ruisu**. The role of RIC8A in mouse development and its function in cell-matrix adhesion and actin cytoskeletal organisation. Tartu, 2016, 129 p.
299. **Janely Pae**. Translocation of cell-penetrating peptides across biological membranes and interactions with plasma membrane constituents. Tartu, 2016, 126 p.
300. **Argo Ronk**. Plant diversity patterns across Europe: observed and dark diversity. Tartu, 2016, 153 p.
301. **Kristiina Mark**. Diversification and species delimitation of lichenized fungi in selected groups of the family Parmeliaceae (Ascomycota). Tartu, 2016, 181 p.
302. **Jaak-Albert Metsoja**. Vegetation dynamics in floodplain meadows: influence of mowing and sediment application. Tartu, 2016, 140 p.
303. **Hedvig Tamman**. The GraTA toxin-antitoxin system of *Pseudomonas putida*: regulation and role in stress tolerance. Tartu, 2016, 154 p.

304. **Kadri Pärtel**. Application of ultrastructural and molecular data in the taxonomy of helotialean fungi. Tartu, 2016, 183 p.
305. **Maris Hindrikson**. Grey wolf (*Canis lupus*) populations in Estonia and Europe: genetic diversity, population structure and -processes, and hybridization between wolves and dogs. Tartu, 2016, 121 p.
306. **Polina Degtjarenko**. Impacts of alkaline dust pollution on biodiversity of plants and lichens: from communities to genetic diversity. Tartu, 2016, 126 p.
307. **Liina Pajusalu**. The effect of CO_2 enrichment on net photosynthesis of macrophytes in a brackish water environment. Tartu, 2016, 126 p.
308. **Stoyan Tankov**. Random walks in the stringent response. Tartu, 2016, 94 p.
309. **Liis Leitsalu**. Communicating genomic research results to population-based biobank participants. Tartu, 2016, 158 p.
310. **Richard Meitern**. Redox physiology of wild birds: validation and application of techniques for detecting oxidative stress. Tartu, 2016, 134 p.
311. **Kaie Lokk**. Comparative genome-wide DNA methylation studies of healthy human tissues and non-small cell lung cancer tissue. Tartu, 2016, 127 p.
312. **Mihhail Kurašin**. Processivity of cellulases and chitinases. Tartu, 2017, 132 p.
313. **Carmen Tali**. Scavenger receptors as a target for nucleic acid delivery with peptide vectors. Tartu, 2017, 155 p.
314. **Katarina Oganjan**. Distribution, feeding and habitat of benthic suspension feeders in a shallow coastal sea. Tartu, 2017, 132 p.
315. **Taavi Paal**. Immigration limitation of forest plants into wooded landscape corridors. Tartu, 2017, 145 p.
316. **Kadri Õunap**. The Williams-Beuren syndrome chromosome region protein WBSCR22 is a ribosome biogenesis factor. Tartu, 2017, 135 p.
317. **Riin Tamm**. In-depth analysis of factors affecting variability in thiopurine methyltransferase activity. Tartu, 2017, 170 p.
318. **Keiu Kask**. The role of RIC8A in the development and regulation of mouse nervous system. Tartu, 2017, 184 p.
319. **Tiia Möller**. Mapping and modelling of the spatial distribution of benthic macrovegetation in the NE Baltic Sea with a special focus on the eelgrass *Zostera marina* Linnaeus, 1753. Tartu, 2017, 162 p.
320. **Silva Kasela**. Genetic regulation of gene expression: detection of tissue- and cell type-specific effects. Tartu, 2017, 150 p.
321. **Karmen Süld**. Food habits, parasites and space use of the raccoon dog *Nyctereutes procyonoides*: the role of an alien species as a predator and vector of zoonotic diseases in Estonia. Tartu, 2017, p.
322. **Ragne Oja**. Consequences of supplementary feeding of wild boar – concern for ground-nesting birds and endoparasite infection. Tartu, 2017, 141 p.
323. **Riin Kont**. The acquisition of cellulose chain by a processive cellobiohydrolase. Tartu, 2017, 117 p.
324. **Liis Kasari**. Plant diversity of semi-natural grasslands: drivers, current status and conservation challenges. Tartu, 2017, 141 p.

325. **Sirgi Saar**. Belowground interactions: the roles of plant genetic relatedness, root exudation and soil legacies. Tartu, 2017, 113 p.
326. **Sten Anslan**. Molecular identification of Collembola and their fungal associates. Tartu, 2017, 125 p.
327. **Imre Taal**. Causes of variation in littoral fish communities of the Eastern Baltic Sea: from community structure to individual life histories. Tartu, 2017, 118 p.
328. **Jürgen Jalak**. Dissecting the Mechanism of Enzymatic Degradation of Cellulose Using Low Molecular Weight Model Substrates. Tartu, 2017, 137 p.
329. **Kairi Kiik**. Reproduction and behaviour of the endangered European mink (*Mustela lutreola*) in captivity. Tartu, 2018, 112 p.
330. **Ivan Kuprijanov**. Habitat use and trophic interactions of native and invasive predatory macroinvertebrates in the northern Baltic Sea. Tartu, 2018, 117 p.
331. **Hendrik Meister**. Evolutionary ecology of insect growth: from geographic patterns to biochemical trade-offs. Tartu, 2018, 147 p.
332. **Ilja Gaidutšik**. Irc3 is a mitochondrial branch migration enzyme in *Saccharomyces cerevisiae.* Tartu, 2018, 161 p.
333. **Lena Neuenkamp**. The dynamics of plant and arbuscular mycorrhizal fungal communities in grasslands under changing land use. Tartu, 2018, 241 p.
334. **Laura Kasak**. Genome structural variation modulating the placenta and pregnancy maintenance. Tartu, 2018, 181 p.
335. **Kersti Riibak**. Importance of dispersal limitation in determining dark diversity of plants across spatial scales. Tartu, 2018, 133 p.
336. **Liina Saar**. Dynamics of grassland plant diversity in changing landscapes. Tartu, 2018, 206 p.
337. **Hanna Ainelo**. Fis regulates *Pseudomonas putida* biofilm formation by controlling the expression of *lapA*. Tartu, 2018, 143 p.
338. **Natalia Pervjakova**. Genomic imprinting in complex traits. Tartu, 2018, 176 p.
339. **Andrio Lahesaare**. The role of global regulator Fis in regulating the expression of *lapF* and the hydrophobicity of soil bacterium *Pseudomonas putida*. Tartu, 2018, 124 p.
340. **Märt Roosaare**. *K*-mer based methods for the identification of bacteria and plasmids. Tartu, 2018, 117 p.
341. **Maria Abakumova**. The relationship between competitive behaviour and the frequency and identity of neighbours in temperate grassland plants. Tartu, 2018, 104 p.
342. **Margus Vilbas**. Biotic interactions affecting habitat use of myrmecophilous butterflies in Northern Europe. Tartu, 2018, 142 p.
343. **Liina Kinkar**. Global patterns of genetic diversity and phylogeography of *Echinococcus granulosus* sensu stricto – a tapeworm species of significant public health concern. Tartu, 2018, 147 p.

344. **Teivi Laurimäe**. Taxonomy and genetic diversity of zoonotic tapeworms in the species complex of *Echinococcus granulosus* sensu lato. Tartu, 2018, 143 p.
345. **Tatjana Jatsenko**. Role of translesion DNA polymerases in mutagenesis and DNA damage tolerance in Pseudomonads. Tartu, 2018, 216 p.
346. **Katrin Viigand**. Utilization of α-glucosidic sugars by *Ogataea* (*Hansenula*) *polymorpha*. Tartu, 2018, 148 p.
347. **Andres Ainelo**. Physiological effects of the *Pseudomonas putida* toxin grat. Tartu, 2018, 146 p.
348. **Killu Timm**. Effects of two genes (DRD4 and SERT) on great tit (*Parus major*) behaviour and reproductive traits. Tartu, 2018, 117 p.
349. **Petr Kohout**. Ecology of ericoid mycorrhizal fungi. Tartu, 2018, 184 p.
350. **Gristin Rohula-Okunev**. Effects of endogenous and environmental factors on night-time water flux in deciduous woody tree species. Tartu, 2018, 184 p.
351. **Jane Oja**. Temporal and spatial patterns of orchid mycorrhizal fungi in forest and grassland ecosystems. Tartu, 2018, 102 p.
352. **Janek Urvik**. Multidimensionality of aging in a long-lived seabird. Tartu, 2018, 135 p.
353. **Lisanna Schmidt**. Phenotypic and genetic differentiation in the hybridizing species pair *Carex flava* and *C. viridula* in geographically different regions. Tartu, 2018, 133 p.
354. **Monika Karmin**. Perspectives from human Y chromosome – phylogeny, population dynamics and founder events. Tartu, 2018, 168 p.
355. **Maris Alver**. Value of genomics for atherosclerotic cardiovascular disease risk prediction. Tartu, 2019, 148 p.
356. **Lehti Saag**. The prehistory of Estonia from a genetic perspective: new insights from ancient DNA. Tartu, 2019, 171 p.
357. **Mari-Liis Viljur**. Local and landscape effects on butterfly assemblages in managed forests. Tartu, 2019, 115 p.
358. **Ivan Kisly**. The pleiotropic functions of ribosomal proteins eL19 and eL24 in the budding yeast ribosome. Tartu, 2019, 170 p.
359. **Mikk Puustusmaa**. On the origin of papillomavirus proteins. Tartu, 2019, 152 p.
360. **Anneliis Peterson**. Benthic biodiversity in the north-eastern Baltic Sea: mapping methods, spatial patterns, and relations to environmental gradients. Tartu, 2019, 159 p.
361. **Erwan Pennarun**. Meandering along the mtDNA phylogeny; causerie and digression about what it can tell us about human migrations. Tartu, 2019, 162 p.
362. **Karin Ernits**. Levansucrase Lsc3 and endo-levanase BT1760: characterization and application for the synthesis of novel prebiotics. Tartu, 2019, 217 p.
363. **Sille Holm**. Comparative ecology of geometrid moths: in search of contrasts between a temperate and a tropical forest. Tartu, 2019, 135 p.

364. **Anne-Mai Ilumäe**. Genetic history of the Uralic-speaking peoples as seen through the paternal haplogroup N and autosomal variation of northern Eurasians. Tartu, 2019, 172 p.
365. **Anu Lepik**. Plant competitive behaviour: relationships with functional traits and soil processes. Tartu, 2019, 152 p.
366. **Kunter Tätte**. Towards an integrated view of escape decisions in birds under variable levels of predation risk. Tartu, 2020, 172 p.
367. **Kaarin Parts**. The impact of climate change on fine roots and root-associated microbial communities in birch and spruce forests. Tartu, 2020, 143 p.
368. **Viktorija Kukuškina**. Understanding the mechanisms of endometrial receptivity through integration of 'omics' data layers. Tartu, 2020, 169 p.
369. **Martti Vasar**. Developing a bioinformatics pipeline gDAT to analyse arbuscular mycorrhizal fungal communities using sequence data from different marker regions. Tartu, 2020, 193 p.
370. **Ott Kangur**. Nocturnal water relations and predawn water potential disequilibrium in temperate deciduous tree species. Tartu, 2020, 126 p.
371. **Helen Post**. Overview of the phylogeny and phylogeography of the Y-chromosomal haplogroup N in northern Eurasia and case studies of two linguistically exceptional populations of Europe – Hungarians and Kalmyks. Tartu, 2020, 143 p.
372. **Kristi Krebs**. Exploring the genetics of adverse events in pharmacotherapy using Biobanks and Electronic Health Records. Tartu, 2020, 151 p.
373. **Kärt Ukkivi**. Mutagenic effect of transcription and transcription-coupled repair factors in *Pseudomonas putida.* Tartu, 2020, 154 p.
374. **Elin Soomets**. Focal species in wetland restoration. Tartu, 2020, 137 p.
375. **Kadi Tilk**. Signals and responses of ColRS two-component system in *Pseudomonas putida.* Tartu, 2020, 133 p.
376. **Indrek Teino**. Studies on aryl hydrocarbon receptor in the mouse granulosa cell model. Tartu, 2020, 139 p.
377. **Maarja Vaikre**. The impact of forest drainage on macroinvertebrates and amphibians in small waterbodies and opportunities for cost-effective mitigation. Tartu, 2020, 132 p.
378. **Siim-Kaarel Sepp**. Soil eukaryotic community responses to land use and host identity. Tartu, 2020, 222 p.
379. **Eveli Otsing**. Tree species effects on fungal richness and community structure. Tartu, 2020, 152 p.
380. **Mari Pent**. Bacterial communities associated with fungal fruitbodies. Tartu, 2020, 144 p.
381. **Einar Kärgenberg**. Movement patterns of lithophilous migratory fish in free-flowing and fragmented rivers. Tartu, 2020, 167 p.
382. **Antti Matvere**. The studies on aryl hydrocarbon receptor in murine granulosa cells and human embryonic stem cells. Tartu, 2021, 163 p.
383. **Jhonny Capichoni Massante**. Phylogenetic structure of plant communities along environmental gradients: a macroecological and evolutionary approach. Tartu, 2021, 144 p.

384. **Ajai Kumar Pathak.** Delineating genetic ancestries of people of the Indus Valley, Parsis, Indian Jews and Tharu tribe. Tartu, 2021, 197 p.
385. **Tanel Vahter.** Arbuscular mycorrhizal fungal biodiversity for sustainable agroecosystems. Tartu, 2021, 191 p.
386. **Burak Yelmen.** Characterization of ancient Eurasian influences within modern human genomes. Tartu, 2021, 134 p.