DISSERTATIONES
BIOLOGICAE
UNIVERSITATIS
TARTUENSIS
421

MATHILDE FRÉDÉRIQUE E. ANDRÉ

New Guinea, a hotspot for Human evolution: settlement history and adaptation in northern Sahul

DISSERTATIONES BIOLOGICAE UNIVERSITATIS TARTUENSIS
421

DISSERTATIONES BIOLOGICAE UNIVERSITATIS TARTUENSIS
421

MATHILDE FRÉDÉRIQUE E. ANDRÉ

New Guinea, a hotspot for Human evolution: settlement history and adaptation in northern Sahul

UNIVERSITY OF TARTU
Press
1632

Institute of Genomics and the Institute of Molecular and Cell Biology, University of Tartu, Estonia

Dissertation was accepted for the commencement of the degree of Doctor of Philosophy in Gene Technology on the 7th of July, 2023, by the Council of the Institute of Molecular and Cell Biology, Faculty of Science and Technology, University of Tartu.

Supervisors: Mayukh Mondal, PhD; Research Fellow of Evolutionary Genetics, Centre for Genomics, Evolution & Medicine Institute of Genomics, University of Tartu, Estonia; Researcher, Institute of Clinical Molecular Biology, Christian-Albrechts-Universität zu Kiel, Germany

François-Xavier Ricaut, PhD; Senior Researcher, National Center for Scientific Research (CNRS), Department of Evolution and Biological Diversity, Université de Toulouse, France.

Opponent: Ian Mathieson, PhD; Associate Professor of Genetics, Department of Genetics, University of Pennsylvania, United States

Commencement: Room No 105, 23B Riia St, Tartu, on 12th of September, 2023, at 15:15

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

This research was funded by the European Union´s Horizon 2020 research and innovation programme under grant agreement No 810645 and by the European Union through the European Regional Development Fund Grant MOBEC008

European Union
European Regional
Development Fund

Investing
in your future

ISSN 1024-6479 (print)
ISBN 978-9916-27-284-8 (print)
ISSN 2806-2140 (pdf)
ISBN 978-9916-27-285-5 (pdf)

University of Tartu Press
www.tyk.ee

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES

Figures

Tables

LIST OF ORIGINAL PUBLICATIONS

I. Nicolas Brucato, **Mathilde André**, Roxanne Tsang, Lauri Saag, Jason Kariwiga, Kylie Sesuki, Teppsy Beni, William Pomat, John Muke, Vincent Meyer, Anne Boland, Jean-François Deleuze, Herawati Sudoyo, Mayukh Mondal, Luca Pagani, Irene Gallego Romero, Mait Metspalu, Murray P Cox, Matthew Leavesley, François-Xavier Ricaut, Papua New Guinean Genomes Reveal the Complex Settlement of North Sahul, *Molecular Biology and Evolution*, Volume 38, Issue 11, October 2021, https://doi.org/10.1093/molbev/msab238

II. Nicolas Brucato, **Mathilde André**, Georgi Hudjashov, Mayukh Mondal, Murray P.Cox, Matthew Leavesley, François-Xavier Ricaut, Chronology of natural selection in Oceanian genomes, *IScience*, Volume 25, Issue 7, June 2022, https://doi.org/10.1016/j.isci.2022.104583

III. **Mathilde André**, Nicolas Brucato, Sébastien Plutniak, Jason Kariwiga, John Muke, Adeline Morez, Matthew Leavesley, Mayukh Mondal, François-Xavier Ricaut, Phenotypic differences between highlanders and lowlanders in Papua New Guinea, *PLOS ONE*, Volume 16, Issue 7, July 2021, https://doi.org/10.1371/journal.pone.0253921

IV. **Mathilde André**, Nicolas Brucato, Georgi Hudjasov, Vasili Pankratov, Danat Yermakovich, Rita Kreevan, Jason Kariwiga, John Muke, Anne Boland, Jean-François Deleuze, Vincent Meyer, Nicholas Evans, Murray P. Cox, Matthew Leavesley, Michael Dannemann, Tõnis Org, Mait Metspalu, Mayukh Mondal, François-Xavier Ricaut, Positive selection in the genomes of two Papua New Guinean populations at distinct altitude levels, *bioRxiv,* [preprint], December 2022, https://doi.org/10.1101/2022.12.15.520226

AUTHOR'S CONTRIBUTION TO THE LISTED ARTICLES

The author's contributions to the listed articles are as follows:

Ref I: I participated in the interpretation of the results and provided input for writing the paper.

Ref II: I participated in the selection dating analysis, the interpretation of the results and provided input for writing the paper..

Ref III: I curated and analysed phenotype measurements, generated figures, interpreted the results and led the manuscript writing team.

Ref IV: I generated selection scans, reconstructed allele frequencies, analysed association studies data and investigated candidate haplotypes for evidence of archaic introgression. In addition, I generated figures, interpreted the results and led the manuscript writing team.

ABBREVIATIONS

a.s.l.	above sea level
AMH	anatomically modern human
ARG	ancestral recombination graph
BMI	body mass index
bp	base pair
CI	confidence interval
cM	centiMorgan
CMS	chronic mountain sickness
DAF	derived allele frequency
DARC	Duffy antigen receptor for chemokines
EHH	Extended Haplotype Homozygosity
FEV1	forced expiratory volume in 1 second
F_{ST}	fixation index
FVC	forced vital capacity
GBP	guanylate-binding protein
GEMMA	Genome-wide Efficient Mixed Model Association
GEVA	Genealogical Estimation of Variant Age
GTEx	Genotype-Tissue Expression
GWAS	genome-wide association studies
HIF	Hypoxia-Inducible Factor
HLA	Human Leukocyte Antigen
HMM	Hidden Markov Model
iHS	integrated haplotype score
IQR	interquartile range
kya	thousand of years ago
LD	linkage disequilibrium
LMM	Linear Mixed Model
LS	Li and Stephen
MCMC	Markov Chain Monte Carlo
MHC	Major Histocompatibility Complex
MRCA	most recent common ancestor
nSL	number of segregating sites by length
PBS	Population Branch Statistic
PCA	Principal Component Analysis

PEF	peak expiratory flow
PNG	Papua New Guinea(n)
SGDP	Simons Genome Diversity Project
SNP	single nucleotide polymorphism
TMRCA	time to the most recent common ancestor
VEP	Ensembl Variant Effect Predictor
WHO	World Health Organization
XP-EHH	cross-population extended haplotype homozygosity
XP-nSL	cross-population number of segregating sites by length
YBP	Years Before Present

1. INTRODUCTION

1.1. New Guinea: exploring the multi-layered diversity

New Guinea is the world's second-largest island. It is separated from Australia by the Arafura Sea and is shared between two countries. Indonesia governs the western half of the island, and Papua New Guinea (PNG), an independent country since 1975, constitutes the Eastern half of the island. The island of New Guinea has around 14 million inhabitants, with 4 million living in the Indonesian provinces in western Papua and 10 million in PNG. New Guineans and Papua New Guineans are used consistently throughout this thesis to name the inhabitants of the island of New Guinea or the country of Papua New Guinea, respectively. The term "Papuan" is exclusively used to describe the ancestry of the indigenous people of New Guinea, in contrast with the Austronesian ancestry or other ancestries brought to New Guinea by recent migrations.

New Guinea is known for its early settlement by anatomically modern humans (AMH) and unique bio-cultural diversity. After AMH left Africa around 100 thousand years ago (kya), they arrived in the ancient continent of Sahul, which included Australia, Tasmania and the current island of New Guinea, as early as 50 kya (O'Connell et al., 2018). The journey to Sahul must have been challenging as AMH faced various unfamiliar environments. After the initial settlement, New Guinea remained relatively isolated from outside influence until the Austronesian expansion 3 kya. Communication between the different regions of the island was limited for most of its history. This relative isolation has led to the unique cultural and biological diversity that arose in New Guinea over time. New Guinea's bio-cultural diversity makes it of great interest to many disciplines. For instance, New Guinea is the most linguistically diverse territory in the world, with more than 800 spoken languages (Palmer, 2017), making it a major area of study for linguistics. Alongside their linguistic diversity, New Guinean groups also exhibit strong genetic variation from one another (Bergström et al., 2017). New Guinean genomes stand out even more by harbouring one of the highest amounts of Denisovan ancestry in their genomes (Larena et al., 2021; Reich et al., 2010).

New Guinean territory is characterised by diverse environments, from the coastal lowlands to the highlands, which host some of the highest mountains in Oceania (4,884 meters above sea level (a.s.l.)). New Guinean highlanders and lowlanders followed different cultural trajectories for thousands of years. New Guinea highland territories located between 1,600 and 2,400 m a.s. l. have been permanently inhabited for 20 kya and are currently the most densely populated area of the country (Brookfield & Allen, 1989; Müller, Bockarie, Alpers, & Smith, 2003). The partial pressure of inspired oxygen decreases with increasing altitude, which results in hypoxia (i.e., reduced oxygen availability to the body). Various detrimental conditions, such as acute mountain sickness, can occur after exposure to hypoxia from elevation as low as 1500 m a.s.l. (Barry & Pollard, 2003). Hence,

New Guinean highlanders have been experiencing hypoxia and its detrimental effects for thousands of years. Some populations, like the Andean Calchaquíes who have lived at intermediate altitude (above 1,500 m a.s.l.), and the Tibetan population who have lived at high-altitude (above 2,500 m a.s.l.) for thousands of years, have adapted to hypoxia and carry genomic signatures of selection to this stressor (Cynthia M. Beall, 2014; Eichstaedt et al., 2015). In contrast, the New Guinean highlands, despite being the most densely populated areas of the island since the Holocene (11 kya) (Brookfield & Allen, 1989; Müller et al., 2003; Trájer, Sebestyén, & Domokos, 2020), have been overlooked in terms of adaptation to hypoxia as these other populations.

Pathogenic pressure is another strong environmental challenge in New Guinea, with pathogens infection being the leading cause of death among people of Papua New Guinea (PNGs) (GBD 2013 Mortality and Causes of Death Collaborators, 2015; Kitur, Adair, Riley, & Lopez, 2019; Naraqi, Feling, & Leeder, 2003). Differences in pathogenic pressure between lowlanders and highlanders could explain why New Guinean highlands are more densely populated (Riley, 1983; Trájer, 2022; Trájer et al., 2020). For example, Malaria is highly prevalent in the lowlands while being rare in the highlands (Müller et al., 2003; Senn et al., 2010).

Population genetics has come a long way since the pioneering work of Cavalli-Sforza and colleagues, who provided the first insights into human populations' genetic diversity and structure (Cavalli-Sforza, Piazza, Menozzi, & Mountain, 1988). With the advent of affordable whole genome sequencing technologies, we have now entered the era of population genomics (Jobling, Hollox, Hurles, Kivisild, & Tyler-Smith, 2013; Pareek, Smoczynski, & Tretyn, 2011). The effort in large-scale sequencing of multiple populations worldwide (e.g., 1000 Genomes Project (The 1000 Genomes Project Consortium et al., 2015), Simons Genome Diversity Project (SGDP) (Mallick et al., 2016), Estonian Biocentre Human Genome Diversity Panel (Pagani et al., 2016)) has further improved our understanding of human genetic diversity. New sequencing tools have also enabled the generation of genomes from ancient human populations (Orlando et al., 2021) and even from other hominin species like Neanderthal (Green et al., 2010; Prüfer et al., 2014) and Denisova (Reich et al., 2010). In the case of the Denisovans, there were so few remains that their existence was identified only through the analysis of ancient DNA (Krause et al., 2010). Together advancements in modern and ancient genomics have widened our understanding of the evolution of human genome diversity, both spatially and temporally. Furthermore, the recent recognition of Svante Pääbo's work with the 2022 Nobel Prize in Physiology or Medicine has brought worldwide attention to evolutionary genomics, highlighting its potential to illuminate the interplay between evolutionary history and human health ("The Nobel Prize in Physiology or Medicine 2022", 2022). Evolutionary genomics remains a hot topic that promises to advance our understanding of the biological mechanisms underlying human evolution and health.

Whole genome sequencing is particularly valuable for studying New Guineans' unique genetic diversity. Indeed, currently available single nucleotide polymorphism (SNP) array designs are missing important genetic components exclusively

found in New Guineans and then only provide a partial and biased picture of their genetic diversity. By analyzing whole genomes, we can better understand the evolutionary forces that have shaped this diversity and uncover unique signatures of selection that may be relevant to human health and disease. This thesis presents new whole genome sequences from PNG highlands and lowlands as well as measurements of diverse phenotypes in these two populations. The first reference aimed to describe the models for northern Sahul settlement based on genomic data (Ref I). The second reference investigated signatures of selection unique to New Guineans witnessing the selection pressure their ancestors faced immediately after settling in new environments when reaching Sahul (Ref II). The last two references explored phenotype differences (Ref III) and selection signatures (Ref IV) unique to PNG highlanders and lowlanders to detect if PNG populations might display signals of local adaptation to their specific environment.

2. LITERATURE OVERVIEW

New Guinea is known for its early settlement by AMH and unique bio-cultural diversity. New Guineans are among the most genetically diverse populations and have lived in challenging environments for tens of thousands of years. Furthermore, their genomes harbour some of the highest levels of Denisovan DNA, an extinct hominin species. For all these reasons, New Guineans are a major interest in human evolution and adaptation studies.

The following literature review provides an overview of the complex settlement of New Guinea and the different migratory waves that participated in the diversity of New Guineans. It also describes the key concepts in studying selection in human genomes and the methods used to detect and fine-map selection.

2.1. New Guinea: a unique settlement history and diversity

2.1.1. Crossing the sea: early human migration to New Guinea

AMH appeared in Africa around 300 kya (Hublin et al., 2017), where they remained confined until 70 kya (Malaspinas et al., 2016; Mallick et al., 2016; Mondal et al., 2016). However, the time AMH expanded out of Africa remains debated, as some studies suggest an early out-of-Africa event around 100 kya (Grove et al., 2015; Kuhlwilm et al., 2016; Pagani et al., 2016).

When AMH left Africa, they began to spread across Eurasia and Southwest Asia. At this time, the sea level was lower, so some territories now separated by the sea were then part of the same landmass. There were two main continents in the South Pacific region: Sunda, which connected many current southeast Asian islands and mainland Asia; and Sahul, which included mainland Australia, Tasmania, and New Guinea. These two ancient continents were separated by Wallacea, many small islands separated by deep-water crossings that limited migrations between Sahul and Sunda (Wallace 1869) and created two distinct biogeographic areas (Figure 1) (Cámara-Leret et al., 2020). Despite this complex landscape, archaeological artefacts, ecological models and mitochondrial DNA indicate that the first settlers reached northern Sahul between 50 and 65 kya (Bradshaw et al., 2021; Clarkson et al., 2017; O'Connell & Allen, 2015; Pedro et al., 2020; Summerhayes et al., 2010). It has been suggested that New Guineans may even carry genomic traces from an early expansion of AMH out of Africa (Pagani et al., 2016). Nevertheless, the exact way how AMH crossed Wallacea and reached Sahul remains a topic of ongoing research and debate among scientists (Allen & O'Connell, 2020). Our current understanding relies on migration models that use paleogeological reconstructions of coastlines, environmental information or archaeological artefacts.

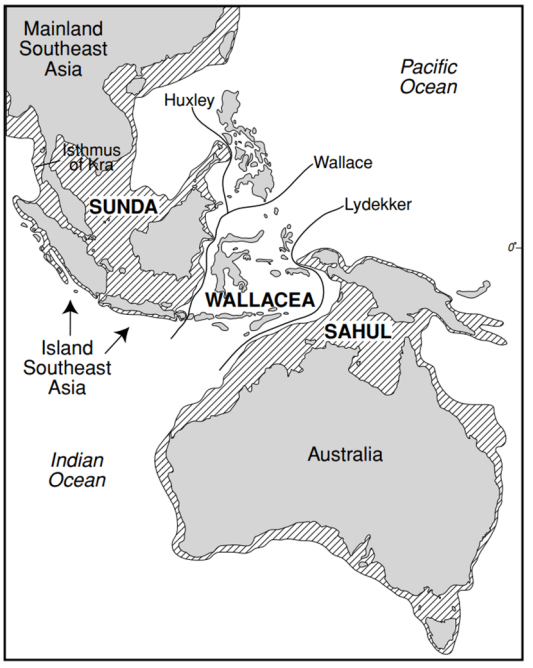

Figure 1: **Map of Australasia showing the extent of the Sunda and Sahul continental landmasses with Wallacea in between.** The contemporary geography of the region is shown in grey. Reprinted from Figure 1 (Harrison, Krigbaum, & Manser, 2006), with permission from Springer Nature. Copyright © 2006, Springer Science Business Media, LLC

2.1.1.1. An accident or a deliberate effort?

There is debate over whether the crossing from Sunda to Sahul was intentional or accidental. Some studies have suggested that early humans accidentally reached Sahul using natural rafts of vegetation following currents (Bird et al., 2019). However, with a minimal founding population size of around 1,300 individuals, the accident hypothesis is unlikely (Allen & O'Connell, 2020; Bird et al., 2019; Bradshaw et al., 2019). The substantial size of the founding groups also suggests that they had cohesion and cognitive and technological abilities to plan a journey, supporting the idea of an intentional and directed migration (Bradshaw et al., 2019).

2.1.1.2. The northern vs the southern route

Because Sunda and Sahul have never been connected by land, AMH must have used seafaring from one Wallacean island to another to reach Sahul. Two pathways were possible: the northern route through Sulawesi into New Guinea and the southern route through Bali and Timor into northwestern Australia (Smith, 2001). Different modelling approaches, using crossing distance as a proxy for crossing difficulty (Kealy, Louys, & O'Connor, 2018) and coastal viewshed analysis (i.e., identifying locations that are visible from the observer's point of

view) (Bird et al., 2019; Norman et al., 2018), suggest that the northern path was the easiest and quickest route, while the southern path was longer and required travelling without having the destination in sight (Bird et al., 2019). However, older archaeological sites have been found along the southern route through the Arafura plain (around 46 kya)) (Hawkins et al., 2017; Sutikna et al., 2018) (or even earlier, 65 kya in northern Australia (Clarkson et al., 2017)) than on the northern route (around 35 kya) (Latinis & Stark, 2005; O'Connor, Louys, Kealy, & Samper Carro, 2017). Nevertheless, these hypotheses cannot rely solely on archaeological evidence as the sea level has risen substantially since the first settlement, and many early sites may now be underwater (Summerhayes, Field, Shaw, & Gaffney, 2017). Finding older sites on the south path when models favour the northern arrival might suggest that first settlers rapidly spread along the coastal areas or a lack of archaeological investigation in western New Guinea (Bird et al., 2019).

Because the journey was intentional and repeatedly taken by several groups, it is also likely that the different groups used different routes during their trip to Sahul (Bird et al., 2019; Bradshaw et al., 2019). More recent models support multiple entry points, with an initial entry through the southern road into the north of modern Australia and a latter entrance via the northern route into New Guinea (Bradshaw et al., 2021). These groups of settlers may have already been genetically distinct (Bird et al., 2019). Indeed, the arrival of a unique human group is unlikely to explain the deep genetic diversity still observed in this region (Allen & O'Connell, 2020). However, though most of the studies on the Indigenous Australian and New Guinean Y chromosomes suggest a split around the time of the initial settlement of Sahul (Bergström et al., 2016; Tobler et al., 2017), there is also a suggestion of more recent connections of paternal lineages from these regions (Karmin et al., 2022).

2.1.2. Settlement within mainland New Guinea

Since its settlement by AMH 50 kya (O'Connell et al., 2018; Summerhayes et al., 2010), New Guinea has remained relatively isolated from outside influences (Terrell, 2004). This isolation, coupled with the diverse landscape found in New Guinea (Pawley & Australian National University, 2005; Summerhayes et al., 2017), has resulted in exceptional cultural and genetic diversity on the island.

2.1.2.1. Facing multiple new environments

The initial settlers of northern Sahul spread across the continent within a few thousand years (Bergström et al., 2016), encountering a diverse range of environments, including savannah grasslands, tropical rainforests, and montane landscapes.

After they entered Sahul, the first settlers rapidly accessed the mountain valleys. Indeed, one of the oldest sites in northern Sahul is located around 2000 m a.s.l. (Summerhayes et al., 2010). Coastal small hunter-gatherer groups were probably

penetrating the higher valleys as temporary foraging camps to access different resources than in the lowland. By the late Pleistocene (around 25kya), some groups shifted to a semi-permanent settlement in the highlands, probably thanks to the ameliorating climate (Gaffney, Ford, & Summerhayes, 2015). New Guinean highlanders have been residing in the highlands for at least 20,000 years (Haas et al., 2017; Madsen et al., 2017). The speed of their establishment suggests a rapid adaptation to the local flora and fauna for sustenance (Summerhayes et al., 2017).

Currently, the PNG highlands are the most densely populated area of the country, comprising approximately 3 million individuals, nearly 40% of the total population (Bourke & Allen, 2021). Many highlander groups have inhabited elevations between 1,600 and 2,400 m a.s.l. since the Holocene (Brookfield & Allen, 1989; Müller et al., 2003; Trájer et al., 2020). One reason for the uneven population distribution might be that intense pathogenic pressure in New Guinean lowlands that prompted the first inhabitant of New Guinea to access the highlands (Müller et al., 2003; Riley, 1983; Trájer, 2022; Trájer et al., 2020).

Since the late Pleistocene, the limited interaction between highland and lowland populations has facilitated the development of specific regional cultures. For instance, local plant cultivation practices appeared independently in the highlands around 10–7 kya (Denham, 2003; Golson, 2017) and have been associated with the demographic growth in that region (Bergström et al., 2017). In contrast with the highlands, the population size in the lowlands remained stable during the Holocene (Bergström et al., 2017), and agricultural practices were brought to the lowlands through the Austronesian influence around 3 kya (Bellwood et al., 2007; Skoglund et al., 2016).

2.1.2.2. From environment diversity to genetic diversity

During the initial settlement of Sahul, Indigenous Australians and New Guineans were probably not yet distinct populations and may have only diverged around 30 kya (Bergström et al., 2016; Karmin et al., 2022; Malaspinas et al., 2016; Tobler et al., 2017). After the rapid peopling of Sahul, New Guineans remained genetically isolated from other populations leading to their current unique genetic makeup (Bergström et al., 2020, 2016; Malaspinas et al., 2016). New Guinean populations also show a striking genetic heterogeneity among themselves (Bergström et al., 2020). The genetic differentiation between New Guinean groups, as measured by the fixation index (F_{ST}) between groups, is bigger than that of European or Asian groups separated by equivalent geographic distances. The high differentiation between New Guinean groups would support a limited dispersal and admixture of New Guinean populations over time (Bergström et al., 2017).

The first split between New Guinean populations is between New Guinea mainland inhabitants and Bismarck and Solomon islanders, shortly after northern Sahul's initial settlement (Bergström et al., 2017; Choin et al., 2021). The most diverging groups inside mainland New Guinea are lowlanders and highlanders. This genetic structure mostly lies in the isolation of these two groups during the Late Pleistocene, which also resulted in the different cultural practices mentioned above.

2.1.3. The Austronesian expansion

During the mid-Holocene (around 3 kya), Austronesians from Southeast Asia mixed with some Papuan populations, adding another layer of complexity to the New Guinean genetic diversity.

The Austronesian expansion, associated with the Lapita culture and Austronesian languages, started 5 kya from present-day Taiwan and reached the last inhabited territories in the Pacific Ocean (Bellwood et al., 2007; Patrick Vinton Kirch, 2002). This migration wave resulted in the modern Pacific population having a mixed Papuan and Austronesian ancestry (Kayser, 2010; Lipson et al., 2018).

The first wave of Austronesian-speakers from Taiwan to remote Oceania left minimal admixture trace within the inhabitants of Papua New Guinea (Figure 2B) (Bellwood et al., 2007; Hung & Carson, 2014; Skoglund et al., 2016). Another episode of gene flow with Papuans occurred around 3000 years ago in the Bismarck archipelago where the Austronesian-related ancestry was carried by the first settlers of Remote Oceania (P. V. Kirch & Hunt, 1988; Peter Bellwood & Eusebio Dizon, 2005; Skoglund et al., 2016) (Figure 2C). The secondary expansion to remote Oceania 2.3 to 2.9 kya included Papuan-Austronesian admixed populations (Lipson et al., 2018; Skoglund et al., 2016) (Figure 2D).

As a result of Austronesian admixture, some lowlanders in coastal New Guinea exhibit varying degrees of Southeast Asian genetic markers. Austronesian admixture in PNG is particularly important among Austronesian-speaking populations compared to non-Austronesian speakers (Bergström et al., 2017).

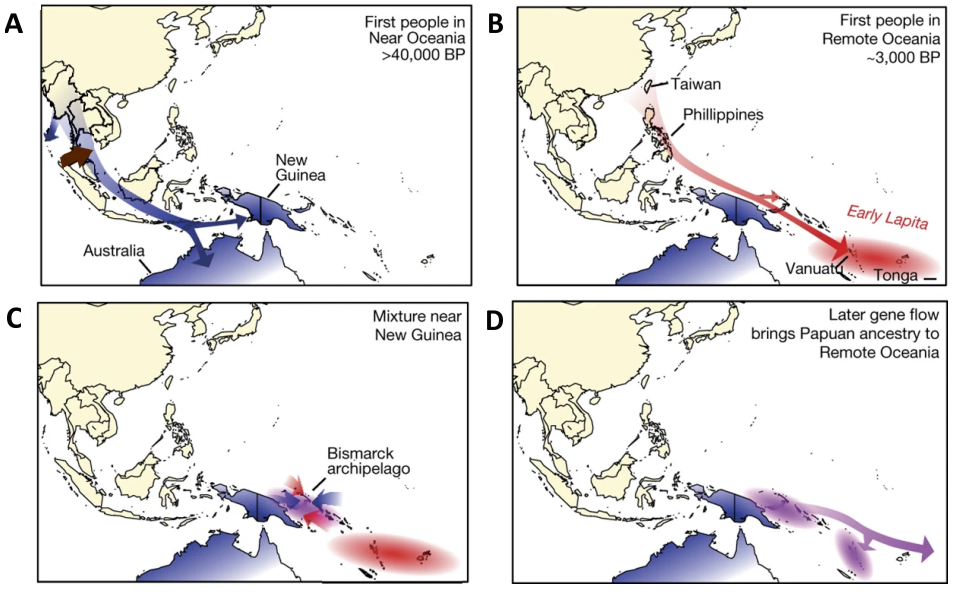

Figure 2: A, A model of population movements more than 40,000 years ago in which modern humans arrived in the Australia–New Guinea region (blue shading) and mixed with archaic Denisovans (brown arrow). **B**, A model of events before 3,000 years ago, in which the First Remote Oceanian population formed by the spread of a population of ultimate East Asian origin to a region including Vanuatu and Tonga, and experienced little or no mixture with the Papuans they encountered along the journey (red shading). Note that geographic routes are speculative. **C**, A model of populations of mixed Papuan–First Remote Oceanian ancestry in Near Oceania less than 3,000 years ago in a patchwork of islands with different proportions of First Remote Oceanian ancestry (pink shading). **D**, A model of secondary expansion of admixed populations bringing Papuan ancestry into Remote Oceania, which was still not complete in Tonga by the date of the Talasiu individual at 2,680–2,340 BP. Modified from Figure 3 (Skoglund et al., 2016) with permission from Springer Nature.

2.1.4. Interbreeding with archaic hominins

While AMH were still constrained to Africa, other hominin populations, such as Neanderthal and Denisova, were already present in Eurasia and East and South Asia, respectively. Both of these archaic hominins diverged from AMH around 520 kya and from each other around 390 kya (Meyer et al., 2016; Prüfer et al., 2017). When AMH left Africa, they met and admixed with other hominin species like Neanderthal (Green et al., 2010; Prüfer et al., 2014), Denisova (Reich et al., 2010) and potentially still unknown other archaic populations (Hammer, Woerner, Mendez, Watkins, & Wall, 2011; Mondal, Bertranpetit, & Lao, 2019; Mondal et al., 2016) (Figure 3). The remains of these admixture events can still be found in the genomes of current populations, including New Guineans. The admixture could have played a role in accelerating AMH adaptation to environments new to them but already inhabited by these archaic hominins for thousands of years (Dannemann & Kelso, 2017; Reilly, Tjahjadi, Miller, Akey, & Tucci, 2022; Sankararaman, Mallick, Patterson, & Reich, 2016; Vernot & Akey, 2014).

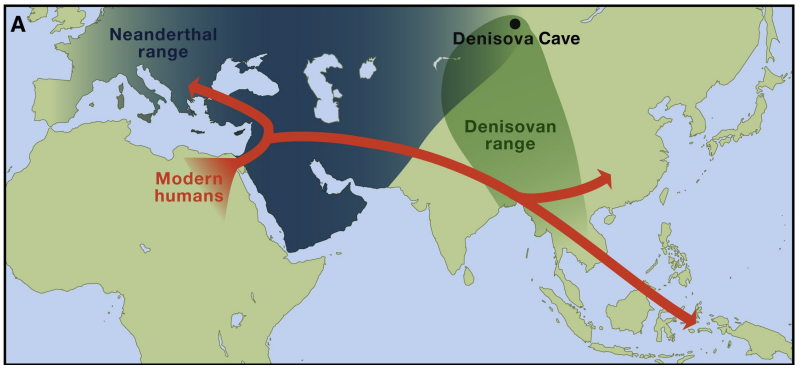

Figure 3: Insights gleaned from Studying Archaic Admixture (**A**) Schematic illustration of the out-of-Africa dispersal of modern humans (red). Blue and green shading roughly indicates the regions inhabited by Neanderthals and Denisovans, respectively. Although fossil evidence of Denisovans has been found only at Denisova Cave, the observed pattern of Denisovan admixture suggests they had a wide range that extended southeast. Reprinted from Figure 1.A (Vattathil & Akey, 2015) with permission from Elsevier. Copyright © 2015 Elsevier Inc. All rights reserved.

2.1.4.1. Neanderthal introgression in New Guineans

Non-African genomes carry around 2% of variants introgressed from Neanderthal (Bergström et al., 2020; Sankararaman et al., 2014). East Asians show a higher rate than Europeans (Figure 4), possibly due to a second pulse of admixture unique to the east Asian population (Vernot et al., 2016). Nonetheless, this enrichment of Neanderthal introgressed variants in East Asians remains relatively modest (Chen, Wolf, Fu, Li, & Akey, 2020). New Guineans would harbour a similar amount of Neanderthal ancestry than other non-African populations (around 2.2 to 2.9%) (Choin et al., 2021).

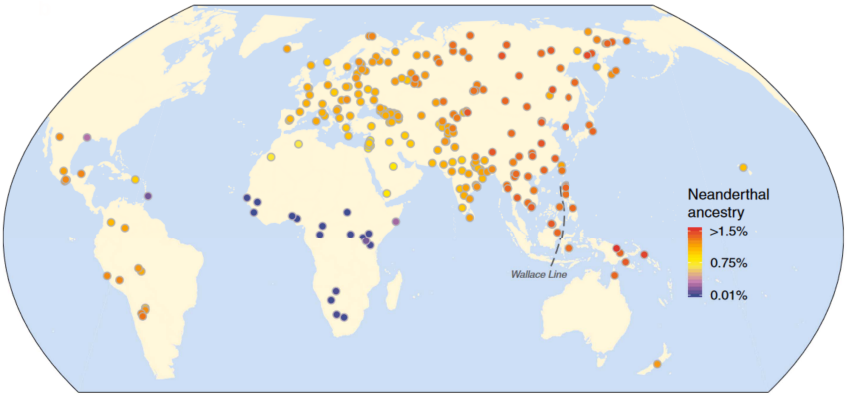

Figure 4: Average proportions of Neanderthal introgression in modern humans. Proportions of >1.5% are presented as 1.5% for visualization. Reprinted from Figure 3.b (Yuan et al., 2021), licensed under Creative Commons CC BY license. Copyright © 2021, Yuan et al.

2.1.4.2. Denisovan introgression in New Guineans

Only populations with Asian or Oceanian ancestry carry Denisovan introgression. New Guineans, along with Indigenous Australians and Philippine Ayta, stand out by carrying one of the highest Denisovan introgression rates (Larena et al., 2021; Reich et al., 2010) (Figure 5). Denisovan introgression in the region is complex because AMH have admixed independently with different Denisovan lineages (Jacobs et al., 2019). Previous studies have suggested that the current Denisovan ancestry within New Guineans results from at least two independent admixture pulses with different Denisovan lineages (Browning, Browning, Zhou, Tucci, & Akey, 2018; Choin et al., 2021; Jacobs et al., 2019). One of these admixture events likely happened before the separation of Near Oceanian and East Asian populations. The other would be more recent and private to New Guineans (Choin et al., 2021; Jacobs et al., 2019).

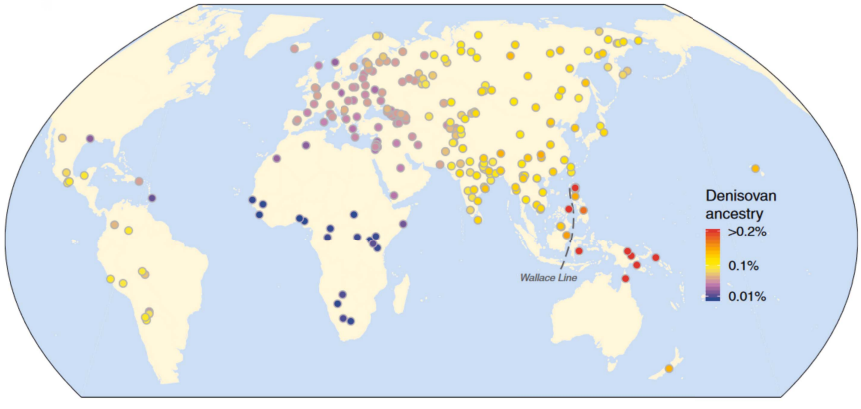

Figure 5: **Average proportions of Denisovan introgression in modern humans.** Proportions of >0.2% are presented as 0.2% for visualization. Reprinted from Figure 3.a (Yuan et al., 2021), licensed under Creative Commons CC BY license. Copyright © 2021, Yuan et al.

2.2. Local adaptation

2.2.1. Survival of the fittest: how natural selection shaped human populations

Natural selection is one of the evolutionary forces that shape human diversity (Key, Fu, Romagne, Lachmann, & Andres, 2016). Humans have colonised various territories where they have experienced different environmental challenges. Under these adaptative pressures, some phenotypes might be more beneficial than others for survival. Consequently, natural selection will lead to changes in the allele frequency associated with the phenotype under selection. Different kinds of natural selection exist that can influence allele frequency in different manners.

First, natural selection is called positive when acting on a genetic variant that confers an advantage in a specific environment. Individuals carrying the beneficial variant will pass it to their offspring. This offspring with higher fitness will have a higher survival rate. Carriers have then more descendants than non-carriers. The descendants of carriers will pass the beneficial variant to their offspring, which will also be more numerous. As a result, the selected variant frequency will increase in the population with time until its fixation (i.e., until it is the only remaining allele). While the selected genetic variant becomes predominant, the frequency of the other alleles on the same genomic site decrease. The loss of diversity as the selected variant frequency increases is called a selective sweep (Figure 6). A famous example of positive selection in the human population is lactase persistence, which is strongly positively selected in different populations (Bersaglieri et al., 2004). Why being able to digest milk is under such strong selective pressure is debated. It has been hypothesized that consuming milk improves survival by increasing the diet's energy content or vitamin D intake (Itan, Powell, Beaumont, Burger, & Thomas, 2009; Ségurel & Bon, 2017).

On the other hand, when natural selection leads to the removal of deleterious mutations, it is called negative or purifying selection. Negative selection decreases deleterious variants frequency in the population with time. Eventually, the deleterious variants will become rare or even disappear in the population, making negative selection hard to detect (Zhai, Nielsen, & Slatkin, 2009). Negative selection plays a major role in limiting polymorphism in loci involved in essential cellular or developmental functions (e.g., protein synthesis) and constrains the modification of the associated genome sequences to maintain them identical (conserved) through generations (Jobling et al., 2013). However, slightly deleterious mutations (i.e., nearly neutral variants) might still reach fixation when selection strength is weaker than the genetic drift, for example, when the population size is small (Ohta, 2017).

Neutral mutations, which do not confer any selective advantages or disadvantages, are the most common type of mutation found throughout the genome. Positive and negative selection indirectly impact neutral mutations through linkage disequilibrium (LD), a phenomenon in which variants located near each other on the genome are more likely to be transmitted together to the next generation. As

a result, a neutral mutation may be transmitted along with advantageous variants positively selected or removed with disadvantageous variants under negative selection. A selective sweep (i.e., variant frequency increases in the population because of positive selection, which lowers the diversity at the locus) generates a loss of polymorphism in the surrounding genomic region of the selected variant. Because of LD, the variants that co-occur nearby the selected one on the same chromosome will be transmitted simultaneously to the offspring (Kaplan, Hudson, & Langley, 1989). This process is called genetic hitchhiking. If selection is strong enough, the frequency of the selected haplotype (i.e., the set of genetic variants inherited together with the selected variant) increases faster than reverse processes, such as recombination or mutations, which break the haplotype or increase haplotype diversity (Figure 6). When a *de novo* mutation is under selection, the linked haplotype will rapidly rise in frequency in the population. This is called a hard sweep. On the other hand, a soft sweep happens when selection acts on a mutation already existing in the population, which is then likely to be included in multiple haplotypes. In this case, all haplotypes carrying the mutation will increase in frequency, which makes it harder to be detected. Similarly, because of LD, the removal of deleterious variants by negative selection results in background selection, which is the loss of non-deleterious variants linked to the variant negatively selected. Because the deleterious variant is not transmitted to the next generation, linked variants nearby this variant is less likely to be transmitted too (D. Charlesworth, Charlesworth, & Morgan, 1995).

While both positive and negative selection reduce genetic variability by favouring or removing genetic variants, respectively, another kind of selection, called balancing selection, maintains multiple alleles (polymorphisms) (Deborah Charlesworth, 2006). The high frequency of sickle-cell disease in Sub-Saharan regions has extensively been used to illustrate balancing selection (Flint, Harding, Boyce, & Clegg, 1998). This disease is caused by a mutation in the *HBB* gene that encodes the β-globin. People carrying the mutated gene, HbS, display an atypical haemoglobin (haemoglobin S) that creates deformed red blood cells (sickle cell). Because homozygotes carrying two copies of HbS suffer from detrimental sickle-cell anaemia, we would imagine that negative selection would have removed the HbS variant. However, heterozygotes carrying only one copy of the mutated gene will have a reduced risk of severe malaria (heterozygote advantage). Because of this advantage, heterozygotes, and then the HbS variant, are maintained in the population (Ackerman et al., 2005).

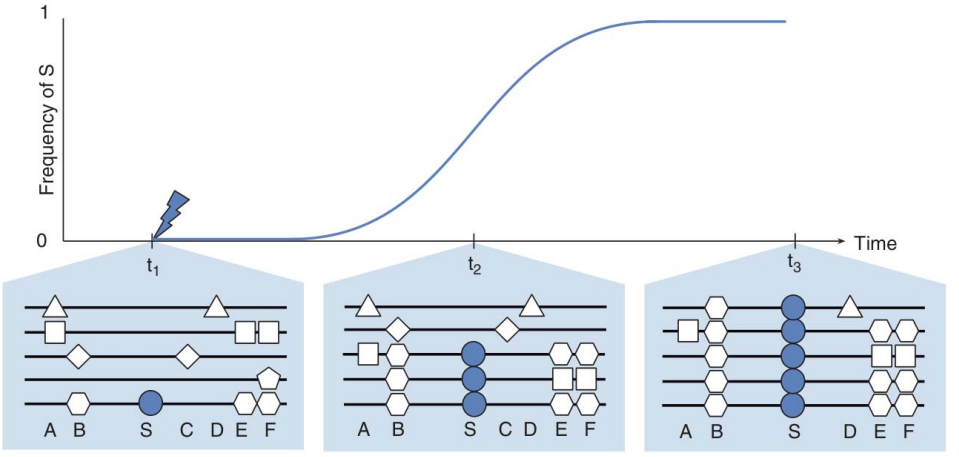

Figure 6: Genetic variation changes as a sweep progresses due to hitch-hiking. Individual haplotypes (n = 5) are denoted using a different shape for each haplotype in the sample at t1, to keep track of recombination events during the sweep. At t1, the selected allele S mutates into the population in the ⬡ background. At this stage, there are six neutral polymorphisms (A–F) in the sample. At t_2, the sweep is ongoing and the frequency of the selected allele is intermediate (incomplete selective sweep). Relative to a neutral allele, fewer recombination events have occurred around the selected allele by the time it reaches this frequency, due to its rapid increase in frequency driven by selection. As a result, diversity within haplotypes carrying *S* is much reduced compared to haplotypes carrying the disfavored ancestral allele, that is, there has been an increase in haplotype homozygosity within the allelic class carrying *S*. Note that at this stage, two recombinations have occurred, both between □ and ⬡. At t_3, *S* has swept to fixation, along with the *B* allele. Another recombination event has occurred between △ and ⬡. Note the increase in high-frequency derived alleles and overall reduced levels of variability at t_3 relative to t_1 and t_2, exemplified by the loss of diversity at three sites (B, C, S). Furthermore, the sample is IBD for the entire tract $B - S$. Reprinted from Figure 14.1 (Stern & Nielsen, 2019) with permission from John Wiley and Sons. Copyright © 2019 John Wiley & Sons Ltd.

2.2.2. Genomic approaches to detecting positive natural selection

While negative and balancing selection (Fijarczyk & Babik, 2015; Zhai et al., 2009) are complex to identify, multiple methods exist for detecting positive selection at the genome level. These techniques screen for genomic patterns compatible with signatures associated with positive selection (Rees, Castellano, & Andrés, 2020).

2.2.2.1. Haplotype-based selection scans

One signature associated with genomic regions that have been a target of positive selection is a reduced diversity on a haplotype. Estimates of haplotype homozygosity have therefore been used to scan for such regions in a population. The Extended Haplotype Homozygosity (EHH) metric is one of the statistics that

estimates such signals. The EHH score reflects the probability that two randomly chosen haplotypes are identical around a SNP up to a distance x. Because of recombination, EHH decreases when the distance x from the SNP increases. Under selection, there is less variation around the SNP – haplotypes carrying the selected SNP are more similar– and EHH will decrease more slowly (Sabeti et al., 2007; Stern & Nielsen, 2019).

The integrated Haplotype Score (iHS) is a later modification of EHH that considers the local recombination rate. Using genetic distance instead of physical distance corrects for recombination rate. Indeed, regions with low recombination tend to display more haplotype homozygosity (Voight, Kudaravalli, Wen, & Pritchard, 2006). EHH can also be modified into cross-population extended haplotype homozygosity (XP-EHH) that detects selective sweep that has occurred in one population but not another (Sabeti et al., 2007). XP-EHH detects selective sweeps in which the selected allele has achieved fixation in one population (complete sweep) but remains polymorphic in the overall human population.

Another haplotype-based statistic calculates the number of segregating sites by length (nSL) and is a measure that does not require a genetic map. Instead, it computes the length of haplotype homozygosity between two haplotypes with respect to the number of mutations within the other haplotypes from the studied dataset (Ferrer-Admetlla, Liang, Korneliussen, & Nielsen, 2014). Similarly to XP-EHH, a cross-population approach using nSL – called XP-nSL– exists that compares haplotypes between two populations. This approach offers a favourable method for detecting soft sweeps compared to XP-EHH (Ferrer-Admetlla et al., 2014; Szpiech, Novak, Bailey, & Stevison, 2021).

2.2.2.2. Allele frequency-based selection scan: Population Branch Statistic

The population branch statistic (PBS) is used to detect another type of genomic signature of positive selection. While XP-EHH or XP-nSL are based on haplotype homozygosity, PBS uses allele frequency differences between populations to identify signatures of selection. Three populations are needed to compute PBS for a target population: a target population, a reference population and an outgroup population (Yi et al., 2010). A population's PBS value represents the amount of allele frequency change at a given genomic position in the history of a target population.

The allele frequency will be used to estimate the fixation index (F_{ST}) between each population at each SNP. F_{ST} is as common neutrality test that characterizes the genetic differentiation between two populations. There are multiple ways to estimate $F_{ST,}$ but they all reflect the amount of genetic variation between populations (V_B) as a proportion of the total variation (V_T) that includes within population and between populations variations (eq. 1) (Beaumont & Wang, 2019).

$$F_{ST} = V_B/V_T \qquad \text{(eq. 1)}$$

When the allele frequency of one variant highly differs between two populations, the corresponding F_{ST} value will be close to 1, while it will be equal to 0 when the allele frequency is identical between two populations. High genetic differentiation of a SNP (high F_{ST}) between two populations might result from selection (Lewontin & Krakauer, 1973; Stern & Nielsen, 2019). However, the F_{ST} value alone cannot decipher within which population the variant is under selection. Using an outgroup population (i.e., a population as distant to the target as the reference population) for the PBS calculation allows for the directional assessment of selection (Yi et al., 2010). When picturing a tree with three branches (one for each population) (Figure 7), F_{ST} is used to estimate the branch length (or the population divergence time T^{xy}) between every population pair (eq. 2). Or for population *x* and *y*:

$$T^{xy} = -log(1-F_{ST}(x,y)) \qquad \text{(eq. 2)}$$

The branch length between each pair of populations is used to calculate the branch length from the inner node of the tree to the target population *x.* This inner branch length is the PBS.

$$PBS = \frac{T^{xy}+T^{xz}-T^{yz}}{2} \qquad \text{(eq.3)}$$

The frequency change of a genetic variant under selection is often more dramatic than that observed in variants that are not under selection, which increases its PBS score (Figure 7) (Yi et al., 2010).

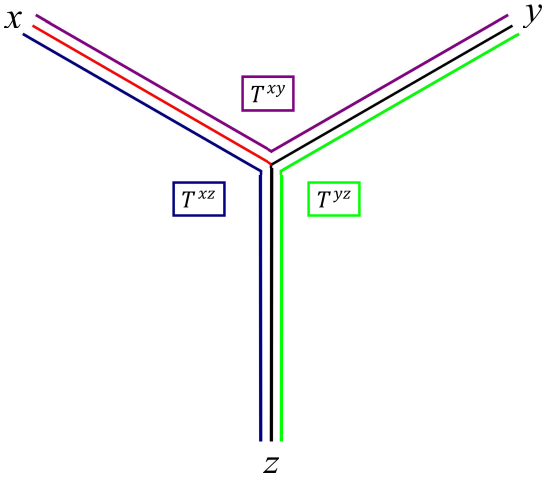

Figure 7: Computing locus-specific population branch statistic (PBS) for the target population *x*, considering a reference population *y* and an outgroup *z*. Log transformed pairwise F_{ST} are used to estimate T^{xy}, T^{xz} and T^{yz} (eq. 2), the branch length (or the divergence time) between the target and reference populations (in purple), the target and outgroup populations (in blue) and the reference and outgroup populations respectively. Pairwise branch lengths are used to estimate PBS (in red), the length of the branch to the target population x since the divergence from the reference population y and the outgroup population z, following eq. 3.

2.2.3. Challenges and limitations in interpreting selection

Although selection scans offer insights into the genomic regions under selection in a population, establishing a link between the associated genomic region and a potential target phenotype is challenging.

One way to identify the biological consequences of the candidate variants for selection is by testing the variants with functional assays to determine their function in a non-human model organism. Still, this approach is limited by the reduced transferability of a gene function between organisms (i.e., from the model organism to humans) (Rees et al., 2020). The growing number of large-scale genome-wide association studies (GWAS) have provided another source to test genomic variants for their phenotypic effects (e.g., UKbiobank (Pan-UKB team, 2020)). Nonetheless, GWAS are performed on a specific population, and the transferability of the summary statistic between populations is limited (Mathieson, 2021).

An additional challenge is that a similar phenotypic adaptation can occur due to different genetic variants, meaning that populations under identical selection pressures can display the same phenotype but with distinct genetic causes (= convergent evolution). For instance, high-altitude populations have independently adapted to low oxygen availability through genes impacting the Hypoxia-Inducible Factor (HIF) pathway. HIFs are transcription factors regulating the hypoxic response (Bigham & Lee, 2014).

It is also important to note that most phenotypes are polygenic, meaning they are impacted by many genetic variants, each of which only has a limited effect. When natural selection acts on these complex phenotypes, the frequency of each genetic variant that impacts the phenotype will only shift by a small degree. This subtle frequency shift makes detecting polygenic selection difficult (Sella & Barton, 2019).

Pleiotropy is another phenomenon, which relates to single genes or single genetic variants, that can have an effect on multiple unrelated traits, which makes it challenging to establish the relationship between a gene that is under selection and a phenotype that may be the target of selection (Stearns, 2010).

Finally, because the environment might have changed since the time of selection, identifying the initial environmental pressure that acted on a population remains complex. An example is the Duffy-null phenotype, protecting against *P. vivax*, found in high frequency in Africa, while *P. vivax* is not endemic in Africa anymore (McManus et al., 2017)). In addition, a phenotype detrimental to health might still benefit survival under specific pressure (ex: sickle cell disease) and maintained in the population by balancing selection (Fan, Hansen, Lo, & Tishkoff, 2016).

The recent omnigenic model elegantly put together these multiple concepts by suggesting there exist only a few "core" genes that directly affect a phenotype (Boyle, Li, & Pritchard, 2017). Most of the genes would be ''peripheral'' and influence the phenotype indirectly through a network of interactions with multiple peripheral and core genes. Because of the dense interconnection between genes, a single variant may affect various traits. This process is called network

pleiotropy. Nonetheless, phenotypes do not only result from genetics but are also impacted by various environmental factors. For example, obesity is associated with multiple genes and environmental factors like diet or smoking (Loos & Yeo, 2022). The omnigenic model can be extended by adding core and peripheral environment effects to the network to take account of the environmental influence on a phenotype. Between populations, the network will vary: core genes are conserved, but interactions between peripheral genes and environmental effects vary. Different omnigenic networks between populations explain the differences in effect size between GWAS in different populations (Figure 8) (Mathieson, 2021). Identifying the phenotypic consequences of variants with signatures of selection in the omnigenic network is challenging and exemplifies why translating selection signatures to their phenotypic targets is often difficult.

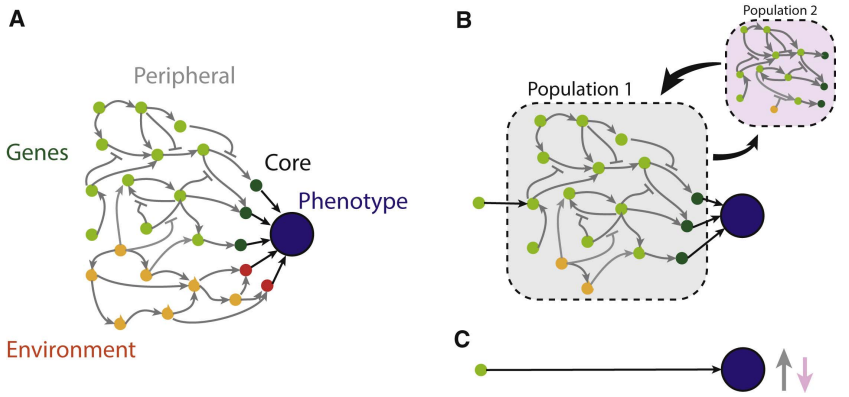

Figure 8: The omnigenic model (**A**) Schematic of the omnigenic model (after Liu, Li, & Pritchard, 2019). Genetic factors are in green and environmental factors are in orange, with arrows showing interactions. Peripheral factors are lightly shaded, whereas core factors with direct effects on phenotype are darkly shaded. (**B**) A gene's-eye perspective. A single peripheral gene's effect on the phenotype is filtered through part of the network (grey box). (**C**) The GWAS perspective. The effect of a causal variant as measured by an association study in a single population (grey arrow in **C**) is the expected value of its effect, with respect to the distribution of genetic (i.e., allele frequencies) and environmental factors in the population. In a different population (pink box in B), the weights and possibly structure of the network change, leading to a different expected effect size (pink arrow in **C**). Reprinted from Figure 1 (Mathieson, 2021) with permission from Elsevier. © 2021 American Society of Human Genetics.

2.2.4. Enhancing fitness through archaic interbreeding: the role of adaptative introgression

Neanderthals inhabited a vast territory from Southwest Europe to South Siberia from around 300 kya until around 40 kya (Higham et al., 2014; J. J. Hublin, 2009). There is a lack of fossil records for Denisovans, but indirect evidence suggests that their territory might have spread from Siberia (Reich et al., 2010) to Sulawesi (van den Bergh et al., 2016). They might have inhabited these territories between

100 and 200 kya (Douka et al., 2019; van den Bergh et al., 2016). The extended time that archaic humans, such as Neanderthals and Denisovans, were exposed to their environments gave selection time to act on their genome and increase their fitness to environmental challenges. When AMH interbred with these archaic groups, their genomes inherited some of these selected archaic variants that might have helped them to enhance their fitness faster than under a selective process acting on a *de novo* mutation (Racimo, Sankararaman, Nielsen, & Huerta-Sánchez, 2015). A noteworthy example is a variant of the *EPAS1* gene introgressed from Denisovans, found in high frequencies in Tibetans, and associated with reduced haemoglobin concentration. The archaic variants at this gene locus might have helped Tibetan to counteract high-altitude hypoxia's detrimental effects (Huerta-Sánchez et al., 2014). Nonetheless, selection might have happened significantly after the admixture event and would not constitute a real adaptative introgression but rather a selection on a standing variant introgressed earlier (X. Zhang et al., 2021). Adaptative introgression has also notably contributed to the modern human immune system and skin pigmentation (Dannemann, Andrés, & Kelso, 2016; Mendez, Watkins, & Hammer, 2012; Reilly et al., 2022; Sams et al., 2016; Sankararaman et al., 2016; Vernot & Akey, 2014).

On the other hand, background selection has also been shown to play an important role in introgressed archaic DNA. The smaller effective population sizes of Neanderthals and Denisovans, compared to AMH, meant that selection against harmful gene variants was weak in these archaic hominin populations. As a result, their genomes accumulated weak deleterious alleles. However, after admixture with AMH, which had a larger effective population size, selection against these introgressed deleterious variants became stronger (Harris & Nielsen, 2016; Juric, Aeschbacher, & Coop, 2016). Negative selection has particularly targeted introgressed variants in conserved regulatory regions (Petr, Pääbo, Kelso, & Vernot, 2019). In addition, there are also large genomic regions in present-day people that are devoid of introgressed Neanderthal and Denisvovan variants, also referred to as archaic deserts. These regions are enriched for genes expressed in the brain and testis. They are also enriched on the X chromosome. While it is difficult to pinpoint the exact target of selection, these results suggest that sex-biased admixture processes, potentially associated with incompatibilities in reproduction or brain-related traits, might have driven the removal of archaic DNA in these regions (Sankararaman et al., 2014; Vernot & Akey, 2014; Vernot et al., 2016).

2.2.4.1. Adaptative introgression in New Guineans

New Guineans with Indigenous Australians and Philippine Ayta carry both Neanderthal introgressions and have one of the highest Denisovan introgression rates (Larena et al., 2021; Reich et al., 2010). Introgressed loci in New Guineans might have enabled them to face the environmental pressure in New Guinea by shaping their immune system (Vespasiani et al., 2022). For example, *TNFAIP3* (Gittelman et al., 2016; Zammit et al., 2019), *STAT2* (Mendez et al., 2012; Yermakovich et al., 2022), and the *GBP* locus (Vernot et al., 2016) are three

introgressed archaic loci candidates for selection in New Guineans that are involved in the interferon-induced cell-autonomous host defence (MacMicking, 2012). Moreover, archaic introgression might also have contributed to New Guinean dietary adaption through changes in the metabolism (e.g., *FADS1* (Jacobs et al., 2019), *SUMF1* (Choin et al., 2021; Jacobs et al., 2019)).

2.2.5. Adaptation to extreme environments

Since they arrived in northern Sahul, New Guinean populations have experienced various selection pressures associated with the variety of environments they have settled in. New Guinean highlanders and lowlanders face different environmental challenges, with the strongest challenges being hypoxic stress and pathogenic pressure, respectively.

2.2.5.1. Altitude pressure

2.2.5.1.1. Altitude pressure worldwide

Altitude is a severe environmental pressure inflicted on human beings (Fan et al., 2016). The human body's biggest challenge at altitude is hypoxia– the lower oxygen availability to body tissues. Hypoxia not only occurs during summit ascent but is also a condition common to various diseases (e.g., heart failure, lung diseases, anaemia, cancer, covid-19) (Lee, Chandel, & Simon, 2020; Rahman et al., 2021). Hypoxia is also a common pregnancy complication (Piešová & Mach, 2020). The hypoxia-inducible factor (HIF) pathway orchestrated most cellular mechanisms that preserve oxygen homeostasis in the hypoxic organism (Lee et al., 2020). But despite its importance in human health, the body's responses to hypoxia are still not fully understood (Lee et al., 2020). Many studies conducted in high-altitude populations have enhanced our knowledge of the human body's strategy to counteract the detrimental effect of hypoxia.

For example, Tibetans, Andeans, and Ethiopian Amharas display phenotypic traits that seem to counterbalance the decrease in oxygen in the blood because of hypoxia (Moore, 2017). Andeans have a wider chest, larger lung capacity, and higher haemoglobin levels compared to their corresponding lowlanders (C. M. Beall et al., 1998; Brutsaert, Soria, Caceres, Spielvogel, & Haas, 1999; Mueller, Schull, Schull, Soto, & Rothhammer, 1978), while Tibetans show lower haemoglobin levels due to higher plasma volume (C. M. Beall et al., 1998; Stembridge et al., 2019). At first glance, because haemoglobin is the molecule transporting oxygen in the blood, lower haemoglobin concentration can seem maladaptive for living at high-altitude. Indeed, in response to the lower oxygen levels, lowlanders at high-altitude will experience an increase in haemoglobin concentration to enhance oxygen delivery to the body tissues. Nonetheless, an increase in haemoglobin concentration also results in higher blood viscosity, which is detrimental in the long term. Higher blood viscosity puts more stress on the cardiovascular system and results in cardiovascular and pregnancy complications. The high

haemoglobin levels of Andean highlanders might even explain their high susceptibility to chronic mountain sickness (CMS) (Villafuerte & Corante, 2016). Ethiopian highlanders also show increased haemoglobin levels and a larger chest (Alkorta-Aranburu et al., 2012; Hoit et al., 2011; Scheinfeldt et al., 2012), although not to the same extent as Andeans. However, not all studies have confirmed the increased haemoglobin levels in Ethiopians (Cynthia M. Beall et al., 2002; Cheong et al., 2016; Huerta-Sánchez et al., 2013).

Genomic studies revealed that many genes considered for high-altitude adaptation are associated with the HIF pathway (Bigham & Lee, 2014). Among these genes are *EGLN1*, which is under selection in both Andean and Tibetan highlanders, and *PRKAA1*, *EPAS1* and *HLHE41,* which are candidates for selection in Andean, Tibetan and Ethiopian Amhara highlanders, respectively (Bigham & Lee, 2014; Huerta-Sánchez et al., 2013). *EPAS1*– associated with the characteristic lower haemoglobin concentration level in Tibetans (Simonson et al., 2010; Yi et al., 2010) – is particularly noteworthy as its selected Tibetan variant is associated with an introgressed Denisovan haplotype (Huerta-Sánchez et al., 2014). Nonetheless, it has been suggested that the selection of this variant postdates the introgression event (X. Zhang et al., 2021). Another matter of discussion is the phenotype targeted by selection in Tibetans. It has been shown that selection might target a higher plasma volume instead of a higher haemoglobin concentration (Stembridge et al., 2019). Moreover, selection to hypoxia also acted on intermediate altitude populations such as Andean Calchaquíes living around 2,300 m a.s.l. that show a strong signal of selection for *PRKG1*, suggesting adaptation of the nitric oxide pathway (Eichstaedt et al., 2015).

Although natural selection may explain part of the specific phenotypes of highlanders worldwide, these traits may result from other non-mutually exclusive factors such as genetic adaption to other high-altitude stressors (e.g., cold) (Payne, Kumar Bc, Pomeroy, Macintosh, & Stock, 2018), ancestry, high-altitude exposure during growth (Frisancho, 2013), lifestyle (Cynthia M. Beall, 2014; Leonard, 1989), and socioeconomic status (Greksa, 2006).

2.2.5.1.2. Altitude pressure in New Guinea highlands

New Guinea has more than 40 summits above 4000 m a.s.l .and is the most mountainous region of the Sahul continent. The first settlers entered New Guinean highlands for semi-permanent settlement as early as 25 kya (Summerhayes et al., 2017). However, despite the extended time they have inhabited the highlands, New Guineans highlanders have been understudied regarding adaptation to hypoxia. Since they permanently settled in the highlands 20 kya (Malaspinas et al., 2016; Mountain, 1991; O'Connell et al., 2018; Summerhayes et al., 2010), New Guinean highlanders remained relatively genetically isolated through time, with only minor influences from outside (Bergström et al., 2017). Nowadays, the most densely populated places in PNG are located at intermediate altitude (above 1500 m a.s.l.) between 1,600 and 2,400 m a.s.l. (Brookfield & Allen, 1989; Müller et al., 2003). Previous studies on the health of PNG populations showed

that highlander groups display an increase in ventilatory lung function and haemoglobin concentration (Cotes, Anderson, & Patrick, 1974; Senn et al., 2010; Woolcock, Colman, & Blackburn, 1972). While these data were not explicitly studied in the context of adaptation to hypoxia, they strongly suggest that PNG highlanders could display similar phenotypic traits as those observed in other intermediate and high-altitude populations worldwide.

2.2.5.2. Pathogenic pressure

As AMH expanded across new territories, they were exposed to different pathogens that exerted strong selective pressure leading to the selection of protective genes (Fumagalli et al., 2011). At the same time, pathogens evolved to counteract the humans' newly acquired adaptations, leading to new selective pressures. Consequently, the immune system has been a prime target of natural selection throughout human evolution (Siddle & Quintana-Murci, 2014). Moreover, environmental changes (e.g., climate change) and cultural practices (e.g., agriculture, medicine) created constantly shifting pathogenic pressures. This dynamic interplay of adaptive traits and selective pressures is theorized by the Red Queen hypothesis (Van Valen, 1973). It explains how pathogenic pressure and diseases are significant and constant driving forces in human evolution. This remains true today, as demonstrated by the recent emergence of novel pathogens such as SARS-CoV-2, the virus responsible for COVID-19 (Souilmi et al., 2021). These constant challenges left distinct signatures of selection in the genomes of human populations. For example, pathogenic pressure has driven the diversity of Human Leukocyte Antigen (HLA) class I genes in humans (Prugnolle et al., 2005). More specifically, some populations show adaptation to a specific disease, like the above-mentioned widespread sickle-cell allele HbS protecting sub-Saharan populations from malaria (Ackerman et al., 2005). Additionally, archaic introgression may have contributed to immune system adaptations in AMH (Dannemann et al., 2016; Enard & Petrov, 2018; Mendez et al., 2012; Quach et al., 2016; Sams et al., 2016; Vespasiani et al., 2022).

2.2.5.2.1. Malaria pressure worldwide

Malaria is the most common parasitic disease in humans (White, 2018) and one of the leading causes of death for children under five years old (World Health Organization, 2022b). This infection of the red blood cells by a protozoan parasite of the *Plasmodium* group is transmitted through bites of infected *Anopheles* mosquitoes. *Plasmodium falciparum* and *Plasmodium vivax* are the leading cause of malaria in humans (Haldar & Mohandas, 2009). Malaria infection induces the destruction of parasitised and unparasitized red blood cells and bone marrow dysfunction (Totino, Daniel-Ribeiro, & Ferreira-da-Cruz, 2016; White, 2018). Malaria is of major health concern worldwide. The World Health Organization (WHO) recorded 247 million malaria cases and 619 000 deaths from malaria in

2021. Most of these cases were in sub-Saharan countries (World Health Organization, 2022a).

The most common effect of malaria is anaemia – a decrease in haemoglobin concentration – which causes substantial infant mortality in the affected territories (Haldar & Mohandas, 2009; White, 2018). Another common side effect of malaria is decreased platelet count (thrombocytopenia) (O'Sullivan, Preston, O'Regan, & O'Donnell, 2016). Malaria can reach more severe stages like cerebral anaemia, when the infected red blood cells stick to the brain endothelium from *P. falciparum* infection (Bafaro, Liu, Xu, & Dempski, 2017; Haldar & Mohandas, 2009). Platelets have been shown to play a crucial role in this process by mediating the adhesion of infected red blood cells together and to the endothelium (O'Sullivan et al., 2016; Pain et al., 2001). This haemostatic dysfunction can even result in vascular occlusion and increase the risk of mortality from *P. falciparum* or *P. vivax* malaria (Lampah et al., 2015). However, platelets also have beneficial effects in regulating malaria symptoms by binding to infected red blood cells and contributing to *Plasmodium* killing (Kho et al., 2018). Thus, platelets play a complex and essential role in the pathogenesis of malaria, but their complete contribution to disease outcomes is subject to ongoing research (McMorran, 2019).

While most of the deaths and severe symptoms are attributable to *P. falciparum*, *P. vivax*, the most widespread type of malaria outside of Africa (Gething et al., 2012), can still lead to severe anaemia (Haldar & Mohandas, 2009; White, 2018). Moreover, *P. vivax* might induce heart pathologies (Bammigatti, Shetty, Shetty, & Kumar, 2011; Bhat, Kumar, & Alva, 2013; Gupta et al., 2021; Holm, Gomes, Biering-Soerensen, Silvestre, & Brainin, 2020).

Because of its high mortality rate among children, malaria exerts strong selective pressures on the human genome (Kwiatkowski, 2005). Malaria-induced selective pressure is linked to the emergence of high-frequency haemoglobinopathies and red blood cell polymorphisms in impacted regions (Flint et al., 1998). A striking example we have addressed in the chapter above is the frequency of the HbS allele, causing deformed red blood cells. While HbS homozygotes suffer from sickle-cell anaemia that can often be fatal, heterozygotes are protected against severe forms of malaria (Ackerman et al., 2005). The heterozygote advantage maintains HbS at around 10% in malaria-endemic regions thanks to balancing selection (Flint et al., 1998). But malaria has kept a huge variety of other red blood cell polymorphisms, such as the Duffy blood group-negative phenotype that offer a complete resistance against malaria caused by *P. vivax*. Duffy-null red blood cells do not carry the Duffy antigen receptor for chemokines (DARC) that P. vivax uses to infect the red blood cell. The allele associated with this phenotype is called FY*O, which is near fixation in western and central Africa but carries the signature of strong selection (McManus et al., 2017).

2.2.5.2.2. Pathogenic pressure in New Guinea lowlands

New Guineans are exposed to a variety of pathogens (e.g., malaria, dysentery and pneumonia, tuberculosis, HIV, etc.) that are the leading cause of death in PNG (GBD 2013 Mortality and Causes of Death Collaborators, 2015; Kitur et al., 2019; Naraqi et al., 2003). Nevertheless, the environmental diversity in PNG leads to a unique situation regarding the epidemiology of malaria. We previously discussed the uneven population distribution in PNG, with most people living between 1600 and 2400 m a.s.l. (Brookfield & Allen, 1989; Müller et al., 2003) since the Holocene (Trájer et al., 2020). A possible explanation for this uneven distribution could be the migration of early settlers of New Guinea towards the highlands due to the high pathogenic pressure in the lowlands (Riley, 1983; Trájer, 2022; Trájer et al., 2020). Malaria might be one of the leading causes of this pathogenic pressure, even though we don not know when it first appeared in the lowlands (Trájer et al., 2020). Indeed, malaria is widespread in New Guinea: PNG accounted for nearly 87% of the malaria cases and 94% of malaria deaths in the Western Pacific Region in 2021 (World Health Organization, 2022a). Malaria estimated mortality rate of children under 10 is between 4 and 17% depending on the region in PNG (Cleary, Hetzel, & Clements, 2022). Despite these numbers, New Guinean highlands are nearly malaria-free (Müller et al., 2003; Senn et al., 2010). An association between altitude level and lower malaria prevalence has also been observed in other populations like Tanzanians and Kenyans (Akhwale et al., 2004; Drakeley et al., 2005).

New Guineans lowlanders carry polymorphisms associated with blood-related traits that might have been selected to face this endemic malaria pressure (Müller et al., 2003), like G6PD deficiency, a-thalassaemia, and the Gerbich (Ge)-negative blood group, that are absent or rare in PNG highlands (Müller et al., 2003). Interestingly, PNG highlanders affected by malaria have a higher risk of thrombocytopenia (decrease in the platelet count) than lowlanders, suggesting the absence of selection in PNG highlanders (Lampah et al., 2015).

2.3. Estimating genealogies

Inferring genealogies, considering coalescent and recombination events for a set of sequences, has become an essential tool in statistical population genomics. However, estimating genealogies for large whole genome datasets can be computationally challenging. This has led to the development of various statistical methods for inferring genealogies, providing insight into the evolutionary history of populations and the forces that have shaped their genomes.

2.3.1. The neutral Wright-Fisher model

To draw meaningful evolutionary inferences from a dataset, it's crucial to understand the genealogy that has led to its diversity. The neutral Wright-Fisher model is widely used to build such genealogy (Hein, Schierup, & Wiuf, 2005). This

model represents the effects of genetic drift and random neutral mutation at a given locus of the studied samples. The Wright-Fisher model is used to infer demographic and evolutionary parameters that have led to the current genomic diversity (Nordborg, 2019). It is an idealized representation of how a population evolves over time and relies on several assumptions that simplify the simulation of genealogies. The first assumption is neutrality, meaning that selection has not influenced the allele frequency in the studied dataset. This assumption implies that no mutation in the population has been selected positively or negatively. Therefore, no mutation in the population will affect the number of offspring produced by carriers of that mutation. Panmixia, or random mating, is a second assumption of this model. Under this assumption, each individual in a given generation is considered to have randomly inherited its genes from the previous generation's pool of parents. In other words, each parent in the previous generation has an equal chance of being the genetic contributor for any given individual in the next generation. Consequently, the model can be applied to a subset of a population without knowledge about the whole population. Finally, the Wright-Fisher model assumes discrete and non-overlapping generations, an absence of migration and constant population size at every generation. Individuals at each generation are obtained by sampling with replacement from the parents (Hein et al., 2005; Nordborg, 2019). Besides its simplicity, the Wright-Fisher model becomes computationally challenging when the population size is much larger than the size of the dataset being studied, as it is usually the case.

2.3.2. The coalescent theory

The coalescent model approximates the Wright-Fisher model for large populations (Kingman, 1982). A key feature of the coalescent theory is that the genealogy is built backwards in time. When we think of a family tree, we instinctively imagine it is going forward in time, from the ancestors to the descendants. Lineages of this tree split into separate new lineages whenever an individual has multiple offspring and end when there are no more descendants. However, coalescent trees are modelled in reverse: they start with the dataset being studied and go backwards to the ancestors. In a coalescent tree, lines converge (coalesce) when two or more individuals share a common ancestor (Figure 9). Because coalescent trees are built backwards in time, lineages without descendants at the basis of the tree are excluded. Not having to do calculations for these extinct lineages enable more efficient computing than the Wright-Fisher model.

Coalescent trees keep track of coalescence events between lineages until, eventually, the most recent common ancestor (MRCA) to the sample is found; this gives the time to the most recent common ancestor (TMRCA) or the first time when the sample had one ancestor only (Hein et al., 2005; Nordborg, 2019). Mutations seen in the dataset are superimposed on the tree branches in a forward-in-time manner. The more mutations there are on a tree branch, the longer the branch is and the time to the next coalescent event (Hein et al., 2005; Nordborg, 2019).

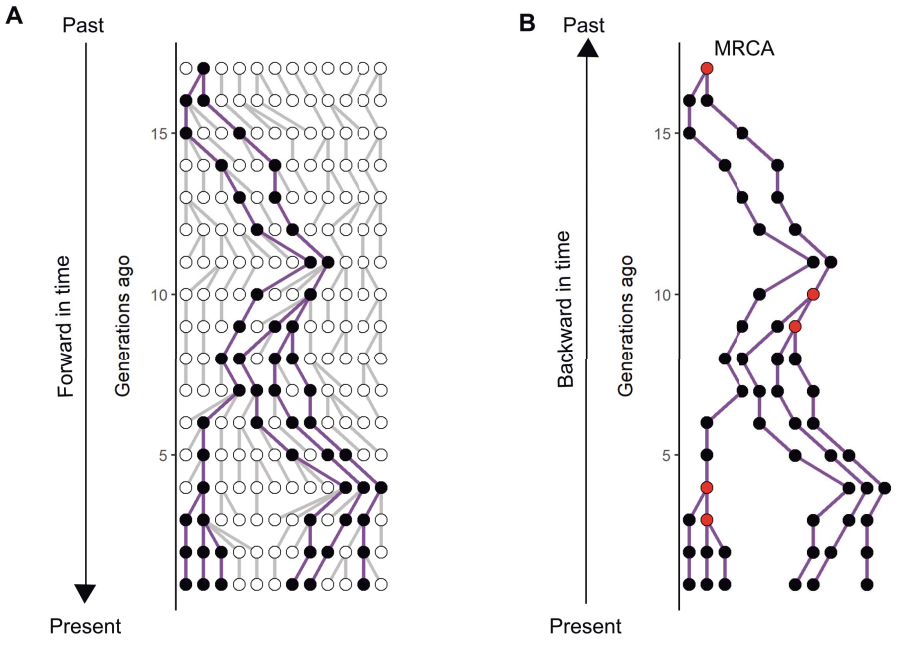

Figure 9: Genealogy of a sample of 6 individuals inferred under the Wright-Fisher model and the coalescent model. A. Genealogy for a population of size 12 under the neutral Wright-Fisher model for 17 generations. The Wright-Fisher model works forward in time, from the past to the present. Individuals that are ancestors to a sample of 6 individuals in the final generation are represented as black dots. The genealogy of the sample is shown in purple. **B**. Tree of the sample inferred with the coalescent model. Because the coalescent theory has a backwards-in-time perspective, the tree only tracks the ancestors that have contributed to the studied sample. This process is more efficient than the Wright-Fisher model – especially when the population size is larger than the studied sample – because it does not have to track individuals that did not contribute to the study sample. Coalescent events are shown in red. The last coalescent event of this tree corresponds to the sample's most recent common ancestor (MRCA).

The coalescent theory is used to understand the statistical properties of a sample of genomes from a population. It is particularly useful to study biological phenomena that affect the genealogical process but do not affect the mutation process (e.g., population subdivision). It is also used to understand how model parameters (mutation rate, generation time, etc.) affect polymorphism data by understanding how they affect genealogies (Nordborg, 2019).

2.3.3. Coalescent with recombination

During gamete formation, a diploid cell in humans containing 23 pairs of chromosomes undergoes a division that produces four haploid cells, each with 23 chromosomes. During this cell division, called meiosis, variants located on different chromosomes will be transmitted independently to different gametes. In contrast, variants on the same chromosome will be inherited together. However, two

chromosomes from the same pair (homologous chromosomes) can crossover during meiosis and exchange some of their genetic material. This crossover creates a recombination event: variants initially located on the same chromosome are not inherited together (Jobling et al., 2013). The more physical distance there is between two variants, the more likely the recombination event will happen between them. In contrast, variants close to each other remain more likely to be co-inherited, resulting in LD between these variants (Altshuler, Donnelly, & The International HapMap Consortium, 2005).

The frequency of recombination between two variants is not linearly associated with the physical distance between these variants. Indeed, the recombination rate is non-uniform along the genome: recombination events are more frequent toward the telomere but less frequent toward the centromere. Moreover, there are recombination hotspots, small genomic regions with a high recombination rate, or coldspots where the recombination rate is way lower (Crawford et al., 2004; McVean et al., 2004; Myers, Bottolo, Freeman, McVean, & Donnelly, 2005). The recombination rate also differs between sexes, populations and individuals (Kong et al., 2010). Because the physical distance, measured in base pair (bp), cannot reflect the frequency of recombination between two variants, a genetic distance has been defined and is expressed in centiMorgan (cM). One cM between two variants equals a recombination event frequency of 1% between these two variants per generation. Genetic distance measurements along the genome are stored in a genetic map (or recombination map) (Haldane, 1919).

Recombination can break down the ancestral relationships between haplotypes, leading to more complex genealogical histories. The classical coalescent theory assumes no recombination between loci, and the genealogy of haplotypes follows a simple branching pattern back to a common ancestor. However, since considering recombination leads to a more complex genealogy, it becomes necessary to consider the probability of different genealogical histories. Therefore, incorporating recombination events into coalescent models is essential to accurately infer the genealogical relationships between haplotypes and their evolutionary history. Considering recombination is particularly important in population genetics research, where understanding the genealogy of haplotypes is essential for inferring demographic and evolutionary processes.

2.3.3.1. Representing coalescent with recombination: Ancestral Recombination Graph

The addition of recombination boosts the complexity of the coalescent model. When going backwards in time, coalescent events are represented as two sequences merging into one common ancestor, while recombination events is a sequence that splits into two different ancestors (Figure 10). Because of these recombination events along the genome, different genomic loci can have different genealogies, and a genomic sequence cannot be represented with a unique tree anymore (Hein et al., 2005; Nordborg, 2019). Instead, one tree will exist for each site of the sequences (local trees). The genealogy of the whole sequence will be

a collection of these local trees in the form of a complex graph called the ancestral recombination graph (ARG) (Griffiths & Marjoram, 1997; Hudson, 1983).

The ARG is precious for studying evolutionary scenarios where recombination has occurred, especially when analyzing whole-genome data. ARG can be used to visualize relationships among lineages, estimate recombination and mutation rate, date mutation and recombination events, extract information about the demography history of the sample or detect selection (Arenas, 2013; Hubisz & Siepel, 2020).

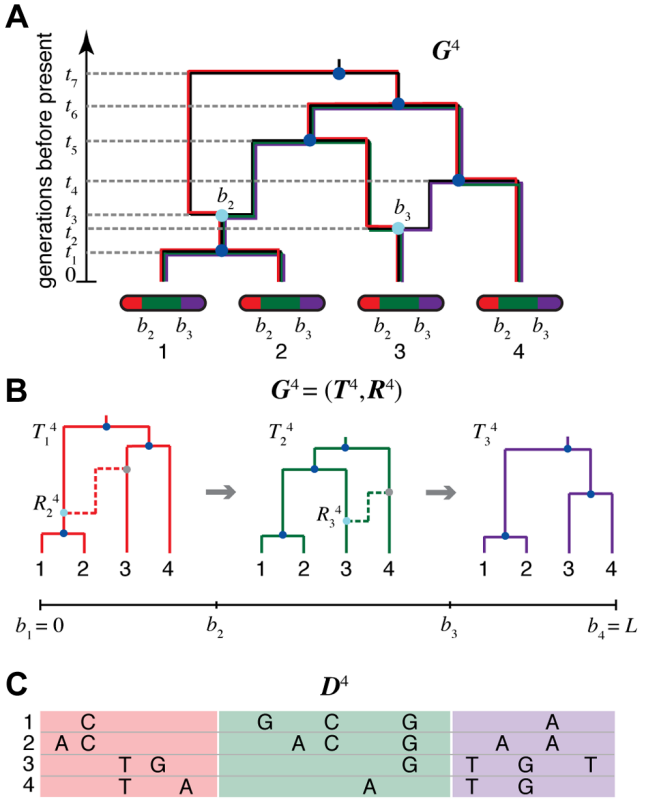

Figure 10: An ancestral recombination graph (ARG) for four sequences. (**A**) Going backwards in time (from bottom to top), the graph shows how lineages that lead to modern-day chromosomes (bottom) either "coalesce" into common ancestral lineages (dark blue circles) or split into the distinct parental chromosomes that were joined (in forward time) by recombination events (light blue circles). Each non-recombining interval of the sequences (shown in red, green, and purple) corresponds to a "local tree" embedded in the ARG (shown in matching colours). Recombinations cause these trees to change along the length of the sequences, making the correlation structure of the data set highly complex (B) Representation of the ARG in terms of a sequence of local trees $\boldsymbol{T}^4$ and recombination events $\boldsymbol{R}^4$. (**C**) An alignment of four sequences, $\boldsymbol{D}^4$, corresponds to the linearized ARG shown in (**B**). Reprinted from Figure 1 (M. D. Rasmussen, Hubisz, Gronau, & Siepel, 2014), licensed under Creative Commons CC BY 4.0 license. Copyright © 2014 Rasmussen et al.

2.3.3.2. Inferring coalescent with recombination with the Hidden Markov model

Coalescent models with recombination can be easily simulated (Hudson, 1983) but using these models on genome-wide data remains computationally demanding. However, some models can now approximate coalescence with recombination on sequenced data. One such model is the Li and Stephen (LS) model, which invokes a Hidden Markov Model (HMM) (Li & Stephens, 2003; Lunter, 2019; M. D. Rasmussen et al., 2014). An HMM models a sequence of values (called states) considering two fundamental properties. First, a future state depends only on the current state, so there is no need to know the full chain of events. Secondly, the states are hidden, meaning they cannot be observed directly but are associated with observable variables. An HMM uses observable states to estimate hidden states. The LS model infers the coalescent and recombination events (i.e., the hidden states) that generated the sampled sequences using the genetic data (i.e., the observable states). In other words, the LS model reconstructs a target haplotype of an individual at a given locus, the hidden state, as a mosaic of all the remaining haplotypes in the dataset. This is an imperfect mosaic because the LS model considers mutation rate, allowing some variation between the reference and target haplotypes. Although the LS model is a significant step in inferring coalescent and recombination events at the whole genome level, the number of generated states remains high, which means it cannot be used on large cohorts (Lunter, 2019; M. D. Rasmussen et al., 2014).

2.3.4. Estimating genome-wide genealogies for large datasets

While full likelihood methods that incorporate all the information stored in ARGs may be up-and-coming, they, unfortunately, remain computationally intractable for large genomic datasets. However, several methods have been developed that infer ARGs using various approximations and assumptions (Kelleher et al., 2019; M. D. Rasmussen et al., 2014; Speidel, Forest, Shi, & Myers, 2019), which can be used to quantify selection or other population genomics statistics. Some methods even infer population genomics statistics under a coalescent model without generating trees (Albers & McVean, 2020).

2.3.4.1. RELATE: inferring marginal trees

The RELATE software addresses the ARG computing problem by inferring marginal coalescence trees instead of full ARG. First, RELATE estimates the tree topology at the first SNP at the 5' end of a chromosome using a hierarchical clustering algorithm that will define the order of coalescence event between haplotype pairs from distance matrices calculated with a modified Li and Stephen HMM (Li & Stephens, 2003). To reduce computing, RELATE does not estimate a tree at every SNP but instead builds a new tree only if a mutation on a SNP cannot be mapped on a unique branch of the previous tree (marginal tree). Once trees have

been assigned to each SNP, RELATE estimates the branch lengths jointly with the coalescence rate with a Markov Chain Monte Carlo (MCMC) algorithm.

RELATE can perform different downstream analyses using these marginal trees as input. It can, for example, estimate within-population and cross-population coalescence rates, two estimates used to understand populations' separation history. RELATE also includes a selection test based on estimating the speed of spread of the lineage carrying the derived allele relative to the non-carriers lineages.

2.3.4.2. Genealogical Estimation of Variant Age

Genealogical Estimation of Variant Age (GEVA) is a method to estimate the age and origin of genetic variants. This method relies on the fact that, under some given assumptions, the time of mutation events is delimited by the TMRCA of lineages carrying the mutation and the most recent time these carrier lineages coalesce with non-carrier sequences. GEVA will first select random concordant pairs (pairs of sequences carrying the mutation) and random discordant pairs (one sequence carrying the mutation and one sequence that does not). GEVA uses an HMM to estimate the region where the MRCA remains unchanged for concordant and discordant pairs. Specifically, the HMM calculates the distance to the closest recombination event on either side of the focal position along the sequence. The inferred sequence is also used to determine the number of mutations that have occurred on the branch to the MRCA of the haplotype. Secondly, GEVA will infer the pairwise TMRCA between lineages (concordant and discordant) pairs for three probabilistic models that account for mutation rate, recombination rate or both. Finally, information from the pairs is combined within a composite-likelihood framework to obtain an approximate posterior distribution on the age of the derived allele.

2.3.5. CLUES: Inferring selection using marginal trees

Selection tests based on RELATE have several limitations: they use single-point estimates of the branch length and do not take uncertainty into account. Moreover, they assume that the tree is derived from a population of constant size, which is usually not true. The Coalescent Likelihood Under Effects of Selection (CLUES) method is an approximate full-likelihood method for inferring selection that has been developed to overcome these limitations (Stern, Wilton, & Nielsen, 2019). CLUES uses the local tree generated with RELATE containing coalescence times for a SNP of interest to compute the ratio of the likelihood of the SNPs being under selection given a coefficient of selection versus the likelihood of the SNP being under neutrality. CLUES also infers the change in derived allele frequency (DAF) over time for a model with recombination and selection using only modern data. This DAF trajectory is an asset because it allows estimating strength and timing selection from modern data, which previously required genomes from different time periods.

3. AIMS OF THE STUDIES

The settlement of Sahul results from one of the most ancient human migrations of AMH. The paths AMH took across Wallacea, before reaching northern Sahul, are still debated. The location of the entrance point to Sahul and the dynamic of settlements within Sahul are also still areas of investigation. Because they faced new environments, the first settlers encountered new adaptative pressures once they reached New Guinea. This thesis aims to investigate New Guinea populations' genetic and phenotypic diversity using a combination of new whole-genome sequences and phenotypic measurements. The first reference explores genomic models for migrations from Sunda to Sahul and the settlement within New Guinea. The second study concentrates on the chronology of selection events during the migration of AMH across Wallacea and their settlement in northern Sahul using published genomes from Wallacea, New Guinea and the Bismarck Archipelago. The third study compares phenotypic differences between PNG highlanders and lowlanders. The fourth study investigates selection signals in the same PNG highlanders and lowlanders. These studies shed light on New Guinean populations' genetic and phenotypic diversity and evolutionary history.

3.1. Aims of the first study (Ref I)

The exact migratory routes taken by the initial settlers of Sahul remain unclear (Allen & O'Connell, 2020). In this study, we sought to address multiple scenarios by reconstructing the key genetic events that led to the settlement of New Guinea using a genomic model. We first investigated whether one or several human groups conducted the migration to Sahul and which was the most likely migratory route. We also explored how New Guineans settled the diverse environments of the current island and the dynamics of this settlement. To accomplish this, we sequenced 58 new whole genome sequences sampled in Port Moresby, filling geographical gaps in previous samplings of PNG sequences and providing the first genomic model for the settlement of northern Sahul, considering one or two migrations from Wallacea.

3.2. Aims of the second study (Ref II)

When AMH migrated through Wallacea and settled in northern Sahul, they had to adapt rapidly to unfamiliar environments. This study aimed to establish the chronology of selection events during this settlement. We analysed selection signals in whole-genome sequences previously published for three populations: Wallaceans, New Guineans, and Bismarck Islanders. We chose these populations due to their relative isolation, as we aimed to identify selection signals specific to each stage of the northern Sahul settlement. We used cross-population selection

tests, PBS, XP-EHH and XP-nSL (Sabeti, 2006; Szpiech & Hernandez, 2014; Yi et al., 2010), to identify unique selection signals in these three populations. By detecting such selection events, we established a relative chronology of the adaption. We complemented these efforts with GEVA (Albers & McVean, 2020) to compute the coalescence time of the variants under selection to obtain an absolute chronology of these events. Finally, we explored the contribution of Denisovan introgression to the loci under selection.

3.3. Aims of the third study (Ref III)

Multiple populations world-wide have adapted to living at high-altitude and exhibit different phenotypes than their lowlander counterparts. However, despite exposure to hypoxia for the last 20,000 years, limited information is available on the impact of selection and phenotypic variation on PNG highlanders. We hypothesised that PNG highlanders differ from PNG lowlanders in phenotypes that have helped them cope with the lower oxygen availability in the highlands. To test this hypothesis, we explored how 13 newly measured phenotypes related to cardiovascular health, body proportions, and lung capacities differ between PNG highlanders and lowlanders. We also compared our findings to those made in high-altitude populations from the Andes, Ethiopia, and Tibet.

3.4. Aims of the fourth study (Ref IV)

PNG highlanders and lowlanders have faced different environmental challenges since they separated from their shared ancestral population. PNG highlanders have been exposed to low oxygen levels for the last 20,000 years, while PNG lowlanders have encountered a unique pathogenic environment. We hypothesized that PNGs might show distinct selection signals depending on their environmental niche. We further hypothesized that these selection signals would be associated with phenotypic differences defined by Ref III (André et al., 2021). To explore this hypothesis, we sequenced and analysed 128 new whole-genomes from PNG lowlands and highlands. We detected unique selection candidates for PNG lowlanders and highlanders using PBS and XP-EHH selection scores (Sabeti, 2006; Yi et al., 2010). We used CLUES (Stern et al., 2019) to identify the most likely SNP to drive selection signal in each region under selection in PNG highlanders and lowlanders. We looked for associations of these candidates SNPs with phenotypes from the UKBiobank (Pan-UKB team, 2020) or measured in PNGs in Ref III (André et al., 2021).

4. MATERIAL AND METHODS

4.1. Ref I

This study includes 58 newly sequenced Papua New Guinean participants sampled in Port Moresby. These participants originate from different parts of the country and then increase the geographical coverage of New Guinea, which is under-represented in previously published whole-genome data sets. We combined the new PNG sequences with published genomes from Indonesia (Jacobs et al., 2019), northeast Australia (Mallick et al., 2016), and the Bismarck Archipelago (Vernot et al., 2016), and with genotyping data from North Maluku Islands located on the northern route of Wallacea (Kusuma et al., 2017) and imputed genotyping data from PNG (Bergström et al., 2017). We used fineSTRUCTURE to identify the regional groups within our dataset (Lawson, Hellenthal, Myers, & Falush, 2012). Because our study focuses on migrations movements before the Lapita people reached New Guinea, we masked the Asian genetic ancestry in our dataset using haplotypic information with PCAdmix (Brisbin et al., 2012). We further explored the diversity of our dataset by performing ADMIXTURE (Alexander, Novembre, & Lange, 2009) and PCA analysis, as well as computing the F_{ST} between each regional group pair. We estimated splits and admixture events in our dataset using TREEMIX. To identify the most likely route taken by the first settlers of Sahul (i.e., southern vs northern route), we computed outgroup *f3*-statistics (Patterson et al., 2012) to show how genetically close New Guinean, northeast Indigenous Australian, and Bismarck Archipelago islander subgroups are to Wallacean subgroups located on the southern or the northern route, using African Yoruba as an outgroup. Additionally, we performed *f4*-statistic (Patterson et al., 2012) to test for gene flow from Wallacean subgroups located on the southern route and Wallacean subgroups located on the northern route to New Guinean, northeast Indigenous Australian and Bismarck Archipelago islander, using African Yoruba as an outgroup again. To explore the likelihood of different migration hypotheses, we tested several models using qpGraph (Patterson et al., 2012). We tested models that consider migrations from Sunda to Sahul along either the southern or the northern route and different possibilities of admixture between Wallacean and New Guinean groups. We also modelled demographic scenarios with one or two migrations from Sunda to Sahul. To improve the resolution of our demographic models, we incorporated a group from the Bismarck Archipelago and tested its relationship with New Guineans and northeast Indigenous Australians. Finally, to determine the most likely entry point into New Guinea, we used qpGraph to determine the best scenario for the settlement within New Guinea, fitting data from the five New Guinean groups and African Yoruba outgroups, with or without admixture events. To refine the various scenario models described in this paper, we also inferred the effective population size of these groups and the divergence dates between them using MSMC2 (Schiffels & Durbin, 2014).

4.2. Ref II

All data used in this publication has been obtained from the literature. It used whole genome sequences from 239 Oceanians. In addition to the 58 PNG sequences from Ref I (Brucato et al., 2021), this study also included populations from Wallacea, New Guinea, and the Bismarck Archipelago (Jacobs et al., 2019; Malaspinas et al., 2016; Mallick et al., 2016; Vernot et al., 2016). To increase the sample size of the New Guinean population, we also included previously published imputed genomes from PNG (Bergström et al., 2017). We added 330 whole-genome sequences from non-Oceanian populations (Jacobs et al., 2019; Mallick et al., 2016) to compare our dating results from GEVA with the initial publication. We used GEVA v1.beta (Albers & McVean, 2020) (Genealogical Estimation of Variant Age; described in the literature review) to estimate the distribution of coalescence ages of SNPs. We performed XP-EHH and XP-nSL analysis using Selscan v.1.3.0 (Szpiech & Hernandez, 2014). To detect signatures of selection unique to New Guineans, Bismarck Islanders, and Wallaceans, we generated these selection scans for three target and reference population pairs: New Guinea compared to Wallacea, Bismarck Archipelago compared to Wallacea and the Bismarck Archipelago compared to New Guinea. PBS scores (Yi et al., 2010) were generated from F_{ST} distances inferred with vcftools v0.1.15 (Danecek et al., 2011). We considered New Guinea or the Bismarck Archipelago as the target population, Wallacea as the reference population and used African genomes from the SGDP dataset as an outgroup. We also perform a PBS analysis defining the Bismarck Archipelago as the target population, New Guinea as the reference population and Africa as the outgroup. Finally, we combined the XP-EHH, XP-nSL and PBS for each SNP in a Fisher Score (Lopez et al., 2019). We estimated the significance of enrichment of selection signals in 1000 years time frames using coalescent time estimates from GEVA for the SNPs with outlier Fisher Score. Finally, we estimated the archaic introgression in the genomes from Wallacea, New Guinea and the Bismarck Archipelago using Skov's HMM method (Skov et al., 2018), which does not require an archaic reference. We also analysed adaptive introgression using the Q95 statistics (Choin et al., 2021; Racimo et al., 2015).

4.3. Ref III

This paper relies on phenotypes measurements performed on PNG highlanders living in Mount Wilhelm between 2,300 and 2,700 m a.s.l. and PNG lowlanders living on Daru island around 100 m a.s.l. Phenotypes were collected on healthy adults. The studied phenotypes were standing height, weight, body mass index (BMI), waist circumference, minimal and maximal chest depth, haemoglobin concentration, heart rate, and diastolic and systolic blood pressure. Spirometer measurements reflecting lungs capacities were also recorded: the volume of air

that is forcibly blown out in the first second after full inspiration (forced expiratory volume in 1 second, FEV1), the maximal flow achieved during the maximal forced expiration initiated at full inspiration (peak expiratory flow, PEF) and the volume of air that is forcibly blown out after complete aspiration (forced vital capacity, FVC). To consider only PNG highlanders with minimal lowland admixture and reciprocally, we excluded all individuals with at least one parent or grandparent not originating from the same altitude level. After this step, we kept 70 highlander and 86 lowlander individuals. We corrected all the phenotypes for age and sex by computing residuals for the corresponding multilinear regression. We also regressed minimal and maximal chest depth, waist circumference, weight and spirometer measurements for height that could be a covariate to these phenotypes. Finally, we compared the residuals of the phenotypes between PNG highlanders and lowlanders with a Mann-Whitney U test, a non-parametric test. The measured phenotypic traits analysed in this study are reportedly adaptive in other highlander groups worldwide (Cynthia M. Beall, 2013; Bigham & Lee, 2014). We then compared our results with previous work on other highlander populations.

4.4. Ref IV

In this publication, we used newly sequenced whole genomes from three villages in Mount Wilhelm located in PNG highlands (n=54, 2300 to 2700 meters a.s.l.) and from Daru Island (n=74, 100 meters a.s.l.). We aimed to detect genomic regions showing signs of positive selection in PNG highlanders and lowlanders. To achieve this, we used three different scores: XP-EHH (Sabeti et al., 2007), PBS (Yi et al., 2010), and a Fisher Score, which combines XP-EHH and PBS scores (Lopez et al., 2019). We used PNG highlanders as the target population and PNG lowlanders as the reference population and vice versa for each score. We included Yorubas from the 1000 Genomes project (The 1000 Genomes Project Consortium et al., 2015) as the outgroup when computing the two PBS scores. With this approach, we identified signatures of selection unique to PNG highlanders and lowlanders. We considered ten genomic regions of interest with the highest scores for each score. Next, we inferred ancestral recombination graphs for the full dataset, using the two PNG populations and the 58 sequences from Ref I (Brucato et al., 2021) and Asian, European and African genomes from The 1000 Genomes Project (The 1000 Genomes Project Consortium et al., 2015) with RELATE (v1.1.8) (Speidel, Forest, Shi, & Myers, 2019). We extracted the local tree for each SNP of the genomic regions under selection from PNG highlanders and lowlander subtrees. We used these local trees as input for CLUES (Stern et al., 2019). CLUES assigned a likelihood ratio (logLR) which indicates the support of the non-neutral model for each tested SNP.

The log LR was computed five times for each SNP, and the average of the five runs was taken. To decide between the top five SNPs, the logLR was generated 50 more times, and the SNP with the highest average logLR (a.k.a. candidate

SNP) was considered the most likely SNP to drive selection within the region of interest. Defining the candidate SNPs allowed us to look for the associated phenotypes in the UK biobank (Pan-UKB team, 2020). We extracted the p-value and the beta for European ancestry from the UK biobank summary statistics for the candidate SNPs and 1,931 phenotypes with more than 10,000 samples. When candidate SNPs were not listed in the UK biobank, we looked for summary statistics for the closet SNP from a 1kb upstream and 1kb downstream region. We considered the association of a candidate SNP with a phenotype of interest if the log_{10}(p-value) is lower than –11.29 to correct for multiple testing. We chose this threshold base on the standard significance p-value threshold in GWAS of 5e-8 (Dudbridge & Gusnanto, 2008) and adjusted it for the 1931 different phenotypes we used in the summary statistics following the Bonferroni correction guideline (Shaffer, 1995).

Moreover, we compared these associations from the UK biobank with associations found for phenotypes measured directly in this publication's two studied PNG populations and published in Ref III (André et al., 2021). We tested if the candidate SNPs were related to any PNG phenotypes using a statistical model called Linear Mixed Model (LMM) with the Genome-wide Efficient Mixed Model Association (GEMMA) (v0.98.4) software (X. Zhou & Stephens, 2012). To increase our sample size of genotypes and phenotypes, we performed these association tests using PNG highlanders, PNG lowlanders, and PNG samples from Port Moresby published with Ref I (Brucato et al., 2021). We incorporated a centred relatedness matrix generated with GEMMA for all the PNG sequences into the LMM to correct for population stratification. We adjusted the p-values of the SNPs to account for multiple tests using the Benjamini-Hochberg procedure (Benjamini & Hochberg, 1995). We grouped the phenotype into the five categories of highly correlated phenotypes defined in Ref III (André et al., 2021): body proportion group (height, weight, waist circumference, BMI), pulmonary function group (chest depth, spirometer measurements), haemoglobin concentration, blood pressure (diastolic and systolic blood pressure) and heart rate. We adjusted the p-value threshold with Bonferroni correction for the five independent phenotype groups tested ($p\text{-value}_{adjusted}$=0.05/5) to determine statistical significance.

As an additional effort to decipher the function of the candidate SNPs, we looked for significant eQTLs for each candidate SNP using the Genotype-Tissue Expression (GTEx) Portal (Lonsdale et al., 2013). Moreover, we used the Ensembl Variant Effect Predictor (VEP) (McLaren et al., 2016) on the regions under selection to detect missense variants in these regions.

Finally, we used haplostrips (v1.3) (Marnetto & Huerta-Sánchez, 2017) to explore similarities between PNG haplotypes and archaic haplotypes for the genomic regions under selection in PNG highlanders and lowlanders. We also analysed archaic allele frequencies in the PNG from the SGDP dataset (Mallick et al., 2016) in the regions of interest with introgressed haplotypes in PNG highlanders and lowlanders.

5. RESULTS AND DISCUSSION

This section summarises the main results and discussions for the four scientific publications constituting this dissertation's original part. The following is a compressed summary of the main results and discussions. More detailed information can be found in the original articles and their respective supplementary information.

5.1. Models for northern Sahul settlement based on genomic data (Ref I)

5.1.1. Genetic structure characterization:

We identify our dataset's main regional genetic groups using fineSTRUCTURE (Lawson et al., 2012). We found that individuals in our dataset clustered into seven main groups, corresponding to their respective geographical locations, including the New Guinean southeast lowlands, New Guinean north and south lowlands, New Guinean east highlands, New Guinean west highlands, northeast Australia, Bismarck Archipelago, and Wallacea. PCA, Admixture and F_{ST} results reflect a strong geographical structure of the human genetic diversity in Sahul. However, they also indicate a relative genetic continuity between New Guinea and Wallacea, implying that the latter was not an insurmountable barrier (Figure 11B).

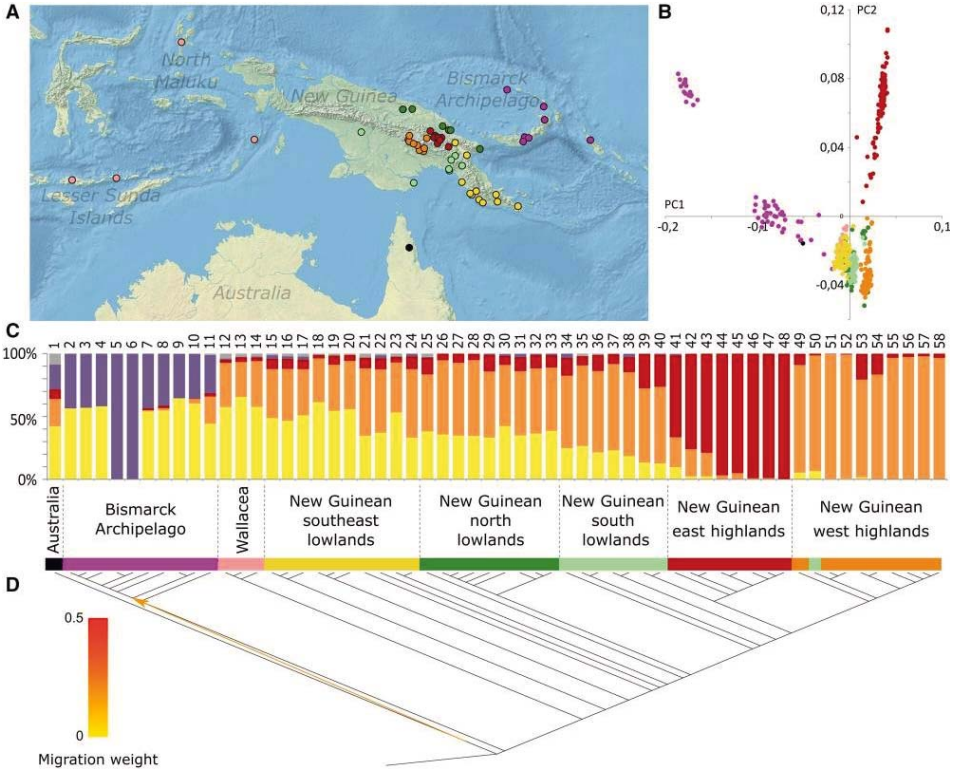

Figure 11 : Genetic diversity of individuals from New Guinea, Wallacea, the Bismarck Archipelago, and northeast Australia. (**A**) Geographic locations of all 58 genetic subgroups defined by the fineSTRUCTURE analysis. (**B**) Principal component analysis representing 26.7% of the total variability of all individual genomes from the 58 subgroups, masked for Asian genetic ancestry. The two first components are represented (PC1 and PC2). (**C**) ADMIXTURE plot for eight genetic components summarized by genetic subgroups (based on K = 8). The grey component combines four components mainly present in African and Eurasian groups. (**D**) TREEMIX plot for all 58 subgroups. All four panels used the same colour scheme to identify the 58 genetic subgroups. Reprinted from Figure 1 (Brucato et al., 2021), licensed under Creative Commons CC BY license. Copyright © 2021, Brucato et. al.

5.1.2. Identification of gene flow

We obtained an optimal TREEMIX model (Pickrell & Pritchard, 2012) when including five gene flows, one of which is from ancestors of northeast Indigenous Australians to a group ancestral to most Bismarck Archipelago islanders. (Figure 11D). This gene flow might explain the component shared between Indigenous Australians and Bismarck Archipelago Islanders observed in the ADMIXTURE plot (Figure 11C). The tree also shows that all the New Guinean subgroups branch together while its root lies in the genetic diversity of New Guinean southeast lowlanders (Figure 11D). Additionally, we observed that subgroups from the New Guinean north and south lowlands constituted two distinct genetic clusters that had genetically diverged from the New Guinean southeast lowlanders (Figure 11D). These results suggest that New Guinean north and south lowlanders originated from separate migrations that were initiated from the southeast region of New Guinea.

5.1.3. Modelling the genetic scenario of the migration in Wallacea

F3-statistics (Patterson et al., 2012) results show that all subgroups from New Guinea, northeast Australia, and the Bismarck Archipelago are equally distant to any Wallacean subgroups located on the southern route to Sahul. However, *f3*-statistics increase when using a Wallacean subgroup located on the northern route instead of Wallacean subgroups from the southern route. These results suggest a stronger genetic link of the Wallacean subgroup located on the north route with all studied subgroups than those on the southern route. *F4*-statistics (Patterson et al., 2012) support that most populations from New Guinea and the Bismarck Archipelago are significantly genetically closer to the Wallacean population located on the northern Route (*f4*(Yoruba, X; southern Wallacea, northern Wallacea)>0.001, Z score > 3). Finally, this genetic link is particularly pronounced in New Guineans: *f3*-statistics and *f4*-statistics values were higher and more significant for New Guineans with north Wallacean subgroups than for the Bismarck Archipelago and northeast Australia.

Two non-exclusive hypotheses might explain the more important genetic link between the north Sahul populations and northern Wallacea subgroups. The first hypothesis is that the northern route was the most likely path for the settlement of Sahul. Alternatively, there may have been secondary gene flows between north Wallacea and northern Sahul groups. To explore the likelihood of these hypotheses, we tested multiple migration and admixture scenarios using qpGraph (Patterson et al., 2012). We obtained two models fitting the data (Figure 12). The first model considers a unique migration from Sunda to Sahul (Z=–2.558). In this model, New Guineans and Indigenous Australians share a common ancestor more recently than Wallaceans. The split from Wallaceans is followed by two important gene flows from New Guinea to Wallacea.

The second model fitting the data displays two migrations from Sunda to Sahul (Z = 2.337). In this model, northern Indigenous Australians and New Guineans descend from groups that have migrated to Sahul independently. These migrations are followed by a gene flow of the ancestors of New Guineans to northern Wallacean groups. This model suggests that New Guineans' ancestors have taken the northern route, which is consistent with previous modelling studies (Bird et al., 2019; Bradshaw et al., 2019; Kealy et al., 2018).

5.1.4. Modelling the genetic scenario of the settlement of Sahul

We improved the previously described qpGraph models for the settlement of Sahul by incorporating a subgroup from the Bismarck Archipelago into the previous two scenarios. For both models, with one (model A, Z=–2.570) or two migrations (model B, Z=2.301) to Sahul, we obtained models fitting the data when the Bismarck Archipelago group results from a separate admixture event between ancestral New Guineans and northeast Indigenous Australian (Figure 12) which correspond to the gene flow between ancestral indigenous Australians group and ancestral Bismarck Archipelago group detected by Treemix. This might again support the continuity hypothesis between groups from Wallacea to New Guinea but stronger genetic barriers between New Guinea and Australia and around the Bismarck Archipelago.

5.1.5. Modelling the genetic scenario of the settlement of New Guinea

We computed qpGraph models to determine the best model for the settlement of New Guinea using the five New Guinean groups (southeast lowlanders, south lowlanders, north lowlanders, east highlanders, and west highlanders) and the African Yoruba outgroup. The unique model fitting our data (Z score =–0.06) presents the southeast lowlands as the most likely entry point to northern Sahul. When adding the west and east highlander groups, the best-fitting model places west highlanders with New Guinean south lowlanders (Z score = 2.759) (Figure 12). *F3*-statistics show a closer link between New Guinean west highlanders and New Guinean south lowlanders. In addition, New Guinean east highlanders are almost at a similar genetic distance to New Guinean south and north lowlanders, which suggests that New Guinean east highlanders came from a place between south and north lowlands. This final model identifies the southeast lowlands as the entry point into New Guinea. Interestingly, New Guinean southeast lowlands correspond to the location of the oldest known archaeological site in New Guinea (45–49kya) (Summerhayes et al. 2010). This supports an initial migration via the ancient plain currently under the Arafura Sea. Our results indicate that this initial migration was followed by two migrations to the south and north lowlands. Later, the New Guinean highlands were settled by two separate migratory waves: one migration from the east lowlands to the east highlands and another from the south

lowlands to the west highlands. We completed this scenario of settlement of New Guinea using MSMC2 (Schiffels & Durbin, 2014). We found a similar divergence time between New Guinean populations (7.6–12.8kya) than previously published results (Bergström et al., 2016). These relatively recent divergence times probably reflect long-term contact between New Guinean populations after the initial settlement.

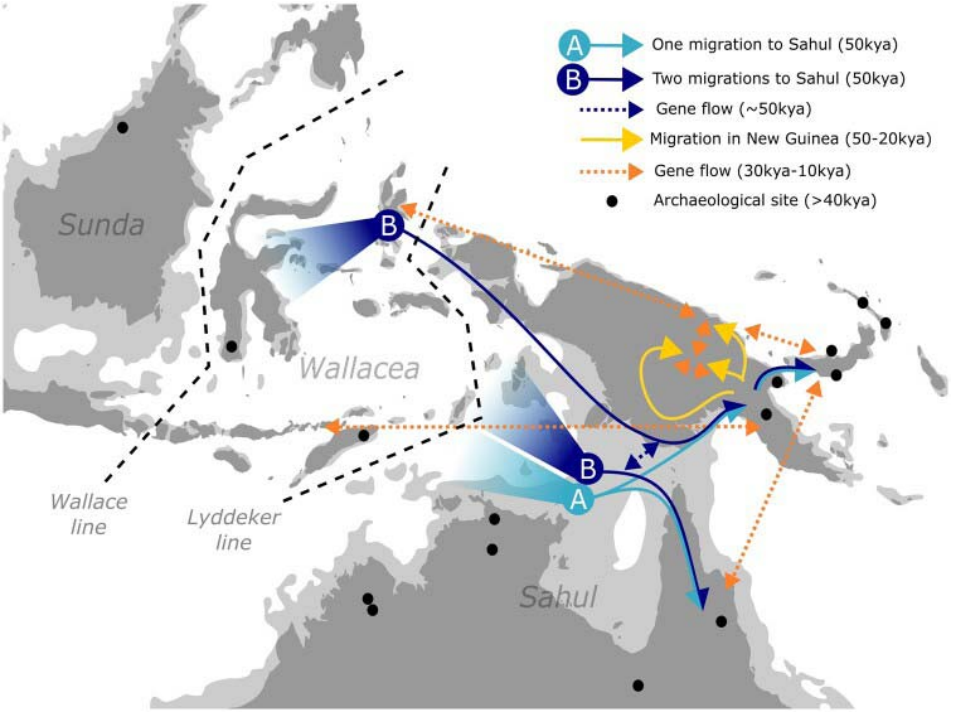

Figure 12: Genomic scenario of the human dynamics in north Sahul during the Upper Pleistocene. Light blue arrows represent the settlement of north Sahul following a scenario with one migration to Sahul (model A). The dark blue arrows represent the settlement of north Sahul following a scenario with two migrations to Sahul (model B). The dotted dark blue arrow represents a gene flow following a scenario with two migrations to Sahul (model B). Yellow arrows represent the settlement of New Guinea. The dotted orange arrows represent gene flows after the initial settlement of Sahul. The black dots represent approximate locations of archaeological sites older than 40 ka (O'Connor, 2007; Summerhayes et al., 2017). The dashed black lines represent ecological lines in Wallacea. The dark grey areas represent the current land area, with light grey areas representing the estimated land area around 50,000 years ago based on paleogeological reconstructions (Coller, 2009). Reprinted from Figure 4 (Brucato et al., 2021), licensed under Creative Commons CC BY license. Copyright © 2021, Brucato et. al.

5.2. Time of selection in Oceanian genomes (Ref II)

5.2.1. Dating the Oceanian genetic diversity

First, we showed that the age of coalescence estimates obtained using GEVA on SGDP whole-genomes from Africa, Asia, and Europe, strongly correlated with the previously published estimates (Albers & McVean, 2020; Mallick et al., 2016). To limit the influence of recent Asian ancestry, we defined the ages of coalescences only for SNPs specific to New Guinea, Bismarck Archipelago or Wallacea populations by excluding SNPs found in the other populations in our dataset with several Asian genomes. The median age of the SNPs specific to Oceanians with a frequency of 0.25 (78,795 Years Before Present (YBP) IQR: 47,383-225,075 YBP, Figure 13B) is older than the median age of private SNPs at the same frequency in other non-African populations. This result is coherent with previous studies, including Ref I (Brucato et al., 2021), on the divergence time of Oceanians from Africans and reflects the extended time these populations have been isolated from other continents (Malaspinas et al., 2016; Pagani et al., 2016).

Most SNPs private to populations from New Guinea, Bismarck Archipelago or Wallacea appeared in the last 10,000 years (Figure 13C), which might correspond to the demographic expansion observed worldwide during the Last Glacial Maximum era (Bergström et al., 2020). We observed that this increase in diversity occurs slightly earlier in New Guineans. This earlier increase in diversity in New Guineans might result from their early farming practices.

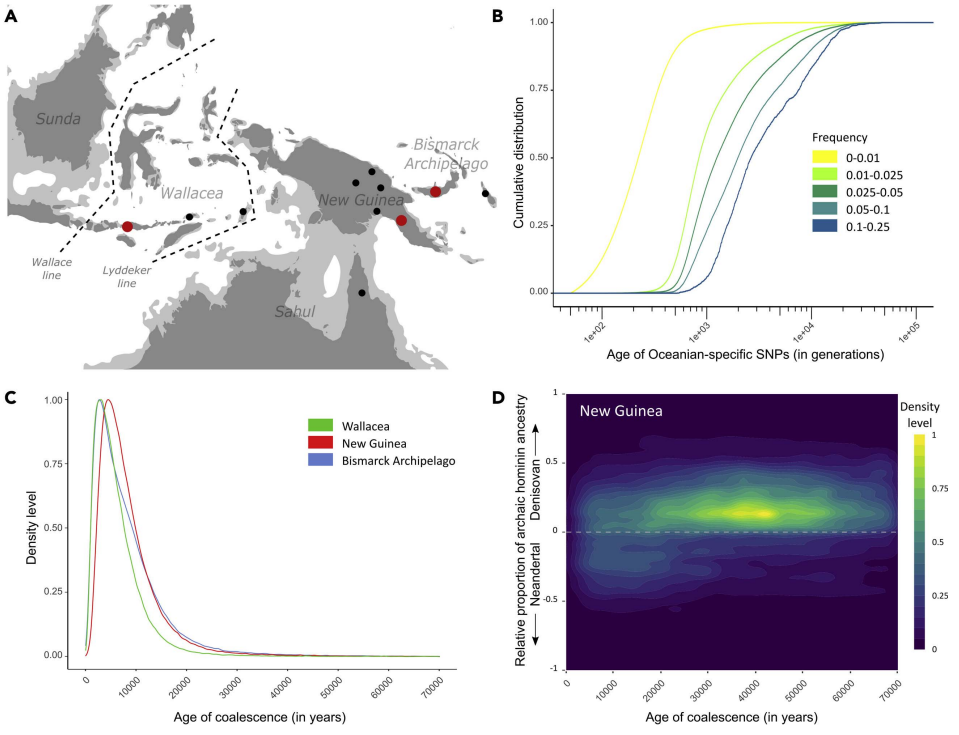

Figure 13: Distribution of coalescence ages in Oceanian genomes. (**A**) Map of analyzed populations. Red dots localize the three studied Oceanian regional groups in Wallacea (Flores Island), New Guinea (Papua New Guinean southeast coast), and the Bismarck Archipelago (New Britain). Black dots represent other groups included in our dataset. (**B**) Cumulative distribution of coalescence ages estimated with GEVA for Oceanian-specific SNPs binned by frequency. (**C**) Density plot of coalescence ages for SNPs specific to each Oceanian regional group. (**D**) Density plot of coalescence ages for SNPs present in archaic fragments in New Guinean genomes. The relative proportion of archaic ancestry, estimated with Skov's HMM method, is calculated as (Denisovan ancestry – Neanderthal ancestry)/(Denisovan ancestry+ Neanderthal ancestry). Reprinted from Figure 1 (Brucato et al., 2022), licensed under Creative Commons CC BY license. Copyright © 2022, Brucato et. al.

5.2.2. Relative and absolute dating of the signal of selection

We defined SNPs significantly under selection in New Guineans and Bismarck islanders with a Fisher Score (Fxp) (Lopez et al., 2019) combining XP-EHH (Sabeti et al., 2007), XP-nSL (Szpiech & Hernandez, 2014) and PBS (Yi et al., 2010). We used Wallaceans as the reference population and African genomes from SGDP as the outgroup population for PBS. Because we used Wallaceans as the reference population, the Fisher Score of New Guineans and Bismarck Islanders characterise genomic regions under selection present only in New Guineans or Bismarck Islanders, respectively, but not in Wallaceans. In other words, we used Wallaceans as the reference population to mask signatures of selection in New Guineans and Bismarck islanders dating before the initial settlement of north Sahul. We used this approach to define a relative timing for selection between the crossing of Wallacea and the settlement of northern Sahul (Figure 14C, D). We also performed the absolute dating of the SNPs under selection using GEVA (Figure 14A, B).

When coupling these two dating approaches, we observed a significant enrichment of selection signals for SNPs dating between 52,000 and 54,000 YBP in both New Guinean and Bismarck Archipelago genomes ($Z > 6$, p-value < 10–9, Figure 14). This timing coincides with the supposed time of the first settlement of Sahul (Summerhayes et al., 2017). This convergence in the time period with enrichment for SNPs under selection in New Guineans and Bismarck, despite the two populations being highly divergent, would suggest that common ancestors of New Guineans and Bismarck Archipelago Islanders faced strong selection pressure around this time period.

The results of the selection scans on the genomes of New Guineans and Bismarck Islanders suggest that selection during the settlement of Sahul mainly affected genes related to the immune system and diet. One of the strongest selection signal in both populations overlaps with the protein-coding gene *ZDHHC20,* which is involved in antiviral response (McMichael et al., 2017). Such selection signals in New Guineans and Bismarck Islanders might result from an ancient viral pathogenic pressure.

Other significant signals of selection in the genomes of New Guineans and Bismarck Islanders overlap genes related to food intakes, such as *RARB*, *KCNMB2*, and *POR*, suggesting that diet (e.g., intake of Vitamin A and folic acid) was a major selective force. The diet of modern New Guineans and Bismarck Islanders considerably relies on the local plant. Selection acting on the first settlers of northern Sahul probably favoured genes supporting the use of the local plant resources of these new territories.

We also observed a significant enrichment for SNPs with signal of selection unique to Bismarck Islanders between 42,000 and 43,000 YBP ($Z > 6$, Figure 14B) when using this population as the target population and New Guineans or Wallaceans as the reference population. This corresponds to the first settlement of the archipelago (Summerhayes et al., 2017).

Moreover, New Guinean and the Bismarck Islanders, but not the Wallaceans, have a significantly lower number of selection signals for SNPs dating between 10,000 and 30,000 YBP ($Z < -6$, p-value $< 10^{-9}$, Figure 14A, B). This lower enrichment would be specific to the descendants of the first settlers of north Sahul and would correspond to the reduced demographic dynamics during the Last Glacial Maximum (Pedro et al., 2020; Summerhayes et al., 2017)

Overall, these results show that the first settlers of northern Sahul were exposed to strong selective pressure, likely associated with exposure to a new diet, rapidly after they reached New Guinea and Bismarck.

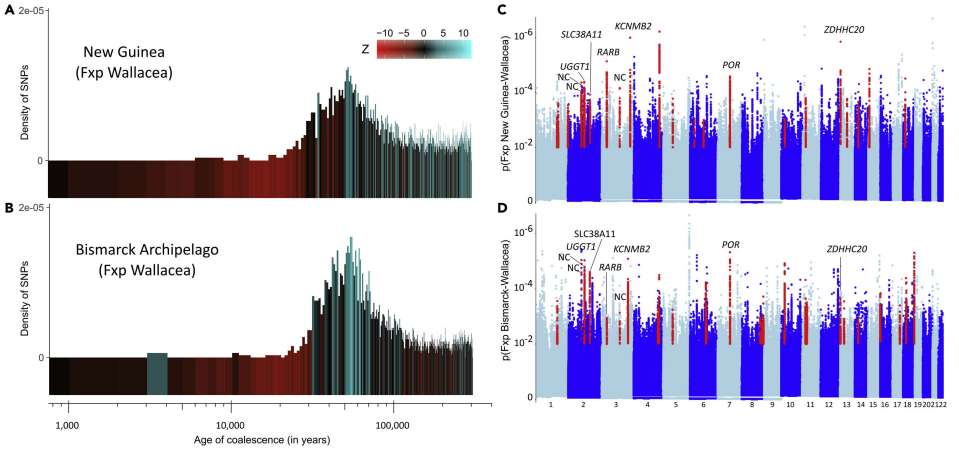

Figure 14: Distribution of coalescence ages for SNPs under selection in the genomes of New Guineans and Bismarck Archipelago Islanders. (**A** and **B**) Density plot of coalescence *ages* for all significant Fxp signals of selection in New Guinea (**A**) and the Bismarck Archipelago ***(B)***, using Wallacea as reference ($p < 0.01$), binned by 1,000 years. The blue-red gradient represents the Z score of enrichment of SNPs with a significant Fxp signal of selection in comparison to the rest of the genome (resampled 1,000 times) for each 1,000-year bin. (**C** and **D**) Manhattan plots of the Fxp p-values for each SNP in the analyses: (***C***) Fxp New Guinea vs Wallacea and (***D***) Bismarck Archipelago vs. Wallacea. Red dots represent variants showing significant Fisher scores in 100kb windows enriched for significant scores (>0.7). Gene names represent loci including at least one SNP with a coalescence age between 52,000 and 54,000 YBP. Reprinted from Figure 2 (Brucato et al., 2022), licensed under Creative Commons CC BY license. Copyright © 2022, Brucato et. al.

5.2.3. Introgression

All New Guinean, Bismarck and Wallacean populations show a higher rate of Denisovan than Neanderthal introgression. Most SNPs introgressed from Denisovan have a coalescence time falling between 35,000 and 45,000 YB. This coalescence age is in the time frame of the admixture events with Denisovans previously reported for the Oceanian region (30,000–46,000 YBP) (Jacobs et al., 2019; Malaspinas et al., 2016). When refining our approach, we noticed that many SNPs introgressed from Denisovans and private to New Guineans have an even later coalescence time of around 42,000 YBP. Nonetheless, most archaic SNPs in any of these three Oceanian populations have a coalescence time of about 10,000 years, suggesting that they were not inherited from Neanderthals or Denisovans.

5.2.4. Introgression and selection

Although the SNPs significantly under selection in New Guineans, Bismarck Islanders, and Wallaceans did not show an overall enrichment for Denisovan ancestry ($Z<2$, p-value >10-2), we found the highest proportion of Denisovan ancestry in SNPs under selection between 52,000–53,000 YBP in New Guineans and 39,000–40,000 YBP in Bismarck Islanders. These results suggest that the SNPs under selection were present in Denisovan introgressed fragments during the main phases of settlement in the New Guinea region.

Of the loci enriched for Denisovan ancestry and under selection in Wallacea, New Guinea and Bismarck populations (e.g., *TNFAIP3*, *IFNGR1*, the *GBP* locus and the *SIGLEC* locus), most have already been reported previously (Jacobs et al., 2019; Sankararaman et al., 2016; Vernot et al., 2016). Interestingly, none of the three populations shows an elevated Fisher score for any of these loci, suggesting that their selection predated the initial settlement of northern Sahul.

It is to be noted that we found a strong signal of selection in a locus enriched for Denisovan ancestry and overlapping with the gene *RARB*. The gene *RARB* encodes a receptor which is a vitamin A derivative and is involved in adaptive intestinal immunity (Bang et al., 2021; Mora et al., 2006). This suggests that diet and, more specifically, the intake of vitamin A were major selective forces.

5.3. Phenotype differences between PNG highlanders and lowlanders (Ref III)

5.3.1. Multitesting correction

We compared 13 phenotypes between PNG lowlanders and highlanders using the Mann-Whitney U test. Because we were performing multiple comparisons, which multiple the risk of false positive results, we adjusted the significance threshold of 0.05 for the number of independent tests performed following the Bonferroni correction to control for type I error. Among the 13 studied phenotypes, some were highly correlated, meaning that tests for the correlated phenotype were not independent. Adjusting the p-value threshold for the 13 phenotypes would lead to a loss of power (Shaffer, 1995). We defined five groups of highly correlated phenotype using the Spearman correlation coefficient (Figure 15): Body proportion group (height, weight, waist circumference, BMI), Pulmonary function group (chest depth, spirometer measurements), haemoglobin concentration, blood pressure (diastolic and systolic blood pressure) and heart rate. We considered that differences between PNG highlanders and lowlanders were significant if the p-value was lower than 0.05 after Bonferroni correction for five independent tests giving an adjusted p-value of 0.01.

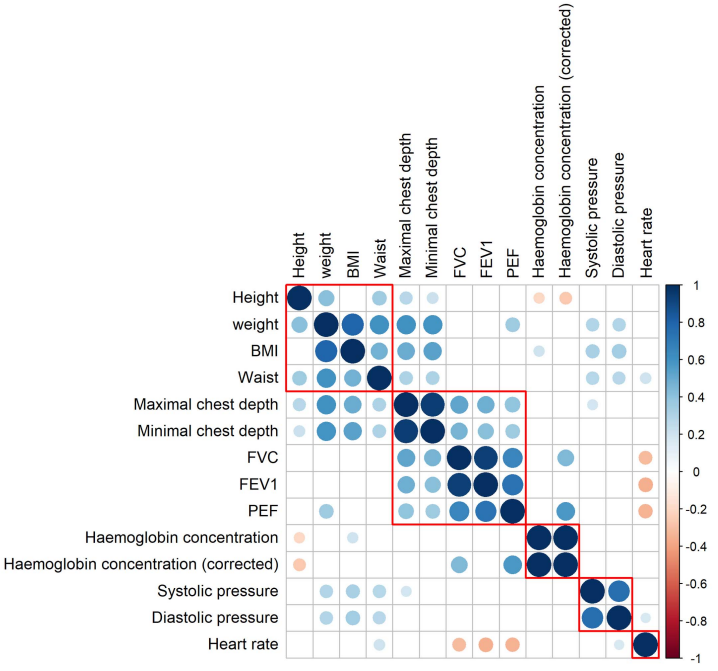

Figure 15: Correlation matrix between the studied phenotype measurements. Spearman correlation was used. Only significant correlations (p-value <0.05) are shown. Five groups of correlated phenotypes are enclosed in red boxes. Reprinted from Figure S1 (André et al., 2021), licensed under Creative Commons CC BY license. Copyright © 2021, André et. al.

5.3.2. Exploring phenotypic differences between PNG highlanders and lowlanders

PNG highlanders are significantly shorter (mean: 164.6 cm [95% CI 163.21–166.03]) than lowlanders (mean: 168.8 cm [95% CI 166.84–170.72]) while having significantly deeper chests ($p\text{-value}_{depth\ min}$ = 2.37e –5, $p\text{-value}_{depth\ max}$ = 3.20e -4). Both theses phenotypes have been shown to differ in a similar direction between Andean, Tibetan and Ethiopian highlanders and the respective lowlander population they are traditionally compared with (Table 1) (Brutsaert et al., 1999; Droma et al., 1991; Eichstaedt et al., 2015; G. A. Harrison et al., 1969; Mueller et al., 1978; Rupert & Hochachka, 2001; Weitz, Garruto, Chin, & Liu, 2004). The combination of larger chest size and shorter height has been associated with enhanced lung capacities in Tibetans and Andeans (Hall & Guyton, 2011). Similarly, we also observed a significantly higher FVC (p-value = 0.008), a proxy for lung volume, in PNG highlanders (mean: 4.09 L [95% CI 3.78–4.40]) in comparison with PNG lowlanders (mean: 3.32 L [95% CI 3.02–3.63]). The significantly increased pulmonary volume observed in highlanders from Papua New Guinea may suggest that they have adapted to lower oxygen availability in the highlands. We also reproduced results from another study showing significantly greater haemoglobin levels (p-value = 5.51e-4) in PNG highlanders (mean: 14.45 g/dl [95% CI 13.84–15.06]) than in PNG lowlanders (mean: 12.63 g/dl [95% CI 12.07–13.20]) (Senn et al., 2010). We further show that this observation remains true (p-value=3.36e-4) even when excluding individuals suffering from anaemia following the WHO standard cut-offs (Department of Nutrition for Health and Development (NHD), 2011; Silubonde et al., 2020). This difference then does not result from the difference in malaria incidence between the highlands and the lowlands. While Andean and Ethiopian highlanders also display elevated haemoglobin levels, Tibetans show a reduced haemoglobin concentration. Moreover, high haemoglobin concentration is also observed in acclimatised lowlanders living at high-altitude and is associated with increased blood viscosity, which is maladaptive in the long term. An increased level of haemoglobin concentration is observed in individuals born with low-altitude ancestry and living at high-altitude and facing lower levels of oxygen saturation caused by hypoxia. Surprisingly, PNG highlanders do not differ from lowlanders for any cardiovascular phenotypes. Because native lowlanders usually show increased heart rate and blood pressure at high-altitude (Brito et al., 2007), the absence of such an increase in PNG highlanders may be another adaptative trait counteracting the detrimental effect of hypoxia.

Table 1 : Phenotypic comparison between PNG highlanders and lowlanders. Modified from Table 1 (André et al., 2021), licensed under Creative Commons CC BY license. Copyright © 2021, André et. al.

	PNG lowlanders N= 86	**PNG highlanders N= 70**		**Mann-Whitney U test residuals**	
	Mean [95% CI]	**Mean [95% CI]**	**raw data**	**age, sex**	**age, sex, height**
Body proportions					
Height (cm)	**168.8 [166.84–170.72]**	**164.6 [163.21–166.03]**	**0.001***	**0.001***	–
Weight (kg)	68.46 [65.86–71.15]	64.82 [63.2–66.53]	0.137	0.198	0.87
BMI (m²/kg)	24.04 [23.19–24.97]	24 [23.33–24.76]	0.661	0.425	–
Waist circumference (cm)	**91.77 [89.44–94.16]**	**83.54 [81.99–85.17]**	**<.001***	**<.001***	**0.001***
Respiratory capacities					
Minimal chest depth (cm)	**19.16 [18.66–19.68]**	**19.91 [19.43–20.41]**	**0.033**	**<.001***	**<.001***
Maximal chest depth (cm)	**20.50 [20.00–21.01]**	**21.01 [20.47–21.56]**	**0.146**	**0.003***	**<.001***
Forced vital capacity (FVC) (L)	3.32 [3.02–3.63]	4.09 [3.78–4.40]	0.005§	0.020§	0.008
Forced expiratory volume after 1 second (FEV1) (L)	2.72 [2.45–3.00]	3.31 [3.01–3.61]	0.008§	0.177§	0.124
Peak expiratory flow (PEF) (L/min)	6.09 [5.28–6.92]	6.32 [5.39–7.34]	0.747§	0.638§	0.962
Circulatory system					
Haemoglobin concentration (g/dL)	**12.63 [12.07–13.20]**	**14.45 [13.84–15.06]**	**<.001***	**<.001***	–
Haemoglobin concentration (g/dL)**	**14.81 [14.31–15.33]**	**16.07 [15.65–16.49]**	**0.001§***	**<.001§***	–
Systolic pressure (mmHg)	131.76 [126.14–137.23]	123.94 [117.48–129.93]	0.071	0.169	–
Diastolic pressure (mmHg)	85.29 [82.33–88.26]	81.11 [77.21–84.99]	0.156	0.196	–
Heart rate (bpm)	67.77 [65.49–70.09]	64.38 [61.67–67.2]	0.042	0.03	–

Significant results are in bold. *: Significant Mann-Whitney U test with Bonferroni correction for 14 multiple tests (adjusted p-value = 0.004). §: data corrected to avoid ties. **: People suffering from anaemia were removed following the World Health Organization (WHO) standard cut-offs (exclusion of non-pregnant women < 12g/dl, men <13g/dl) (Department of Nutrition for Health and Development (NHD), 2011).

5.4. Differential positive selection in PNG highlanders and lowlanders (Ref IV)

5.4.1. Genomic regions under selection in PNG highlanders

PNG highlanders face several challenges related to hypoxia, including increased blood viscosity, a high risk of stroke, and neurodevelopmental disorders. To identify genomic regions with signatures of selection in PNG highlanders, we used PBS, XP-EHH and a Fisher Score, which combines the two first scores, with PNG highlanders as the target population, PNG lowlanders as the reference population and Yorubans from the 1000 Genomes as the outgroup population in PBS. We explored further the top ten regions in each of these scores. When merging the overlapping regions in these 30 top regions, we ended up with 23 regions under selection in PNG highlanders that overlap with genes that regulate platelet adhesion, HIF pathway, or neurodevelopment (Figure 16a,b,c). For example, the *FBLN1* gene regulates platelet adhesion (Godyna, Diaz-Ricart, & Argraves, 1996) and thrombosis (Tran et al., 1995). Selection for *FBLN1* in PNG highlanders might counteract the higher risk of thrombosis with increasing altitude because of increased blood viscosity (Ortiz-Prado & Dunn, 2011). The region with the highest Fisher Score overlaps the *LINC02388* intergenic RNA that may play a role in the formation of new blood vessels and increase available oxygen to tissues through regulation of the HIF-pathway (Gudjonsson et al., 2022; Peng et al., 2021; H. Zhou et al., 2021). Several genes regulating brain activity and causing various neurodevelopmental disorders when malfunctioning, such as *DLGAP1* (A. H. Rasmussen, Rasmussen, & Silahtaroglu, 2017), are also under positive selection in PNG highlanders. These genes might protect the brain from hypoxic damage. Additionally, we identified candidate genes under selection related to the immune system, including the Major Histocompatibility Complex (MHC) genomic region that contains HLA genes, which might play a central role in the adaptation of PNG highlanders' immune system.

5.4.2. Genomic regions under selection in PNG lowlanders

We explored 21 regions with signature of positive selection in PNG lowlanders. These regions were in the top 10 regions with the highest PBS, XP-EHH or Fisher Score, which combines the two first scores, when using PNG lowlanders as the target population, PNG highlanders as the reference population and Yorubans from the 1000 Genomes as the outgroup population in PBS. PNG lowlanders show strong selection in immunity-related regions, specifically in the guanine-binding protein (GBP) family locus. The genomic region overlapping these genes shows the highest XP-EHH, PBS and Fisher Score in PNG lowlanders (Figure 16a,d,e).

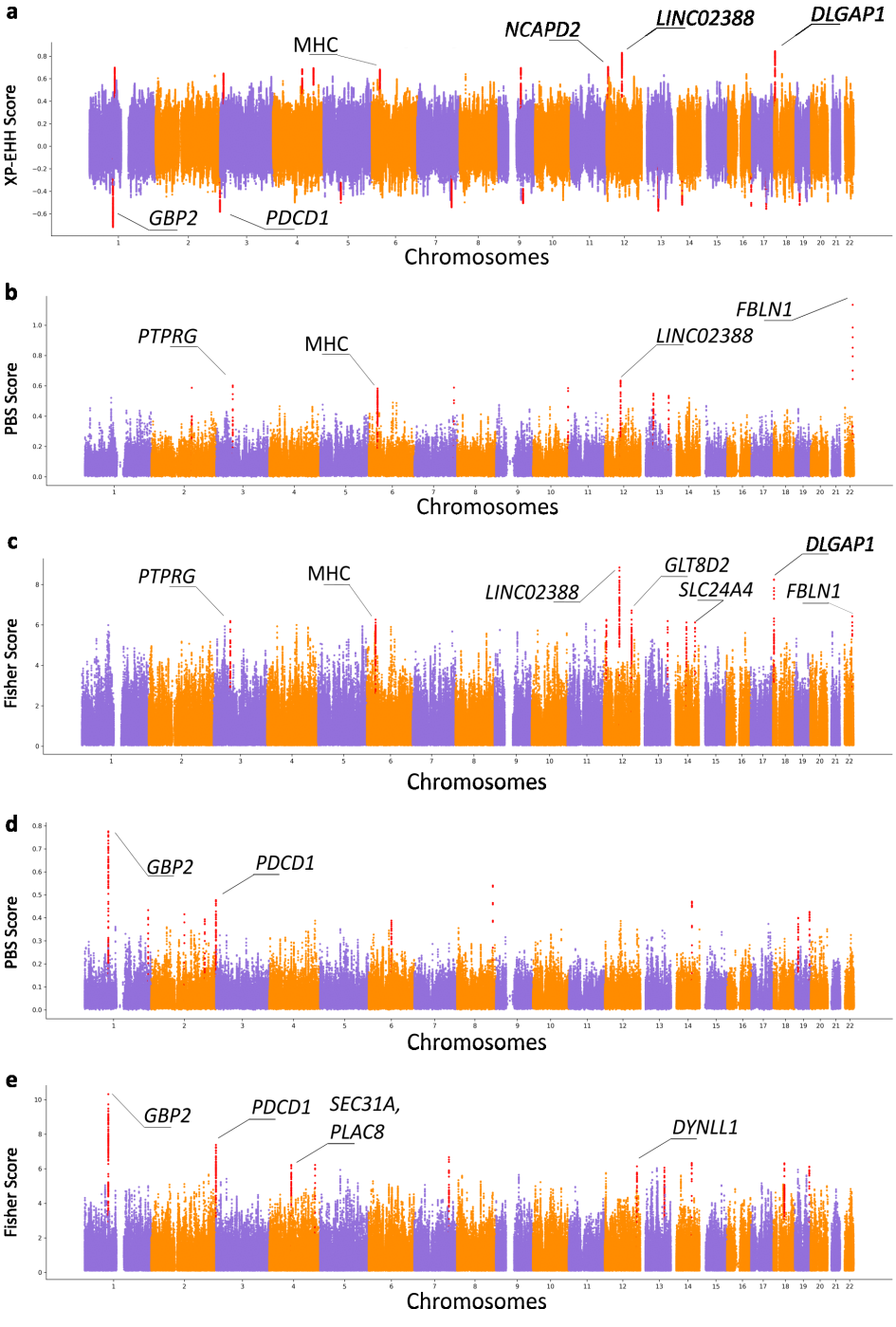

Figure 16: Manhattan plots for the three selection scans among PNG highlanders and lowlanders. Candidate genes discussed in the paper are shown. (**a**) XP-EHH scores using PNG highlanders as the target population and PNG lowlanders as the reference population. Genomic regions with the highest score indicate selection in PNG highlanders. Genomic regions with the lowest score indicate selection in PNG lowlanders.

(**b**) PBS scores using PNG highlanders as the target population, PNG lowlanders as the reference population, and Yorubas from 1000G as the outgroup. (**c**) Fisher Scores combining the PBS and XP-EHH scores of PNG highlanders. (**d**) PBS scores using PNG lowlanders as the target population, PNG highlanders as the reference population, and Yorubas from 1000G as the outgroup. (e) Fisher Scores combining the PBS and XP-EHH scores of PNG lowlanders. Reprinted from Figure 1 (André et al., 2022), licensed under Creative Commons CC BY license. Copyright © 2022, André et. al.

These genes protect against diverse pathogens (Tretina, Park, Maminska, & MacMicking, 2019). For example, one of the GBP7 variants is associated with the strength of malaria symptoms in Cameroon (Apinjoh et al., 2014). *NRF1*, a regulator of intracellular redox homeostasis (Hu et al., 2022; Y. Zhang & Xiang, 2016), may also have been selected in PNG lowlanders due to pathogenic pressure. Increased oxidative stress during the intraerythrocytic life cycle of the malaria parasite protects the host against severe malaria (Rahbari et al., 2017). We hypothesize these genes might have helped PNG lowlanders to face malaria infection.

5.4.3. Associations of the candidate SNPs with phenotypes in UK biobank and PNG populations

Seven candidate SNPs of PNG highlanders exhibit significant associations with at least one phenotype in the UK Biobank. Among these, three SNPs are significantly associated with haematological phenotypes. Similarly, eight candidate SNPs from PNG lowlanders demonstrate significant associations with phenotypes in the UK Biobank, with four associated with haematological phenotypes. Furthermore, we reported two candidate SNPs associated with heart rate measured in the PNG population (Figure 17). Specifically, the derived G-allele of rs74576183, an intronic variant of NCAPD2 under selection in PNG highlanders, is associated with a slower heart rate ($p\text{-value}_{adjusted} = 0.041$, beta=–2.981). On the other hand, the derived allele-T of rs4693058, an intronic variant of SEC31A under selection in PNG lowlanders, is associated with a faster heart rate ($p\text{-value}_{adjusted} = 0.041$, beta=3.137). However, these associations did not survive the correction for the number of phenotypes ($p\text{-value}_{adjusted} > 0.01$). It's worth noting that both these SNPs showed significant association with different type of haematological phenotype in the UK biobank, pointing out that the suggestive associations with heart rate may reflect an association with other haematological components that were not measured in our PNG samples. We hypothesize that selection acted on various haematological traits to optimize oxygen transport in PNG highlanders and to counteract the detrimental effects of malaria in PNG lowlanders, which also affects the red blood cells.

5.4.4. Functional roles of the candidate SNPs

We investigated the putative regulatory role and effects on the protein structure of the candidate SNPs under selection in PNG highlanders and lowlanders. Five candidate SNPs in PNG highlanders and three in PNG lowlanders, including the two candidates SNPs associated with heart rate, show significant eQTLs in various GTEx tissues. In addition, one of the regions under selection in PNG highlanders overlaps with one missense variant with an exceptionally high DAF in PNG highlanders (0.7) but below 0.12 in African, Asian or European populations. Moreover, this missense variant is in high LD (R2=0.952297) with the candidate SNP, rs74576183, associated with a slower heart rate (Figure 15).

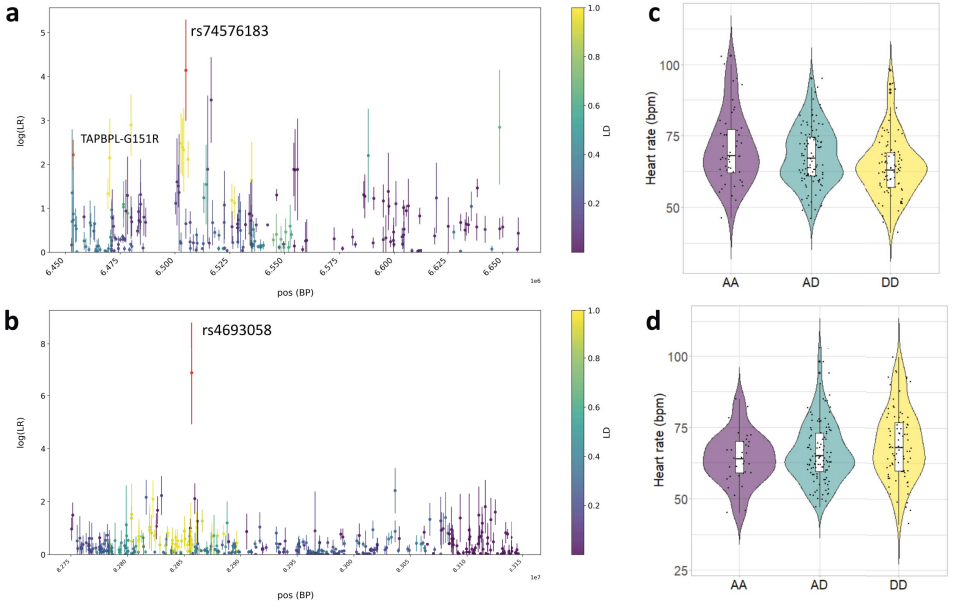

Figure 17: **Clues log(LR) and violin plots of the heart rate distribution depending on the genotype of the candidate SNP for the regions chr12:6452552-6662260 and chr4:82750503-83146792.** (**a, b**) log(LR) for SNPs in regions under selection after five runs of CLUES or 50 runs of CLUES for each of the five top SNPs in the candidate region. The candidate SNP driving selection for the region is shown in red. The colour scale indicates linkage disequilibrium with the candidate SNP. (**a**) Region chr12:6452552-6662260, which is under selection in PNG highlanders. The candidate SNP for the region is rs74576183. The missense variant (rs7295376, TAPBPL-G151R) in high LD with rs74576183 is shown in orange. (**b**) Region chr4:82750503-83146792, which is under selection in PNG lowlanders. Candidate SNP is rs4693058. (**c, d**) Violin plots of the heart rate distribution in PNG depending on their genotype for the candidate SNPs (A = ancestral allele, D = derived allele (under selection)) (**c**) rs74576183, AA=AA, AD=AG, DD=GG and (d) rs4693058, AA=CC, AD=CT, DD=TT. Reprinted from Figure 2 (André et al., 2022), licensed under Creative Commons CC BY license. Copyright © 2022, André et. al.

The GBP locus under selection in PNG lowlanders also has a missense variant absent in non-PNG populations, with a DAF of 82% in PNG lowlanders. This variant is also in LD (R2=0.57) with the candidate SNP for the region (rs368120563). While we expect CLUES top results to be enriched for the causal SNPs of selection, we suggest that for these two cases where the candidate SNPs are in LD with a missense variant, the linked missense variant modifying protein sequence is likely to be the real target of selection for the genomic region.

5.4.5. Archaic introgression in the genomic regions under selection

Within PNG highlanders, the region with the strongest signature of selection carries an haplotype sharing similarities with the Altai neanderthal. This region overlaps with *LINC02388*, which might regulate the HIF/VEGF pathway and is under selection in PNG highlanders. Rs74576183, the candidate SNP associated with a slower heart rate and under selection in PNG highlanders, is located in a region carrying a Denisovan-like haplotype.

The six regions under selection in PNG lowlanders and with evidence for archaic introgression include the immunity-related GBP locus, which exhibited the highest selection peak in PNG lowlanders and showed haplotypes with similarities to Denisovan and Altai Neanderthal.

Interestingly, two candidate SNPs for each studied PNG population (a total of four SNPs) are found on introgressed haplotypes and absent on non-archaic haplotypes, but they are not found on the archaic genomes. These observations suggest that the selected mutations appeared after the introgression event, and the selection of the mutation led to an increase in the introgressed haplotype in PNG populations. Another hypothesis might be that the archaic hominin might show polymorphism for this site but that the currently available whole genome sequences for Denisova or Neanderthal do not carry this candidate variant and that the PNG population would carry both the segregating haplotypes.

6. CONCLUSIONS

The four references in this thesis contributed to a deeper understanding of the evolutionary history and the genetic and phenotypic diversity of New Guinean populations from both a global perspective of the Sahul settlement and a local perspective of the diverse environments within PNG.

Ref I – We present the first models of northern Sahul and New Guinea settlement based on whole-genome data. We described two likely models for the migration across Wallacea to northern Sahul. The first model supports a unique migration from Sunda to Sahul by the ancestors of Indigenous Australians and New Guineans. In contrast, the second model suggests an early split of this ancestral group. We explored additional models describing New Guinea's settlement after the AMH's arrival in northern Sahul. The settlement of New Guinea was probably initiated from its southeast region and followed by two migrations into the highlands.

Ref II – By combining absolute and relative chronologies, we show that the first settlers of northern Sahul, the ancestors of New Guineans and Bismarck Islanders, overcame a strong selection almost immediately after reaching this new territory. Selection mostly acted on genes related to the immune response (e.g., *ZDHHC20*), the diet (e.g., *KCNMB2*, *POR*) or both (e.g., *RARB)*. Indeed, Sahul's first settlers encountered novel pathogenic organisms and faced the challenge of exploiting unfamiliar food resources, potentially requiring adaptations in nutrient absorption.

Ref III – PNG highlanders significantly differ from PNG lowlanders for multiple phenotypes. Specifically, we found that PNG highlanders have a shorter standing height, smaller waist, deeper chest, and higher haemoglobin concentration. There might be likely several processes other than natural selection contributing to these differences. However, given that most of these phenotype differences are similar to those found in other high-altitude populations, we suggest they are part of an adaptive strategy of PNG highlanders to cope with the lower oxygen levels in the highlands. Further genomic analysis has been pursued to investigate this hypothesis in Ref III.

Ref IV – This study provides evidence of genomic selection signatures specific to PNG highlanders and lowlanders resulting from distinct environmental pressures such as hypoxia and pathogenic pressure. Our findings suggest that natural selection has shaped genes associated with HIFs, brain development, blood composition, and immunity in PNG highlanders and genes related to immunity and blood composition in lowlanders. Furthermore, we identified one SNP driving selection in highlanders that lower the heart rate, whereas one in lowlanders increases the heart rate. Moreover, we observed that both SNPs are associated with blood composition phenotypes in the UK biobank. Our results suggest that selection acted on haematological components in PNG highlanders and lowlanders and that heart rate might be a proxy. Finally, we also highlighted the potential role of archaic introgression in local adaptation in PNG populations.

This study suggests that natural selection in response to hypoxia has resulted in distinct genetic adaptations in PNG highlanders compared to other high-altitude populations. Nevertheless, our findings indicate that, consistent with observations in other high-altitude populations, cardiovascular phenotypes remain a target of selection in PNG highlanders. PNG lowlanders also showed a strong signal of local adaptation to pathogenic pressure. Our study provides important insights into the genetic mechanisms underlying local adaptation in two PNG populations and highlights the complex interplay between genetic and environmental factors shaping human genetic diversity.

The studies presented in this thesis offer a comprehensive perspective on the interplay between the unique diversity in the populations of PNG and its diverse environment. They emphasize the complexity of the settlement within New Guinea. This thesis also highlights the crucial role of natural selection in shaping genetic variation, which has acted on PNG genomes since the early settlement of Sahul. Moreover, it sheds light on the differences in selection signals between PNG populations using newly sequenced whole-genome sequences and phenotypic measurements. These studies stress the complex interplay between genetic variation, environmental pressures, and local adaptation in PNG populations.

SUMMARY IN ESTONIAN

Anatoomiliselt kaasaegse inimese evolutsioon Uus-Guineas: asustamise, arhailiste inimlastega segunemise ja kohastumuste ajalugu

Uus-Guinea on üks pikimat aega järjest inimese poolt asustatud piirkondadest väljaspool Aafrikat (O'Connell et al., 2018). Esmaasustamise ajal, kui anatoomiliselt kaasaegsed inimesed (AKI) Aafrikast välja rändasid, oli merepind praegusega võrreldes oluliselt madalam, mistõttu mõned alad, mis tänapäeval on merega eraldatud, moodustasid ühtseid maismaamassiive. Vaikse-Ookeani lõunaosas oli kaks peamist kontinenti: Sunda, mis ühendas paljusid tänaseid Kagu-Aasia saari Aasia kontinendiga, ja Sahul, mis ühendas Austraalia, Tasmaania ja Uus-Guinea. Need kaks iidset kontinenti olid üksteisest eraldatud Wallacea üleminekupiirkonnaga, paljude väikeste saartega pikitud sügava merega, mis takistas inimeste liikumist Sunda ja Sahuli vahel (Wallace 1869). Vaatamata neile takistustele jõudsid esimesed inimesed Sahuli pinnale 65 kuni 50 tuhat aastat tagasi (Bradshaw et al., 2021; Clarkson et al., 2017; Pedro et al., 2020). Kuidas täpselt esimesed AKId Wallacea veelahtme ületasid ja Sahuli jõudsid, on siiani vaidlusalane uurimisteema. Ilmselt kavandati teekond ette ja seda tehti mitmes jaos. On ka tõenäoline, et erinevad grupid jõudsid Sahuli eri teid pidi (Bird et al., 2019; Bradshaw et al., 2019).

Lisaks varasele asustamisele kaasaegse inimese poolt paistab Uus-Guinea silma ka ainulaadse bioloogilise ja kultuurilise mitmekesisusega. Kuna suurema osa ajast viimase 50 000 aasta jooksul olid uusguinealased eraldatud nii Euroopast kui Aasiast, on seal praeguseks välja kujunenud rikkalik kohalik geneetiline mitmekesisus (Bergström et al., 2020, 2016; Malaspinas et al., 2016). Üksteisest geneetiliselt kõige eristunumad inimgrupid Uus-Guineas on mäestike ja tasandike elanikud. Nende eristumine algas holotseeni ajal, mil grupid jäid üksteise suhtes isolatsiooni (Bergström et al., 2017). Selle aja jooksul on nad pidanud kohanema ka väga erinevate keskkonnamõjudega. Uus-Guineas asuvad Okeaania kõrgeimad tipud, mis ulatuvad kuni 4900 meetrini merepinnast (Summerhayes et al., 2017).

Uus-Guinea mäestikuelanikud asusid mägedesse elama vähemalt 20 000 aastat tagasi (Haas et al., 2017; Madsen et al., 2017). Vaatamata kõrgusest tekkivale hapnikuvaegusele e hüpoksiale, mis on väga tugev väliskeskkonnast tulenev stressiallikas, on alates holotseenist mäestikualad kõige tihedamini asustatud Uus-Guinea piirkonnad (Brookfield & Allen, 1989; Müller et al., 2003; Trájer et al., 2020). Vaatamata pikale asustamisajale, pole uuritud sealsete rahvaste kohastumust hüpoksiaga. Üks hüpotees Paapua Uus-Guinea rahvastiku väga ebaühtlasele asustusjaotusele on väga tugev patogeenide surve tasandikel (Riley, 1983; Trájer, 2022; Trájer et al., 2020). Malaaria võib olla üks peamisi selle surve põhjustajaid.

Veel üheks uusguinealaste eripäraks on nende genoomides arvukalt säilinud jäljed hiljutisest segunemisest tänaseks väljasurnud hominiini – denisi inimesega (Larena et al., 2021; Reich et al., 2010). Arhailiste inimestega segunemise tulemusel saadud geneetiline materjal võis kiirendada Uus-Guineasse jõudnud AKIde

kohastumist võrreldes *de novo tekkinud* mutatsioonidele mõjuvate valikuprotsessidega (Racimo et al., 2015). Näiteks on tiibetlastel seostatud denisi inimeselt saadud *EPAS1* geeni varianti madalama hemoglobiini kontsentratsiooniga veres, mis võib olla kohastumus hüpoksia kahjustava mõju tasakaalustamiseks (Huerta-Sánchez et al., 2014).

Kui AKId liikusid Sundast Sahuli ning jõudsid ka praeguse Uus-Guinea aladele, seisid nad silmitsi paljude uute keskkonnatingimuste ning maastikega. See töö analüüsib Paapua Uus-Guinea populatsioonide geneetilist ja fenotüüpilist mitmekesisust, kombineerides täisgenoomide sekveneerimist ja fenotüüpide mõõtmist. Esimene uurimus kirjeldab genoomikaandmetel põhinevat esmast Põhja-Sahuli asustamise mudelit. Teise uurimuse keskmes on loodusliku valiku kronoloogia AKI rändel läbi Wallacea ja Põhja-Sahuli asustamisel. Kolmas uurimus võrdleb fenotüüpilisi erinevusi Paapua Uus-Guinea mäestiku- ja tasandikuelanike vahel. Neljandas töös uuritakse loodusliku valiku mõju Paapua Uus-Guinea mäestiku- ja tasandikuelanikele. Käesoleva dissertatsiooni uuringud avavad Paapua Uus-Guinea elanike evolutsioonilise ajaloo uued tahud nii laiemalt Sahuli asustamise seisukohast kui ka kitsamalt, keskendudes Paapua Uus-Guinea inimeste kohastumistele sealses mitmekesises keskkonnas. Teadustöö tulemused toovad tähelepanu keskmesse keeruka vahekorra geneetilise varieeruvuse, keskkonnamõjude ja kohastumuslike vastuste vahel, mis kõik mõjutavad inimese evolutsiooni.

Esimeses uuringus (Ref I) tuvastasime genoomiandmete põhjal koostatud mudelite abil Uus-Guinea asustamise ajal toimunud peamised geneetilised sündmused. Kirjeldasime meie andmestikuga sobivat kahte Sahuli asustamise mudelit. Esimene mudel toetab Austraalia ja Uus-Guinea põlisasukate eellaste rännet Sundast Sahuli. Teine mudel näitab Austraalia ja Uus-Guinea põlisasukate eellasgruppide varast lahknemist, võimalik, et Wallaceas viibimise ajal. See mudel toetab ka põhjateed kui ühte võimalikku inimeste rändeteed Sahuli. Uurisime ka lisamudeleid, mis kirjeldavad Uus-Guinea asustamise kulgu peale esimeste AKIde saabumist Põhja-Sahuli. Uus-Guinea asustamine sai alguse ilmselt saare loodeosast, levis sealt rännetega lõunas ja põhjas asuvatele madalamatele aladele jõudes lõpuks lääne ja ida mäestikupiirkondadesse.

Teises uuringus (Ref II) seadsime eesmärgiks tuvastada loodusliku valiku ülegenoomseid mustreid. Taastasime valikusündmuste kronoloogia AKI rändel läbi Wallacea ja Põhja-Sahuli asustamise. Tuvastasime loodusliku valiku geneetilisi mustreid Wallaceast, Uus-Guineast ja Bismarcki saarestikust pärit inimeste varem avaldatud täisgenoomidest. Valisime need populatsioonid omavahelise suhtelise eraldatuse tõttu, mis võimaldab tuvastada erinevatele rände- ja asustamisetappidele iseloomulikke valikusignaale. Igale populatsioonile unikaalsete valikusignaalide tuvastamiseks kasutasime populatsioonidevahelisi valikuteste: PBS, XP-EHH ja XP-nSL (Sabeti, 2006; Szpiech & Hernandez, 2014; Yi et al., 2010). Tuvastades valikusignaale, mis olid omased vaid ühele populatsioonile kolmest, saime reastada suhtelise valikusündmuste kronoloogia. Valikus olevate geneetiliste variantide koalestsentsiaegade arvutamine GEVA meetodiga (Albers & McVean, 2020) võimaldas paika panna valikusündmuste absoluutse ajaskaala. Kombineerides suhtelist ja absoluutset ajaskaalat, näitasime, et Sahuli esmaasukad – uusguinealaste ja bismarklaste esivanemad, jäid väga tugeva loodusliku valiku surve alla

peaaegu kohe peale uuele territooriumile jõudmist. Valik mõjus peamiselt geenidele, mis on seotud immuunvastusega (*ZDHHC20*), toitumisega (*KCNMB2, POR*) või mõlemaga (*RARB*). Sahuli esmaasukad olid silmitsi uute patogeenidega ning pidid kasutusele võtma ka uusi ja tundmatuid toiduallikaid, mis võisid vajada kohastumist toitainete imendumisel.

Kolmandas uuringus (Ref III) püstitasime hüpoteesi, et Paapua Uus-Guinea mäestikuelanikud erinevad sealsetest tasandikuelanikest teatud fenotüüpide poolest, mis võimaldavad neil hakkama saada mäestiku madalama hapnikusisaldusega. Hüpoteesi testimiseks analüüsisime, kuidas 13 selle uuringu jaoks mõõdetud fenotüüpi on seotud südameveresoonkonna tervise, keha proportsioonide ja kopsumahuga, ning mil moel erinevad need Paapua Uus-Guinea mäestiku- ja tasandikuelanike vahel. Võrdlesime oma leide ka varem avaldatud vaatlustulemustega Andide, Etioopia ja Tiibeti mäestikuelanikelt. Leidsime, et Paapua Uus-Guinea mäestikuelanikud on lühemad, väiksema vööümbermõõduga, sügavama rinnakorviga ning madalama hemoglobiinitasemega veres võrreldes tasandikuelanikega. Neid erinevusi võivad põhjustada ka mitmed muud protsessid lisaks looduslikule valikule. Samas, enamik neid fenotüübierinevusi on sarnased teistele maailma mäestikupiirkondade elanikele, seega on tõenäoline, et need omadused on osa Paapua Uus-Guinea mäestikuelanike kohastumusest madala hapnikutasemega.

Neljanda uuringus (Ref IV) püstitasime hüpoteesi, et Paapua Uus-Guinea mäestiku- ja tasandikuelanike genoomides võib leida erinevaid loodusliku valiku signaale vastavalt nende keskkonnanišile. Lisahüpotees oli, et need signaalid on seotud varem näidatud fenotüüpide erinevustega (Ref III, André et al., 2021). Nende hüpoteeside uurimiseks sekveneerisime üle 200 Paapua Uus-Guinea mäestiku- ja tasandikuelaniku genoomiproovi, millest tuvastasime populatsioonidele iseloomulikud loodusliku valiku signaalid PBS ja XP-EHH valikuskooridega (Sabeti, 2006; Yi et al., 2010). Tuvastamaks SNPsid, mis neis populatsioonides on kõige tõenäolisemalt valiku all, kasutasime CLUES meetodit (Stern et al., 2019). Lõpuks seostasime need SNPd konkreetsete fenotüüpidega UK Biopanga andmebaasist (Pan-UKB team, 2020) ja paapua-uusguinealastel mõõdetutega (Ref III, André et al., 2021). Uuring tuvastas vastavalt Paapua Uus-Guinea mäestiku- ja tasandikuelanike genoomidest neile iseloomulikud positiivse loodusliku valiku signaalid, mis on põhjustatud erinevatest keskkonnatingimustest nagu näiteks hüpoksia mäestikus ja patogeenide surve tasandikul. Meie tulemused näitavad, et Paapua Uus-Guinea mäestikuelanikel on looduslik valik mõjutanud geene, mis on seotud HIFi (hüpoksiast indutseeritud faktor), aju arengu, vere koostise ja immuunsusega, samas kui tasandikuelanikel vaid immuunsuse ja vere koostisega. Lisaks tuvastasime ühe SNP, mis põhjustab mäestikuelanikel madalamat südame löögisagedust ja teise SNP, mis alandab tasandikuelanikel südame löögisagedust. Mõlemad SNPd on UK Biopangas seostatud verekoostise fenotüüpidega. Meie tulemused viitavad, et meie poolt mõõdetud südame löögisagedus on asendustunnus, ja tegelikult mõjutas valik hematoloogilisi komponente ka Paapua Uus-Guinea mäestiku- ja tasandikuelanikel. Samuti hindasime uuringus arhailistelt inimlastelt pärit geneetilise materjali potentsiaalset mõju Paapua Uus-Guinea populatsioonide kohastumisel keskkonnatingimustega.

REFERENCES

Ackerman, H., Usen, S., Jallow, M., Sisay-Joof, F., Pinder, M., & Kwiatkowski, D. P. (2005). A comparison of case-control and family-based association methods : The example of sickle-cell and malaria. *Annals of Human Genetics*, *69*(Pt 5), 559–565. https://doi.org/10.1111/j.1529-8817.2005.00180.x

Akhwale, W. S., Lum, J. K., Kaneko, A., Eto, H., Obonyo, C., Björkman, A., & Kobayakawa, T. (2004). Anemia and malaria at different altitudes in the western highlands of Kenya. *Acta Tropica*, *91*(2), 167–175. https://doi.org/10.1016/j.actatropica.2004.02.010

Albers, P. K., & McVean, G. (2020). Dating genomic variants and shared ancestry in population-scale sequencing data. *PLOS Biology*, *18*(1), e3000586. https://doi.org/10.1371/journal.pbio.3000586

Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, *19*(9), 1655–1664. https://doi.org/10.1101/gr.094052.109

Alkorta-Aranburu, G., Beall, C. M., Witonsky, D. B., Gebremedhin, A., Pritchard, J. K., & Rienzo, A. D. (2012). The Genetic Architecture of Adaptations to High Altitude in Ethiopia. *PLOS Genetics*, *8*(12), e1003110. https://doi.org/10.1371/journal.pgen.1003110

Allen, J., & O'Connell, J. F. (2020). A different paradigm for the initial colonisation of Sahul. *Archaeology in Oceania*, *55*(1), 1–14. https://doi.org/10.1002/arco.5207

Altshuler, D., Donnelly, P., & The International HapMap Consortium. (2005). A haplotype map of the human genome. *Nature*, *437*(7063), 1299–1320. https://doi.org/10.1038/nature04226

André, M., Brucato, N., Hudjasov, G., Pankratov, V., Yermakovich, D., Kreevan, R., … Ricaut, F.-X. (2022, décembre 15). *Positive selection in the genomes of two Papua New Guinean populations at distinct altitude levels* (p. 2022.12.15.520226). p. 2022.12.15.520226. bioRxiv. https://doi.org/10.1101/2022.12.15.520226

André, M., Brucato, N., Plutniak, S., Kariwiga, J., Muke, J., Morez, A., … Ricaut, F.-X. (2021). Phenotypic differences between highlanders and lowlanders in Papua New Guinea. *PLOS ONE*, *16*(7), e0253921. https://doi.org/10.1371/journal.pone.0253921

Apinjoh, T. O., Anchang-Kimbi, J. K., Njua-Yafi, C., Ngwai, A. N., Mugri, R. N., Clark, T. G., … in collaboration with The MalariaGEN Consortium. (2014). Association of candidate gene polymorphisms and TGF-beta/IL-10 levels with malaria in three regions of Cameroon : A case–control study. *Malaria Journal*, *13*(1), 236. https://doi.org/10.1186/1475-2875-13-236

Arenas, M. (2013). The importance and application of the ancestral recombination graph. *Frontiers in Genetics*, *4*. Accessed at https://www.frontiersin.org/articles/10.3389/fgene.2013.00206

Bafaro, E., Liu, Y., Xu, Y., & Dempski, R. E. (2017). The emerging role of zinc transporters in cellular homeostasis and cancer. *Signal Transduction and Targeted Therapy*, *2*, 17029. https://doi.org/10.1038/sigtrans.2017.29

Bammigatti, C., Shetty, S., Shetty, S., & Kumar, A. (2011). Benign tertian malaria – a misnomer? *Tropical Doctor*, *41*(3), 168–169. https://doi.org/10.1258/td.2011.110025

Barry, P. W., & Pollard, A. J. (2003). Altitude illness. *BMJ : British Medical Journal*, *326*(7395), 915–919.

Beall, C. M., Brittenham, G. M., Strohl, K. P., Blangero, J., Williams-Blangero, S., Goldstein, M. C., … Gonzales, C. (1998). Hemoglobin concentration of high-altitude Tibetans and Bolivian Aymara. *American Journal of Physical Anthropology*, *106*(3), 385–400. https://doi.org/10.1002/(SICI)1096-8644(199807)106:3<385::AID-AJPA10>3.0.CO;2-X

Beall, Cynthia M. (2013). Human adaptability studies at high altitude : Research designs and major concepts during fifty years of discovery. *American Journal of Human Biology*, *25*(2), 141–147. https://doi.org/10.1002/ajhb.22355

Beall, Cynthia M. (2014). Adaptation to High Altitude : Phenotypes and Genotypes. *Annual Review of Anthropology*, *43*(1), 251–272. https://doi.org/10.1146/annurev-anthro-102313-030000

Beall, Cynthia M., Decker, M. J., Brittenham, G. M., Kushner, I., Gebremedhin, A., & Strohl, K. P. (2002). An Ethiopian pattern of human adaptation to high-altitude hypoxia. *Proceedings of the National Academy of Sciences*, *99*(26), 17215–17218. https://doi.org/10.1073/pnas.252649199

Beaumont, M., & Wang, J. (2019). Conservation Genetics. In *Handbook of Statistical Genomics* (p. 457–40). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781119487845.ch16

Bellwood, P., Gamble, C., Blanc, S. A. L., Pluciennik, M., Richards, M., & Terrell, J. E. (2007). First Farmers : The Origins of Agricultural Societies, by Peter Bellwood. Malden (MA): Blackwell, 2005; ISBN 0-631-20565-9 hardback £60; ISBN 0-631-20566-7 paperback £17.99, xix+360 pp., 59 figs., 3 tables. *Cambridge Archaeological Journal*, *17*(1), 87–109. https://doi.org/10.1017/S0959774307000078

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate : A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, *57*(1), 289–300.

Bergström, A., McCarthy, S. A., Hui, R., Almarri, M. A., Ayub, Q., Danecek, P., … Tyler-Smith, C. (2020). Insights into human genetic variation and population history from 929 diverse genomes. *Science*, *367*(6484). https://doi.org/10.1126/science.aay5012

Bergström, A., Nagle, N., Chen, Y., McCarthy, S., Pollard, M. O., Ayub, Q., … Tyler-Smith, C. (2016). Deep Roots for Aboriginal Australian Y Chromosomes. *Current Biology: CB*, *26*(6), 809–813. https://doi.org/10.1016/j.cub.2016.01.028

Bergström, A., Oppenheimer, S. J., Mentzer, A. J., Auckland, K., Robson, K., Attenborough, R., … Tyler-Smith, C. (2017). A Neolithic expansion, but strong genetic structure, in the independent history of New Guinea. *Science*, *357*(6356), 1160–1163. https://doi.org/10.1126/science.aan3842

Bersaglieri, T., Sabeti, P. C., Patterson, N., Vanderploeg, T., Schaffner, S. F., Drake, J. A., … Hirschhorn, J. N. (2004). Genetic signatures of strong recent positive selection at the lactase gene. *American Journal of Human Genetics*, *74*(6), 1111–1120. https://doi.org/10.1086/421051

Bhat, S., Kumar, M., & Alva, J. (2013). Malaria and the conducting system of the heart. *BMJ Case Reports*, *2013*, bcr2012007462. https://doi.org/10.1136/bcr-2012-007462

Bigham, A. W., & Lee, F. S. (2014). Human high-altitude adaptation : Forward genetics meets the HIF pathway. *Genes & Development*, *28*(20), 2189–2204. https://doi.org/10.1101/gad.250167.114

Bird, M. I., Condie, S. A., O'Connor, S., O'Grady, D., Reepmeyer, C., Ulm, S., … Bradshaw, C. J. A. (2019). Early human settlement of Sahul was not an accident. *Scientific Reports*, *9*, 8220. https://doi.org/10.1038/s41598-019-42946-9

Bourke, M. (Mike), & Allen, B. (2021). *Estimating the population of Papua New Guinea in 2020* (SSRN Scholarly Paper Nº ID 3770356). Rochester, NY: Social Science Research Network. Accessed at Social Science Research Network website: https://papers.ssrn.com/abstract=3770356

Boyle, E. A., Li, Y. I., & Pritchard, J. K. (2017). An expanded view of complex traits : From polygenic to omnigenic. *Cell*, *169*(7), 1177–1186. https://doi.org/10.1016/j.cell.2017.05.038

Bradshaw, C. J. A., Norman, K., Ulm, S., Williams, A. N., Clarkson, C., Chadœuf, J., … Saltré, F. (2021). Stochastic models support rapid peopling of Late Pleistocene Sahul. *Nature Communications*, *12*(1), 2440. https://doi.org/10.1038/s41467-021-21551-3

Bradshaw, C. J. A., Ulm, S., Williams, A. N., Bird, M. I., Roberts, R. G., Jacobs, Z., … Saltré, F. (2019). Minimum founding populations for the first peopling of Sahul. *Nature Ecology & Evolution*, *3*(7), 1057–1063. https://doi.org/10.1038/s41559-019-0902-6

Brisbin, A., Bryc, K., Byrnes, J., Zakharia, F., Omberg, L., Degenhardt, J., … Bustamante, C. D. (2012). PCAdmix : Principal Components-Based Assignment of Ancestry Along Each Chromosome in Individuals with Admixed Ancestry from Two or More Populations. *Human Biology*, *84*(4), 343–364. https://doi.org/10.3378/027.084.0401

Brito, J., Siqués, P., León-Velarde, F., De La Cruz, J. J., López, V., & Herruzo, R. (2007). Chronic Intermittent Hypoxia at High Altitude Exposure for over 12 Years : Assessment of Hematological, Cardiovascular, and Renal Effects. *High Altitude Medicine & Biology*, *8*(3), 236–244. https://doi.org/10.1089/ham.2007.8310

Brookfield, H., & Allen, B. (1989). High-Altitude Occupation and Environment. *Mountain Research and Development*, *9*(3), 201–209. JSTOR. https://doi.org/10.2307/3673510

Browning, S. R., Browning, B. L., Zhou, Y., Tucci, S., & Akey, J. M. (2018). Analysis of Human Sequence Data Reveals Two Pulses of Archaic Denisovan Admixture. *Cell*, *173*(1), 53-61.e9. https://doi.org/10.1016/j.cell.2018.02.031

Brucato, N., André, M., Hudjashov, G., Mondal, M., Cox, M. P., Leavesley, M., & Ricaut, F.-X. (2022). Chronology of natural selection in Oceanian genomes. *IScience*, 104583. https://doi.org/10.1016/j.isci.2022.104583

Brucato, N., André, M., Tsang, R., Saag, L., Kariwiga, J., Sesuki, K., … Ricaut, F.-X. (2021). Papua New Guinean Genomes Reveal the Complex Settlement of North Sahul. *Molecular Biology and Evolution*, (msab238). https://doi.org/10.1093/molbev/msab238

Brutsaert, T. D., Soria, R., Caceres, E., Spielvogel, H., & Haas, J. D. (1999). Effect of developmental and ancestral high altitude exposure on chest morphology and pulmonary function in Andean and European/North American natives. *American Journal of Human Biology*, *11*(3), 383–395. https://doi.org/10.1002/(SICI)1520-6300(1999)11:3<383::AID-AJHB9>3.0.CO;2-X

Cámara-Leret, R., Frodin, D. G., Adema, F., Anderson, C., Appelhans, M. S., Argent, G., … van Welzen, P. C. (2020). New Guinea has the world's richest island flora. *Nature*, *584*(7822), 579–583. https://doi.org/10.1038/s41586-020-2549-5

Cavalli-Sforza, L. L., Piazza, A., Menozzi, P., & Mountain, J. (1988). Reconstruction of human evolution: Bringing together genetic, archaeological, and linguistic data. *Proceedings of the National Academy of Sciences*, *85*(16), 6002–6006. https://doi.org/10.1073/pnas.85.16.6002

Charlesworth, D., Charlesworth, B., & Morgan, M. T. (1995). The pattern of neutral molecular variation under the background selection model. *Genetics*, *141*(4), 1619–1632. https://doi.org/10.1093/genetics/141.4.1619

Charlesworth, Deborah. (2006). Balancing Selection and Its Effects on Sequences in Nearby Genome Regions. *PLOS Genetics*, *2*(4), e64. https://doi.org/10.1371/journal.pgen.0020064

Chen, L., Wolf, A. B., Fu, W., Li, L., & Akey, J. M. (2020). Identifying and Interpreting Apparent Neanderthal Ancestry in African Individuals. *Cell*. https://doi.org/10.1016/j.cell.2020.01.012

Cheong, H. I., Janocha, A. J., Monocello, L. T., Garchar, A. C., Gebremedhin, A., Erzurum, S. C., & Beall, C. M. (2016). Alternative hematological and vascular adaptive responses to high-altitude hypoxia in East African highlanders. *American Journal of Physiology-Lung Cellular and Molecular Physiology*, *312*(2), L172–L177. https://doi.org/10.1152/ajplung.00451.2016

Choin, J., Mendoza-Revilla, J., Arauna, L. R., Cuadros-Espinoza, S., Cassar, O., Larena, M., … Quintana-Murci, L. (2021). Genomic insights into population history and biological adaptation in Oceania. *Nature*, 1–7. https://doi.org/10.1038/s41586-021-03236-5

Clarkson, C., Jacobs, Z., Marwick, B., Fullagar, R., Wallis, L., Smith, M., … Pardoe, C. (2017). Human occupation of northern Australia by 65,000 years ago. *Nature*, *547*(7663), 306–310. https://doi.org/10.1038/nature22968

Cleary, E., Hetzel, M. W., & Clements, A. C. A. (2022). A review of malaria epidemiology and control in Papua New Guinea 1900 to 2021 : Progress made and future directions. *Frontiers in Epidemiology*, *2*. Accessed at https://www.frontiersin.org/articles/10.3389/fepid.2022.980795

Coller, M. (2009). SahulTime : Rethinking Archaeological Representation in the Digital Age. *Archaeologies*, *5*(1), 110–123. https://doi.org/10.1007/s11759-009-9096-x

Cotes, J. E., Anderson, H. R., & Patrick, J. M. (1974). Lung Function and the Response to Exercise in New Guineans : Role of Genetic and Environmental Factors. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *268*(893), 349–361. https://doi.org/10.1098/rstb.1974.0034

Crawford, D. C., Bhangale, T., Li, N., Hellenthal, G., Rieder, M. J., Nickerson, D. A., & Stephens, M. (2004). Evidence for substantial fine-scale variation in recombination rates across the human genome. *Nature Genetics*, *36*(7), 700–706. https://doi.org/10.1038/ng1376

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., … 1000 Genomes Project Analysis Group. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

Dannemann, M., Andrés, A. M., & Kelso, J. (2016). Introgression of Neandertal- and Denisovan-like Haplotypes Contributes to Adaptive Variation in Human Toll-like Receptors. *The American Journal of Human Genetics*, *98*(1), 22–33. https://doi.org/10.1016/j.ajhg.2015.11.015

Dannemann, M., & Kelso, J. (2017). The Contribution of Neanderthals to Phenotypic Variation in Modern Humans. *American Journal of Human Genetics*, *101*(4), 578–589. https://doi.org/10.1016/j.ajhg.2017.09.010

Denham, T. P. (2003). Origins of Agriculture at Kuk Swamp in the Highlands of New Guinea. *Science*, *301*(5630), 189–193. https://doi.org/10.1126/science.1085255

Department of Nutrition for Health and Development (NHD). (2011). *Haemoglobin concentrations for the diagnosis of anaemia and assessment of severity.* World Health Organization. Accessed at http://www.who.int/vmnis/indicators/haemoglobin. pdf

Douka, K., Slon, V., Jacobs, Z., Ramsey, C. B., Shunkov, M. V., Derevianko, A. P., … Higham, T. (2019). Age estimates for hominin fossils and the onset of the Upper Palaeolithic at Denisova Cave. *Nature*, *565*(7741), 640–644. https://doi.org/10.1038/s41586-018-0870-z

Drakeley, C. J., Carneiro, I., Reyburn, H., Malima, R., Lusingu, J. P. A., Cox, J., … Riley, E. M. (2005). Altitude-dependent and -independent variations in Plasmodium falciparum prevalence in northeastern Tanzania. *The Journal of Infectious Diseases*, *191*(10), 1589–1598. https://doi.org/10.1086/429669

Droma, T., McCullough, R. G., McCullough, R. E., Zhuang, J., Cymerman, A., Sun, S., … Moore, L. G. (1991). Increased vital and total lung capacities in Tibetan compared to Han residents of Lhasa (3,658 m). *American Journal of Physical Anthropology*, *86*(3), 341–351. https://doi.org/10.1002/ajpa.1330860303

Dudbridge, F., & Gusnanto, A. (2008). Estimation of significance thresholds for genome-wide association scans. *Genetic Epidemiology*, *32*(3), 227–234. https://doi.org/10.1002/gepi.20297

Eichstaedt, C. A., Antão, T., Cardona, A., Pagani, L., Kivisild, T., & Mormina, M. (2015). Genetic and phenotypic differentiation of an Andean intermediate altitude population. *Physiological Reports*, *3*(5), e12376. https://doi.org/10.14814/phy2.12376

Enard, D., & Petrov, D. A. (2018). Evidence that RNA Viruses Drove Adaptive Introgression between Neanderthals and Modern Humans. *Cell*, *175*(2), 360-371.e13. Scopus. https://doi.org/10.1016/j.cell.2018.08.034

Fan, S., Hansen, M. E. B., Lo, Y., & Tishkoff, S. A. (2016). Going global by adapting local : A review of recent human adaptation. *Science*, *354*(6308), 54–59. https://doi.org/10.1126/science.aaf5098

Ferrer-Admetlla, A., Liang, M., Korneliussen, T., & Nielsen, R. (2014). On Detecting Incomplete Soft or Hard Selective Sweeps Using Haplotype Structure. *Molecular Biology and Evolution*, *31*(5), 1275–1291. https://doi.org/10.1093/molbev/msu077

Fijarczyk, A., & Babik, W. (2015). Detecting balancing selection in genomes : Limits and prospects. *Molecular Ecology*, *24*(14), 3529–3545. https://doi.org/10.1111/mec.13226

Flint, J., Harding, R. M., Boyce, A. J., & Clegg, J. B. (1998). 1 The population genetics of the haemoglobinopathies. *Baillière's Clinical Haematology*, *11*(1), 1–51. https://doi.org/10.1016/S0950-3536(98)80069-3

Frisancho, A. R. (2013). Developmental Functional Adaptation to High Altitude : Review. *American Journal of Human Biology*, *25*(2), 151–168. https://doi.org/10.1002/ajhb.22367

Fumagalli, M., Sironi, M., Pozzoli, U., Ferrer-Admettla, A., Pattini, L., & Nielsen, R. (2011). Signatures of Environmental Genetic Adaptation Pinpoint Pathogens as the Main Selective Pressure through Human Evolution. *PLOS Genetics*, *7*(11), e1002355. https://doi.org/10.1371/journal.pgen.1002355

Gaffney, D., Ford, A., & Summerhayes, G. (2015). Crossing the Pleistocene–Holocene transition in the New Guinea Highlands : Evidence from the lithic assemblage of Kiowa rockshelter. *Journal of Anthropological Archaeology*, *39*, 223–246. https://doi.org/10.1016/j.jaa.2015.04.006

GBD 2013 Mortality and Causes of Death Collaborators. (2015). Global, regional, and national age–sex specific all-cause and cause-specific mortality for 240 causes of death, 1990–2013 : A systematic analysis for the Global Burden of Disease Study 2013. *The Lancet*, *385*(9963), 117–171. https://doi.org/10.1016/S0140-6736(14)61682-2

Gething, P. W., Elyazar, I. R. F., Moyes, C. L., Smith, D. L., Battle, K. E., Guerra, C. A., … Hay, S. I. (2012). A Long Neglected World Malaria Map : Plasmodium vivax Endemicity in 2010. *PLoS Neglected Tropical Diseases*, *6*(9), e1814. https://doi.org/10.1371/journal.pntd.0001814

Godyna, S., Diaz-Ricart, M., & Argraves, W. (1996). Fibulin-1 mediates platelet adhesion via a bridge of fibrinogen. *Blood*, *88*(7), 2569–2577. https://doi.org/10.1182/blood.V88.7.2569.bloodjournal8872569

Golson, J. (Éd.). (2017). *Ten thousand years of cultivation at Kuk Swamp in the highlands of Papua New Guinea*. Acton, ACT, Australia: Australian National University Press.

Green, R. E., Krause, J., Briggs, A. W., Maricic, T., Stenzel, U., Kircher, M., … Pääbo, S. (2010). A Draft Sequence of the Neandertal Genome. *Science*, *328*(5979), 710–722. https://doi.org/10.1126/science.1188021

Greksa, L. P. (2006). Growth and Development of Andean High Altitude Residents. *High Altitude Medicine & Biology*, *7*(2), 116–124. https://doi.org/10.1089/ham.2006.7.116

Griffiths, R. C., & Marjoram, P. (1997). An Ancestral Recombination Graph. In P. Donnelly & S. Tavaré (Éds.), *Progress in Population Genetics and Human Evolution* (p. 257–270). New York, NY: Springer New York. https://doi.org/10.1007/978-1-4757-2609-1_16

Grove, M., Lamb, H., Roberts, H., Davies, S., Marshall, M., Bates, R., & Huws, D. (2015). Climatic variability, plasticity, and dispersal : A case study from Lake Tana, Ethiopia. *Journal of Human Evolution*, *87*, 32–47. https://doi.org/10.1016/j.jhevol.2015.07.007

Gudjonsson, A., Gudmundsdottir, V., Axelsson, G. T., Gudmundsson, E. F., Jonsson, B. G., Launer, L. J., … Gudnason, V. (2022). A genome-wide association study of serum proteins reveals shared loci with common diseases. *Nature Communications*, *13*(1), 480. https://doi.org/10.1038/s41467-021-27850-z

Gupta, S., Gazendam, N., Farina, J. M., Saldarriaga, C., Mendoza, I., López-Santi, R., … Baranchuk, A. (2021). Malaria and the Heart : JACC State-of-the-Art Review. *Journal of the American College of Cardiology*, *77*(8), 1110–1121. https://doi.org/10.1016/j.jacc.2020.12.042

Haas, R., Stefanescu, I. C., Garcia-Putnam, A., Aldenderfer, M. S., Clementz, M. T., Murphy, M. S., … Watson, J. T. (2017). Humans permanently occupied the Andean highlands by at least 7 ka. *Royal Society Open Science*, *4*(6), 170331. https://doi.org/10.1098/rsos.170331

Haldane, J. B. S. (1919). The combination of linkage values, and the calculation of distances between the loci of linked factors. *Journal of Genetics*, *8*(4), 299–309.

Haldar, K., & Mohandas, N. (2009). Malaria, erythrocytic infection, and anemia. *Hematology. American Society of Hematology. Education Program*, 87–93. https://doi.org/10.1182/asheducation-2009.1.87

Hall, J. E., & Guyton, A. C. (2011). *Guyton and Hall textbook of medical physiology*. Philadelphia, PA: Saunders Elsevier. Accessed at http://www.clinicalkey.com/dura/browse/bookChapter/3-s2.0-C20090602506

Hammer, M. F., Woerner, A. E., Mendez, F. L., Watkins, J. C., & Wall, J. D. (2011). Genetic evidence for archaic admixture in Africa. *Proceedings of the National Academy of Sciences*, *108*(37), 15123–15128. https://doi.org/10.1073/pnas.1109300108

Harris, K., & Nielsen, R. (2016). The Genetic Cost of Neanderthal Introgression. *Genetics*, *203*(2), 881–891. https://doi.org/10.1534/genetics.116.186890

Harrison, G. A., Küchemann, C. F., Moore, M. a. S., Boyce, A. J., Baju, T., Mourant, A. E., … Clegg, E. J. (1969). The effects of altitudinal variation in Ethiopian populations. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, *256*(805), 147–182. https://doi.org/10.1098/rstb.1969.0040

Harrison, T., Krigbaum, J., & Manser, J. (2006). Primate Biogeography and Ecology on the Sunda Shelf Islands : A Paleontological and Zooarchaeological Perspective. In S. M. Lehman & J. G. Fleagle (Éds.), *Primate Biogeography : Progress and Prospects* (p. 331–372). Boston, MA: Springer US. https://doi.org/10.1007/0-387-31710-4_12

Hawkins, S., O'Connor, S., Maloney, T. R., Litster, M., Kealy, S., Fenner, J. N., … Louys, J. (2017). Oldest human occupation of Wallacea at Laili Cave, Timor-Leste, shows broad-spectrum foraging responses to late Pleistocene environments. *Quaternary Science Reviews*, *171*, 58–72. https://doi.org/10.1016/j.quascirev.2017.07.008

Hein, J., Schierup, M. H., & Wiuf, C. (2005). *Gene genealogies, variation and evolution : A primer in coalescent theory*. Oxford ; New York: Oxford University Press.

Higham, T., Douka, K., Wood, R., Ramsey, C. B., Brock, F., Basell, L., … Jacobi, R. (2014). The timing and spatiotemporal patterning of Neanderthal disappearance. *Nature*, *512*(7514), 306–309. https://doi.org/10.1038/nature13621

Hoit, B. D., Dalton, N. D., Gebremedhin, A., Janocha, A., Zimmerman, P. A., Zimmerman, A. M., … Beall, C. M. (2011). Elevated pulmonary artery pressure among Amhara highlanders in Ethiopia. *American journal of human biology : the official journal of the Human Biology Council*, *23*(2). https://doi.org/10.1002/ajhb.21130

Holm, A., Gomes, L. C., Biering-Soerensen, T., Silvestre, O., & Brainin, P. (2020). Malaria and cardiovascular disease : A systematic review. *European Heart Journal*, *41*(Supplement_2), ehaa946.2817. https://doi.org/10.1093/ehjci/ehaa946.2817

Hu, S., Feng, J., Wang, M., Wufuer, R., Liu, K., Zhang, Z., & Zhang, Y. (2022, juin 14). *Nrf1 is an indispensable redox-determining factor for mitochondrial homeostasis by integrating multi-hierarchical regulatory networks* (p. 2022.05.04.490622). p. 2022.05.04.490622. bioRxiv. https://doi.org/10.1101/2022.05.04.490622

Hubisz, M., & Siepel, A. (2020). Inference of Ancestral Recombination Graphs Using ARGweaver. In J. Y. Dutheil (Éd.), *Statistical Population Genomics* (p. 231–266). New York, NY: Springer US. https://doi.org/10.1007/978-1-0716-0199-0_10

Hublin, J. J. (2009). The origin of Neandertals. *Proceedings of the National Academy of Sciences*, *106*(38), 16022–16027. https://doi.org/10.1073/pnas.0904119106

Hublin, J.-J., Ben-Ncer, A., Bailey, S. E., Freidline, S. E., Neubauer, S., Skinner, M. M., … Gunz, P. (2017). New fossils from Jebel Irhoud, Morocco and the pan-African origin of Homo sapiens. *Nature*, *546*(7657), 289–292. https://doi.org/10.1038/nature22336

Hudson, R. R. (1983). Properties of a neutral allele model with intragenic recombination. *Theoretical Population Biology*, *23*(2), 183–201. https://doi.org/10.1016/0040-5809(83)90013-8

Huerta-Sánchez, E., DeGiorgio, M., Pagani, L., Tarekegn, A., Ekong, R., Antao, T., … Nielsen, R. (2013). Genetic Signatures Reveal High-Altitude Adaptation in a Set of Ethiopian Populations. *Molecular Biology and Evolution*, *30*(8), 1877–1888. https://doi.org/10.1093/molbev/mst089

Huerta-Sánchez, E., Jin, X., Asan, Bianba, Z., Peter, B. M., Vinckenbosch, N., … Nielsen, R. (2014). Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. *Nature*, *512*(7513), 194–197. https://doi.org/10.1038/nature13408

Hung, H., & Carson, M. T. (2014). Foragers, fishers and farmers : Origins of the Taiwanese Neolithic. *Antiquity*, *88*(342), 1115–1131. https://doi.org/10.1017/S0003598X00115352

Itan, Y., Powell, A., Beaumont, M. A., Burger, J., & Thomas, M. G. (2009). The Origins of Lactase Persistence in Europe. *PLoS Computational Biology*, *5*(8), e1000491. https://doi.org/10.1371/journal.pcbi.1000491

Jacobs, G. S., Hudjashov, G., Saag, L., Kusuma, P., Darusallam, C. C., Lawson, D. J., … Cox, M. P. (2019). Multiple Deeply Divergent Denisovan Ancestries in Papuans. *Cell*, *177*(4), 1010-1021.e32. https://doi.org/10.1016/j.cell.2019.02.035

Jobling, M., Hollox, E., Hurles, M., Kivisild, T., & Tyler-Smith, C. (2013). *Human Evolutionary Genetics* (2e éd.). Garland Science. https://doi.org/10.1201/9781317952268

Juric, I., Aeschbacher, S., & Coop, G. (2016). The Strength of Selection against Neanderthal Introgression. *PLoS Genetics*, *12*(11), e1006340. https://doi.org/10.1371/journal.pgen.1006340

Kaplan, N. L., Hudson, R. R., & Langley, C. H. (1989). The «hitchhiking effect» revisited. *Genetics*, *123*(4), 887–899. https://doi.org/10.1093/genetics/123.4.887

Karmin, M., Flores, R., Saag, L., Hudjashov, G., Brucato, N., Crenna-Darusallam, C., … Cox, M. P. (2022). Episodes of Diversification and Isolation in Island Southeast Asian and Near Oceanian Male Lineages. *Molecular Biology and Evolution*, *39*(3), msac045. https://doi.org/10.1093/molbev/msac045

Kayser, M. (2010). The human genetic history of Oceania : Near and remote views of dispersal. *Current Biology: CB*, *20*(4), R194–201. https://doi.org/10.1016/j.cub.2009.12.004

Kealy, S., Louys, J., & O'Connor, S. (2018). Least-cost pathway models indicate northern human dispersal from Sunda to Sahul. *Journal of Human Evolution*, *125*, 59–70. https://doi.org/10.1016/j.jhevol.2018.10.003

Key, F. M., Fu, Q., Romagne, F., Lachmann, M., & Andres, A. M. (2016). Human adaptation and population differentiation in the light of ancient genomes. *Nature Communications*, *7*. Scopus. https://doi.org/10.1038/ncomms10775

Kho, S., Barber, B. E., Johar, E., Andries, B., Poespoprodjo, J. R., Kenangalem, E., … McMorran, B. J. (2018). Platelets kill circulating parasites of all major Plasmodium species in human malaria. *Blood*, *132*(12), 1332–1344. https://doi.org/10.1182/blood-2018-05-849307

Kingman, J. F. C. (1982). The coalescent. *Stochastic Processes and Their Applications*, *13*(3), 235–248. https://doi.org/10.1016/0304-4149(82)90011-4

Kirch, P. V., & Hunt, T. L. (1988). Radiocarbon Dates from the Mussau Islands and the Lapita Colonization of the Southwestern Pacific. *Radiocarbon*, *30*(2), 161–169. https://doi.org/10.1017/S0033822200044106

Kirch, Patrick Vinton. (2002). *On the road of the winds : An archaeological history of the Pacific islands before European contact* (1. paperback print). Berkeley, Calif.: Univ. of California Press.

Kitur, U., Adair, T., Riley, I., & Lopez, A. D. (2019). Estimating the pattern of causes of death in Papua New Guinea. *BMC Public Health*, *19*(1), 1322. https://doi.org/10.1186/s12889-019-7620-5

Kong, A., Thorleifsson, G., Gudbjartsson, D. F., Masson, G., Sigurdsson, A., Jonasdottir, A., … Stefansson, K. (2010). Fine-scale recombination rate differences between sexes, populations and individuals. *Nature*, *467*(7319), 1099–1103. https://doi.org/10.1038/nature09525

Krause, J., Fu, Q., Good, J. M., Viola, B., Shunkov, M. V., Derevianko, A. P., & Pääbo, S. (2010). The complete mitochondrial DNA genome of an unknown hominin from southern Siberia. *Nature*, *464*(7290), 894–897. https://doi.org/10.1038/nature08976

Kuhlwilm, M., Gronau, I., Hubisz, M. J., de Filippo, C., Prado-Martinez, J., Kircher, M., … Castellano, S. (2016). Ancient gene flow from early modern humans into Eastern Neanderthals. *Nature*, *530*(7591), 429–433. https://doi.org/10.1038/nature16544

Kusuma, P., Brucato, N., Cox, M. P., Letellier, T., Manan, A., Nuraini, C., … Ricaut, F.-X. (2017). The last sea nomads of the Indonesian archipelago : Genomic origins and dispersal. *European Journal of Human Genetics*, *25*(8), 1004–1010. https://doi.org/10.1038/ejhg.2017.88

Kwiatkowski, D. P. (2005). How Malaria Has Affected the Human Genome and What Human Genetics Can Teach Us about Malaria. *The American Journal of Human Genetics*, *77*(2), 171–192. https://doi.org/10.1086/432519

Lampah, D. A., Yeo, T. W., Malloy, M., Kenangalem, E., Douglas, N. M., Ronaldo, D., … Price, R. N. (2015). Severe Malarial Thrombocytopenia : A Risk Factor for Mortality in Papua, Indonesia. *The Journal of Infectious Diseases*, *211*(4), 623–634. https://doi.org/10.1093/infdis/jiu487

Larena, M., McKenna, J., Sanchez-Quinto, F., Bernhardsson, C., Ebeo, C., Reyes, R., … Jakobsson, M. (2021). Philippine Ayta possess the highest level of Denisovan ancestry in the world. *Current Biology: CB*, S0960-9822(21)00977-5. https://doi.org/10.1016/j.cub.2021.07.022

Latinis, D. K., & Stark, K. (2005). Cave Use Variability in Central Maluku, Eastern Indonesia. *Asian Perspectives*, *44*(1), 119–136.

Lawson, D. J., Hellenthal, G., Myers, S., & Falush, D. (2012). Inference of Population Structure using Dense Haplotype Data. *PLOS Genetics*, *8*(1), e1002453. https://doi.org/10.1371/journal.pgen.1002453

Lee, P., Chandel, N. S., & Simon, M. C. (2020). Cellular adaptation to hypoxia through hypoxia inducible factors and beyond. *Nature Reviews Molecular Cell Biology*, *21*(5), 268–283. https://doi.org/10.1038/s41580-020-0227-y

Leonard, W. R. (1989). Nutritional determinants of high-altitude growth in Nuñoa, Peru. *American Journal of Physical Anthropology*, *80*(3), 341–352. https://doi.org/10.1002/ajpa.1330800308

Lewontin, R. C., & Krakauer, J. (1973). Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. *Genetics*, *74*(1), 175–195. https://doi.org/10.1093/genetics/74.1.175

Li, N., & Stephens, M. (2003). Modeling Linkage Disequilibrium and Identifying Recombination Hotspots Using Single-Nucleotide Polymorphism Data. *Genetics*, *165*(4), 2213–2233. https://doi.org/10.1093/genetics/165.4.2213

Lipson, M., Skoglund, P., Spriggs, M., Valentin, F., Bedford, S., Shing, R., … Reich, D. (2018). Population Turnover in Remote Oceania Shortly after Initial Settlement. *Current Biology: CB*, *28*(7), 1157–1165.e7. https://doi.org/10.1016/j.cub.2018.02.051

Liu, X., Li, Y. I., & Pritchard, J. K. (2019). Trans Effects on Gene Expression Can Drive Omnigenic Inheritance. *Cell*, *177*(4), 1022–1034.e6. https://doi.org/10.1016/j.cell.2019.04.014

Lonsdale, J., Thomas, J., Salvatore, M., Phillips, R., Lo, E., Shad, S., … Moore, H. F. (2013). The Genotype-Tissue Expression (GTEx) project. *Nature Genetics*, *45*(6), 580–585. https://doi.org/10.1038/ng.2653

Loos, R. J. F., & Yeo, G. S. H. (2022). The genetics of obesity : From discovery to biology. *Nature Reviews Genetics*, *23*(2), 120–133. https://doi.org/10.1038/s41576-021-00414-z

Lopez, M., Choin, J., Sikora, M., Siddle, K., Harmant, C., Costa, H. A., … Quintana-Murci, L. (2019). Genomic Evidence for Local Adaptation of Hunter-Gatherers to the African Rainforest. *Current Biology*, *29*(17), 2926–2935.e4. https://doi.org/10.1016/j.cub.2019.07.013

Lunter, G. (2019). Haplotype matching in large cohorts using the Li and Stephens model. *Bioinformatics*, *35*(5), 798–806. https://doi.org/10.1093/bioinformatics/bty735

MacMicking, J. D. (2012). Interferon-inducible effector mechanisms in cell-autonomous immunity. *Nature Reviews Immunology*, *12*(5), 367–382. https://doi.org/10.1038/nri3210

Madsen, D. B., Perreault, C., Rhode, D., Sun, Y., Yi, M., Brunson, K., & Brantingham, P. J. (2017). Early foraging settlement of the Tibetan Plateau highlands. *Archaeological Research in Asia*, *11*, 15–26. https://doi.org/10.1016/j.ara.2017.04.003

Malaspinas, A.-S., Westaway, M. C., Muller, C., Sousa, V. C., Lao, O., Alves, I., … Willerslev, E. (2016). A genomic history of Aboriginal Australia. *Nature*, *538*(7624), 207–214. https://doi.org/10.1038/nature18299

Mallick, S., Li, H., Lipson, M., Mathieson, I., Gymrek, M., Racimo, F., … Reich, D. (2016). The Simons Genome Diversity Project : 300 genomes from 142 diverse populations. *Nature*, *538*(7624), 201–206. https://doi.org/10.1038/nature18964

Marnetto, D., & Huerta-Sánchez, E. (2017). Haplostrips : Revealing population structure through haplotype visualization. *Methods in Ecology and Evolution*, *8*(10), 1389–1392. https://doi.org/10.1111/2041-210X.12747

Mathieson, I. (2021). The omnigenic model and polygenic prediction of complex traits. *The American Journal of Human Genetics*, *108*(9), 1558–1563. https://doi.org/10.1016/j.ajhg.2021.07.003

McLaren, W., Gil, L., Hunt, S. E., Riat, H. S., Ritchie, G. R. S., Thormann, A., … Cunningham, F. (2016). The Ensembl Variant Effect Predictor. *Genome Biology*, *17*(1), 122. https://doi.org/10.1186/s13059-016-0974-4

McManus, K. F., Taravella, A. M., Henn, B. M., Bustamante, C. D., Sikora, M., & Cornejo, O. E. (2017). Population genetic analysis of the DARC locus (Duffy) reveals adaptation from standing variation associated with malaria resistance in humans. *PLoS Genetics*, *13*(3), e1006560. https://doi.org/10.1371/journal.pgen.1006560

McMichael, T. M., Zhang, L., Chemudupati, M., Hach, J. C., Kenney, A. D., Hang, H. C., & Yount, J. S. (2017). The palmitoyltransferase ZDHHC20 enhances interferon-induced transmembrane protein 3 (IFITM3) palmitoylation and antiviral activity. *Journal of Biological Chemistry*, *292*(52), 21517–21526. https://doi.org/10.1074/jbc.M117.800482

McMorran, B. J. (2019). Immune role of platelets in malaria. *ISBT Science Series*, *14*(1), 67–76. https://doi.org/10.1111/voxs.12451

McVean, G. A. T., Myers, S. R., Hunt, S., Deloukas, P., Bentley, D. R., & Donnelly, P. (2004). The Fine-Scale Structure of Recombination Rate Variation in the Human Genome. *Science*, *304*(5670), 581–584. https://doi.org/10.1126/science.1092500

Mendez, F. L., Watkins, J. C., & Hammer, M. F. (2012). A Haplotype at STAT2 Introgressed from Neanderthals and Serves as a Candidate of Positive Selection in Papua New Guinea. *The American Journal of Human Genetics*, *91*(2), 265–274. https://doi.org/10.1016/j.ajhg.2012.06.015

Meyer, M., Arsuaga, J.-L., de Filippo, C., Nagel, S., Aximu-Petri, A., Nickel, B., … Pääbo, S. (2016). Nuclear DNA sequences from the Middle Pleistocene Sima de los Huesos hominins. *Nature*, *531*(7595), 504–507. https://doi.org/10.1038/nature17405

Mondal, M., Bertranpetit, J., & Lao, O. (2019). Approximate Bayesian computation with deep learning supports a third archaic introgression in Asia and Oceania. *Nature Communications*, *10*(1), 246. https://doi.org/10.1038/s41467-018-08089-7

Mondal, M., Casals, F., Xu, T., Dall'Olio, G. M., Pybus, M., Netea, M. G., … Bertranpetit, J. (2016). Genomic analysis of Andamanese provides insights into ancient human migration into Asia and adaptation. *Nature Genetics*, *48*(9), 1066–1070. https://doi.org/10.1038/ng.3621

Moore, L. G. (2017). Measuring high-altitude adaptation. *Journal of Applied Physiology*, *123*(5), 1371–1385. https://doi.org/10.1152/japplphysiol.00321.2017

Mountain, M.-J. (1991). Bulmer phase I: environmental change and human activity through the late Pleistoceneinto the Holocene in the highlands of New Guinea : A scenario. In R. Bulmer & A. Pawley (Éds.), *Man and a half : Essays in Pacific anthropology and ethnobiology in honour of Ralph Bulmer*. Auckland: Polynesian Society.

Mueller, W. H., Schull, V. N., Schull, W. J., Soto, P., & Rothhammer, F. (1978). A multinational Andean Genetic and Health Program : Growth and development in an hypoxic environment. *Annals of Human Biology*, *5*(4), 329–352. https://doi.org/10.1080/03014467800002981

Müller, I., Bockarie, M., Alpers, M., & Smith, T. (2003). The epidemiology of malaria in Papua New Guinea. *Trends in Parasitology*, *19*(6), 253–259. https://doi.org/10.1016/S1471-4922(03)00091-6

Myers, S., Bottolo, L., Freeman, C., McVean, G., & Donnelly, P. (2005). A Fine-Scale Map of Recombination Rates and Hotspots Across the Human Genome. *Science*, *310*(5746), 321–324. https://doi.org/10.1126/science.1117196

Naraqi, S., Feling, B., & Leeder, S. R. (2003). Disease and death in Papua New Guinea. *Medical Journal of Australia*, *178*(1), 7–8. https://doi.org/10.5694/j.1326-5377.2003.tb05030.x

Nordborg, M. (2019). Coalescent Theory. In D. Balding, I. Moltke, & J. Marioni (Éds.), *Handbook of Statistical Genomics* (1re éd., p. 145–30). Wiley. https://doi.org/10.1002/9781119487845.ch5

Norman, K., Inglis, J., Clarkson, C., Faith, J. T., Shulmeister, J., & Harris, D. (2018). An early colonisation pathway into northwest Australia 70–60,000 years ago. *Quaternary Science Reviews*, *180*, 229–239. https://doi.org/10.1016/j.quascirev.2017.11.023

O'Connell, J. F., & Allen, J. (2015). The process, biotic impact, and global implications of the human colonization of Sahul about 47,000 years ago. *Journal of Archaeological Science*, *56*, 73–84. https://doi.org/10.1016/j.jas.2015.02.020

O'Connell, J. F., Allen, J., Williams, M. A. J., Williams, A. N., Turney, C. S. M., Spooner, N. A., … Cooper, A. (2018). When did Homo sapiens first reach Southeast Asia and Sahul? *Proceedings of the National Academy of Sciences*, *115*(34), 8482–8490. https://doi.org/10.1073/pnas.1808385115

O'Connor, S. (2007). New evidence from East Timor contributes to our understanding of earliest modern human colonisation east of the Sunda Shelf. *Antiquity*, *81*(313), 523–535. https://doi.org/10.1017/S0003598X00095569

O'Connor, S., Louys, J., Kealy, S., & Samper Carro, S. C. (2017). Hominin Dispersal and Settlement East of Huxley's Line : The Role of Sea Level Changes, Island Size, and Subsistence Behavior. *Current Anthropology*, *58*(S17), S567–S582. https://doi.org/10.1086/694252

Ohta, T. (2017). Nearly Neutral Theory☆. In *Reference Module in Life Sciences*. Elsevier. https://doi.org/10.1016/B978-0-12-809633-8.06779-0

Orlando, L., Allaby, R., Skoglund, P., Der Sarkissian, C., Stockhammer, P. W., Ávila-Arcos, M. C., … Warinner, C. (2021). Ancient DNA analysis. *Nature Reviews Methods Primers*, *1*(1), 1–26. https://doi.org/10.1038/s43586-020-00011-0

Ortiz-Prado, E., & Dunn, J. F. (2011). *High altitude exposure and ischemic stroke : A literature review*. https://doi.org/10.11575/PRISM/10691

O'Sullivan, J. M., Preston, R. J. S., O'Regan, N., & O'Donnell, J. S. (2016). Emerging roles for hemostatic dysfunction in malaria pathogenesis. *Blood*, *127*(19), 2281–2288. https://doi.org/10.1182/blood-2015-11-636464

Pagani, L., Lawson, D. J., Jagoda, E., Mörseburg, A., Eriksson, A., Mitt, M., … Metspalu, M. (2016). Genomic analyses inform on migration events during the peopling of Eurasia. *Nature*, *538*(7624), 238–242. https://doi.org/10.1038/nature19792

Pain, A., Ferguson, D. J. P., Kai, O., Urban, B. C., Lowe, B., Marsh, K., & Roberts, D. J. (2001). Platelet-mediated clumping of Plasmodium falciparum-infected erythrocytes is a common adhesive phenotype and is associated with severe malaria. *Proceedings of the National Academy of Sciences*, *98*(4), 1805–1810. https://doi.org/10.1073/pnas.98.4.1805

Palmer, B. (2017). *The Languages and Linguistics of the New Guinea Area : A Comprehensive Guide*. Walter de Gruyter GmbH & Co KG.

Pan-UKB team. (2020). Accessed at https://pan.ukbb.broadinstitute.org

Pareek, C. S., Smoczynski, R., & Tretyn, A. (2011). Sequencing technologies and genome sequencing. *Journal of Applied Genetics*, *52*(4), 413–435. https://doi.org/10.1007/s13353-011-0057-x

Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., … Reich, D. (2012). Ancient Admixture in Human History. *Genetics*, *192*(3), 1065–1093. https://doi.org/10.1534/genetics.112.145037

Pawley, A., & Australian National University (Éds.). (2005). *Papuan pasts : Cultural, linguistic and biological histories of Papuan-speaking peoples*. Canberra: Pacific Linguistics, Research School of Pacific and Asian Studies, Australian National University.

Payne, S., Kumar Bc, R., Pomeroy, E., Macintosh, A., & Stock, J. (2018). Thrifty phenotype versus cold adaptation : Trade-offs in upper limb proportions of Himalayan populations of Nepal. *Royal Society Open Science*, *5*(6), 172174. https://doi.org/10.1098/rsos.172174

Pedro, N., Brucato, N., Fernandes, V., André, M., Saag, L., Pomat, W., … Ricaut, F.-X. (2020). Papuan mitochondrial genomes and the settlement of Sahul. *Journal of Human Genetics*, 1–13. https://doi.org/10.1038/s10038-020-0781-3

Peng, C., Chen, H., Li, Y., Yang, H., Qin, P., Ma, B., … Guo, D. (2021). LRIG3 Suppresses Angiogenesis by Regulating the PI3K/AKT/VEGFA Signaling Pathway in Glioma. *Frontiers in Oncology*, *11*. Accessed at https://www.frontiersin.org/article/10.3389/fonc.2021.621154

Peter Bellwood & Eusebio Dizon. (2005). The Batanes Archaeological Project and the Out of Taiwan Hypothesis for Austronesian Dispersal. 南島研究學報, *1*(1). https://doi.org/10.29884/JAS.200506.0001

Petr, M., Pääbo, S., Kelso, J., & Vernot, B. (2019). Limits of long-term selection against Neandertal introgression. *Proceedings of the National Academy of Sciences*, *116*(5), 1639–1644. https://doi.org/10.1073/pnas.1814338116

Pickrell, J. K., & Pritchard, J. K. (2012). Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genetics*, *8*(11), e1002967. https://doi.org/10.1371/journal.pgen.1002967

Piešová, M., & Mach, M. (2020). Impact of Perinatal Hypoxia on the Developing Brain. *Physiological Research*, *69*(2), 199–213. https://doi.org/10.33549/physiolres.934198

Prüfer, K., de Filippo, C., Grote, S., Mafessoni, F., Korlević, P., Hajdinjak, M., … Pääbo, S. (2017). A high-coverage Neandertal genome from Vindija Cave in Croatia. *Science (New York, N.Y.)*, *358*(6363), 655–658. https://doi.org/10.1126/science.aao1887

Prüfer, K., Racimo, F., Patterson, N., Jay, F., Sankararaman, S., Sawyer, S., … Pääbo, S. (2014). The complete genome sequence of a Neandertal from the Altai Mountains. *Nature*, *505*(7481), 43–49. https://doi.org/10.1038/nature12886

Prugnolle, F., Manica, A., Charpentier, M., Guégan, J. F., Guernier, V., & Balloux, F. (2005). Pathogen-Driven Selection and Worldwide HLA Class I Diversity. *Current Biology*, *15*(11), 1022–1027. https://doi.org/10.1016/j.cub.2005.04.050

Quach, H., Rotival, M., Pothlichet, J., Loh, Y.-H. E., Dannemann, M., Zidane, N., … Quintana-Murci, L. (2016). Genetic Adaptation and Neandertal Admixture Shaped the Immune System of Human Populations. *Cell*, *167*(3), 643-656.e17. https://doi.org/10.1016/j.cell.2016.09.024

Racimo, F., Sankararaman, S., Nielsen, R., & Huerta-Sánchez, E. (2015). Evidence for archaic adaptive introgression in humans. *Nature Reviews Genetics*, *16*(6), 359–371. https://doi.org/10.1038/nrg3936

Rahman, A., Tabassum, T., Araf, Y., Al Nahid, A., Ullah, M. A., & Hosen, M. J. (2021). Silent hypoxia in COVID-19 : Pathomechanism and possible management strategy. *Molecular Biology Reports*, *48*(4), 3863–3869. https://doi.org/10.1007/s11033-021-06358-1

Rasmussen, A. H., Rasmussen, H. B., & Silahtaroglu, A. (2017). The DLGAP family : Neuronal expression, function and role in brain disorders. *Molecular Brain*, *10*(1), 43. https://doi.org/10.1186/s13041-017-0324-9

Rasmussen, M. D., Hubisz, M. J., Gronau, I., & Siepel, A. (2014). Genome-Wide Inference of Ancestral Recombination Graphs. *PLOS Genetics*, *10*(5), e1004342. https://doi.org/10.1371/journal.pgen.1004342

Rees, J. S., Castellano, S., & Andrés, A. M. (2020). The Genomics of Human Local Adaptation. *Trends in Genetics*, *36*(6), 415–428. https://doi.org/10.1016/j.tig.2020.03.006

Reich, D., Green, R. E., Kircher, M., Krause, J., Patterson, N., Durand, E. Y., … Pääbo, S. (2010). Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature*, *468*(7327), 1053–1060. https://doi.org/10.1038/nature09710

Reilly, P. F., Tjahjadi, A., Miller, S. L., Akey, J. M., & Tucci, S. (2022). The contribution of Neanderthal introgression to modern human traits. *Current Biology*, *32*(18), R970–R983. https://doi.org/10.1016/j.cub.2022.08.027

Riley, I. D. (1983). Population change and distribution in Papua New Guinea : An epidemiological approach. *Journal of Human Evolution*, *12*(1), 125–132. https://doi.org/10.1016/S0047-2484(83)80017-7

Rupert, J. L., & Hochachka, P. W. (2001). The Evidence for Hereditary Factors Contributing to High Altitude Adaptation in Andean Natives : A Review. *High Altitude Medicine & Biology*, *2*(2), 235–256. https://doi.org/10.1089/152702901750265332

Sabeti, P. C. (2006). Positive Natural Selection in the Human Lineage. *Science*, *312*(5780), 1614–1620. https://doi.org/10.1126/science.1124309

Sabeti, Pardis C., Varilly, P., Fry, B., Lohmueller, J., Hostetter, E., Cotsapas, C., … Lander, E. S. (2007). Genome-wide detection and characterization of positive selection in human populations. *Nature*, *449*(7164), 913–918. https://doi.org/10.1038/nature06250

Sams, A. J., Dumaine, A., Nédélec, Y., Yotova, V., Alfieri, C., Tanner, J. E., … Barreiro, L. B. (2016). Adaptively introgressed Neandertal haplotype at the OAS locus functionally impacts innate immune responses in humans. *Genome Biology*, *17*(1), 246. https://doi.org/10.1186/s13059-016-1098-6

Sankararaman, S., Mallick, S., Dannemann, M., Prüfer, K., Kelso, J., Pääbo, S., … Reich, D. (2014). The genomic landscape of Neanderthal ancestry in present-day humans. *Nature*, *507*(7492), 354–357. https://doi.org/10.1038/nature12961

Sankararaman, S., Mallick, S., Patterson, N., & Reich, D. (2016). The Combined Landscape of Denisovan and Neanderthal Ancestry in Present-Day Humans. *Current Biology*, *26*(9), 1241–1247. https://doi.org/10.1016/j.cub.2016.03.037

Scheinfeldt, L. B., Soi, S., Thompson, S., Ranciaro, A., Woldemeskel, D., Beggs, W., … Tishkoff, S. A. (2012). Genetic adaptation to high altitude in the Ethiopian highlands. *Genome Biology*, *13*(1), R1. https://doi.org/10.1186/gb-2012-13-1-r1

Schiffels, S., & Durbin, R. (2014). Inferring human population size and separation history from multiple genome sequences. *Nature genetics*, *46*(8), 919–925. https://doi.org/10.1038/ng.3015

Ségurel, L., & Bon, C. (2017). On the Evolution of Lactase Persistence in Humans. *Annual Review of Genomics and Human Genetics*, *18*(1), 297–319. https://doi.org/10.1146/annurev-genom-091416-035340

Sella, G., & Barton, N. H. (2019). Thinking About the Evolution of Complex Traits in the Era of Genome-Wide Association Studies. *Annual Review of Genomics and Human Genetics*, *20*(1), 461–493. https://doi.org/10.1146/annurev-genom-083115-022316

Senn, N., Maraga, S., Sie, A., Rogerson, S. J., Reeder, J. C., Siba, P., & Mueller, I. (2010). Population Hemoglobin Mean and Anemia Prevalence in Papua New Guinea : New Metrics for Defining Malaria Endemicity? *PLOS ONE*, *5*(2), e9375. https://doi.org/10.1371/journal.pone.0009375

Shaffer, J. P. (1995). Multiple Hypothesis Testing. *Annual Review of Psychology*, *46*(1), 561–584. https://doi.org/10.1146/annurev.ps.46.020195.003021

Siddle, K. J., & Quintana-Murci, L. (2014). The Red Queen's long race : Human adaptation to pathogen pressure. *Current Opinion in Genetics & Development*, *29*, 31–38. https://doi.org/10.1016/j.gde.2014.07.004

Silubonde, T. M., Baumgartner, J., Ware, L. J., Malan, L., Smuts, C. M., & Norris, S. (2020). Adjusting Haemoglobin Values for Altitude Maximizes Combined Sensitivity and Specificity to Detect Iron Deficiency among Women of Reproductive Age in Johannesburg, South Africa. *Nutrients*, *12*(3), 633. https://doi.org/10.3390/nu12030633

Simonson, T. S., Yang, Y., Huff, C. D., Yun, H., Qin, G., Witherspoon, D. J., … Ge, R. (2010). Genetic Evidence for High-Altitude Adaptation in Tibet. *Science*, *329*(5987), 72–75. https://doi.org/10.1126/science.1189406

Skoglund, P., Posth, C., Sirak, K., Spriggs, M., Valentin, F., Bedford, S., … Reich, D. (2016). Genomic insights into the peopling of the Southwest Pacific. *Nature*, *538*(7626), 510–513. https://doi.org/10.1038/nature19844

Skov, L., Hui, R., Shchur, V., Hobolth, A., Scally, A., Schierup, M. H., & Durbin, R. (2018). Detecting archaic introgression using an unadmixed outgroup. *PLOS Genetics*, *14*(9), e1007641. https://doi.org/10.1371/journal.pgen.1007641

Smith, J. M. B. (2001). Did early hominids cross sea gaps on natural rafts. In *Faunal and Floral Migration and Evolution in SE Asia-Australasia* (p. 409–416). CRC Press.

Souilmi, Y., Lauterbur, M. E., Tobler, R., Huber, C. D., Johar, A. S., Moradi, S. V., … Enard, D. (2021). An ancient viral epidemic involving host coronavirus interacting genes more than 20,000 years ago in East Asia. *Current Biology*, *31*(16), 3504–3514.e9. https://doi.org/10.1016/j.cub.2021.05.067

Speidel, L., Forest, M., Shi, S., & Myers, S. R. (2019). A method for genome-wide genealogy estimation for thousands of samples. *Nature Genetics*, *51*(9), 1321–1329. https://doi.org/10.1038/s41588-019-0484-x

Stearns, F. W. (2010). One Hundred Years of Pleiotropy : A Retrospective. *Genetics*, *186*(3), 767–773. https://doi.org/10.1534/genetics.110.122549

Stembridge, M., Williams, A. M., Gasho, C., Dawkins, T. G., Drane, A., Villafuerte, F. C., … Ainslie, P. N. (2019). The overlooked significance of plasma volume for successful adaptation to high altitude in Sherpa and Andean natives. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(33), 16177–16179. https://doi.org/10.1073/pnas.1909002116

Stern, A. J., & Nielsen, R. (2019). Detecting Natural Selection. In *Handbook of Statistical Genomics* (p. 397–40). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781119487845.ch14

Stern, A. J., Wilton, P. R., & Nielsen, R. (2019). An approximate full-likelihood method for inferring selection and allele frequency trajectories from DNA sequence data. *PLOS Genetics*, *15*(9), e1008384. https://doi.org/10.1371/journal.pgen.1008384

Summerhayes, G. R., Field, J. H., Shaw, B., & Gaffney, D. (2017). The archaeology of forest exploitation and change in the tropics during the Pleistocene : The case of Northern Sahul (Pleistocene New Guinea). *Quaternary International*, *448*, 14–30. https://doi.org/10.1016/j.quaint.2016.04.023

Summerhayes, G. R., Leavesley, M., Fairbairn, A., Mandui, H., Field, J., Ford, A., & Fullagar, R. (2010). Human Adaptation and Plant Use in Highland New Guinea 49,000 to 44,000 Years Ago. *Science*, *330*(6000), 78–81. https://doi.org/10.1126/science.1193130

Sutikna, T., Tocheri, M. W., Faith, J. T., Jatmiko, Due Awe, R., Meijer, H. J. M., … Roberts, R. G. (2018). The spatio-temporal distribution of archaeological and faunal finds at Liang Bua (Flores, Indonesia) in light of the revised chronology for Homo floresiensis. *Journal of Human Evolution*, *124*, 52–74. https://doi.org/10.1016/j.jhevol.2018.07.001

Szpiech, Z. A., & Hernandez, R. D. (2014). selscan : An efficient multithreaded program to perform EHH-based scans for positive selection. *Molecular Biology and Evolution*, *31*(10), 2824–2827. https://doi.org/10.1093/molbev/msu211

Szpiech, Z. A., Novak, T. E., Bailey, N. P., & Stevison, L. S. (2021). Application of a novel haplotype-based scan for local adaptation to study high-altitude adaptation in rhesus macaques. *Evolution Letters*, *5*(4), 408–421. https://doi.org/10.1002/evl3.232

Terrell, J. E. (2004). The « sleeping giant » hypothesis and New Guinea's place in the prehistory of Greater Near Oceania. *World Archaeology*, *36*(4), 601–609. https://doi.org/10.1080/0043824042000303782

The 1000 Genomes Project Consortium, Gibbs, R. A., Boerwinkle, E., Doddapaneni, H., Han, Y., Korchina, V., … Rasheed, A. (2015). A global reference for human genetic variation. *Nature*, *526*(7571), 68–74. https://doi.org/10.1038/nature15393

The Nobel Prize in Physiology or Medicine 2022. (2022). Consulté 8 mars 2023, à l'adresse NobelPrize.org website: https://www.nobelprize.org/prizes/medicine/2022/summary/

Tobler, R., Rohrlach, A., Soubrier, J., Bover, P., Llamas, B., Tuke, J., … Cooper, A. (2017). Aboriginal mitogenomes reveal 50,000 years of regionalism in Australia. *Nature*, *544*(7649), 180–184. https://doi.org/10.1038/nature21416

Totino, P. R. R., Daniel-Ribeiro, C. T., & Ferreira-da-Cruz, M. de F. (2016). Evidencing the Role of Erythrocytic Apoptosis in Malarial Anemia. *Frontiers in Cellular and Infection Microbiology*, *6*. https://doi.org/10.3389/fcimb.2016.00176

Trájer, A. J. (2022). Late Quaternary changes in malaria-free areas in Papua New Guinea and the future perspectives. *Quaternary International*. https://doi.org/10.1016/j.quaint. 2022. 04.003

Trájer, A. J., Sebestyén, V., & Domokos, E. (2020). The potential impacts of climate factors and malaria on the Middle Palaeolithic population patterns of ancient humans. *Quaternary International*, *565*, 94–108. https://doi.org/10.1016/j.quaint.2020.10.056

Tran, H., Tanaka, A., Litvinovich, S. V., Medved, L. V., Haudenschild, C. C., & Argraves, W. S. (1995). The Interaction of Fibulin-1 with Fibrinogen : A POTENTIAL ROLE IN HEMOSTASIS AND THROMBOSIS (∗). *Journal of Biological Chemistry*, *270*(33), 19458–19464. https://doi.org/10.1074/jbc.270.33.19458

Tretina, K., Park, E.-S., Maminska, A., & MacMicking, J. D. (2019). Interferon-induced guanylate-binding proteins : Guardians of host defense in health and disease. *The Journal of Experimental Medicine*, *216*(3), 482–500. https://doi.org/10.1084/jem.20182031

van den Bergh, G. D., Li, B., Brumm, A., Grün, R., Yurnaldi, D., Moore, M. W., … Morwood, M. J. (2016). Earliest hominin occupation of Sulawesi, Indonesia. *Nature*, *529*(7585), 208–211. https://doi.org/10.1038/nature16448

Van Valen, L. (1973). A New Evolutionary Law. *Evolutionary Theory*, *1*, 1–30.

Vattathil, S., & Akey, J. M. (2015). Small Amounts of Archaic Admixture Provide Big Insights into Human History. *Cell*, *163*(2), 281–284. https://doi.org/10.1016/j.cell.2015.09.042

Vernot, B., & Akey, J. M. (2014). Resurrecting Surviving Neandertal Lineages from Modern Human Genomes. *Science*, *343*(6174), 1017–1021. https://doi.org/10.1126/science.1245938

Vernot, B., Tucci, S., Kelso, J., Schraiber, J. G., Wolf, A. B., Gittelman, R. M., … Akey, J. M. (2016). Excavating Neandertal and Denisovan DNA from the genomes of Melanesian individuals. *Science (New York, N.Y.)*, *352*(6282), 235–239. https://doi.org/10.1126/ science.aad9416

Vespasiani, D. M., Jacobs, G. S., Cook, L. E., Brucato, N., Leavesley, M., Kinipi, C., … Romero, I. G. (2022). Denisovan introgression has shaped the immune system of present-day Papuans. *PLOS Genetics*, *18*(12), e1010470. https://doi.org/10.1371/journal.pgen.1010470

Villafuerte, F. C., & Corante, N. (2016). Chronic Mountain Sickness : Clinical Aspects, Etiology, Management, and Treatment. *High Altitude Medicine & Biology*, *17*(2), 61–69. https://doi.org/10.1089/ham.2016.0031

Voight, B. F., Kudaravalli, S., Wen, X., & Pritchard, J. K. (2006). A Map of Recent Positive Selection in the Human Genome. *PLOS Biology*, *4*(3), e72. https://doi.org/10.1371/journal.pbio.0040072

Weitz, C. A., Garruto, R. M., Chin, C. T., & Liu, J. C. (2004). Morphological growth and thorax dimensions among Tibetan compared to Han children, adolescents and young adults born and raised at high altitude. *Annals of Human Biology*, *31*(3), 292–310. https://doi.org/10.1080/03014460420000196316

White, N. J. (2018). Anaemia and malaria. *Malaria Journal*, *17*(1), 371. https://doi.org/10.1186/s12936-018-2509-9

Woolcock, A. J., Colman, M. H., & Blackburn, C. R. B. (1972). Factors Affecting Normal Values for Ventilatory Lung Function1,2. *American Review of Respiratory Disease*. (world). https://doi.org/10.1164/arrd.1972.106.5.692

World Health Organization. (2022a). *World malaria report 2022*. Geneva: World Health Organization. Accessed at https://www.who.int/teams/global-malaria-programme/reports/world-malaria-report-2022

World Health Organization. (2022b, janvier 28). Child mortality (under 5 years). Consulté 27 février 2023, à l'adresse https://www.who.int/news-room/fact-sheets/detail/levels-and-trends-in-child-under-5-mortality-in-2020

Yermakovich, D., Pankratov, V., Võsa, U., Yunusbayev, B., Estonian Biobank Research Team, & Dannemann, M. (2022). Long-range regulatory effects of Neandertal DNA in modern humans. *Genetics*, iyac188. https://doi.org/10.1093/genetics/iyac188

Yi, X., Liang, Y., Huerta-Sanchez, E., Jin, X., Cuo, Z. X. P., Pool, J. E., … Wang, J. (2010). Sequencing of 50 Human Exomes Reveals Adaptation to High Altitude. *Science*, *329*(5987), 75–78. https://doi.org/10.1126/science.1190371

Yuan, K., Ni, X., Liu, C., Pan, Y., Deng, L., Zhang, R., … Xu, S. (2021). Refining models of archaic admixture in Eurasia with ArchaicSeeker 2.0. *Nature Communications*, *12*(1), 6232. https://doi.org/10.1038/s41467-021-26503-5

Zhai, W., Nielsen, R., & Slatkin, M. (2009). An Investigation of the Statistical Power of Neutrality Tests Based on Comparative and Population Genetic Data. *Molecular Biology and Evolution*, *26*(2), 273–283. https://doi.org/10.1093/molbev/msn231

Zhang, X., Witt, K. E., Bañuelos, M. M., Ko, A., Yuan, K., Xu, S., … Huerta-Sanchez, E. (2021). The history and evolution of the Denisovan-EPAS1 haplotype in Tibetans. *Proceedings of the National Academy of Sciences*, *118*(22), e2020803118. https://doi.org/10.1073/pnas.2020803118

Zhang, Y., & Xiang, Y. (2016). Molecular and cellular basis for the unique functioning of Nrf1, an indispensable transcription factor for maintaining cell homoeostasis and organ integrity. *Biochemical Journal*, *473*(8), 961–1000. https://doi.org/10.1042/BJ20151182

Zhou, H., Cao, J., Yang, F., Fan, D., Li, H., Fan, T., & Sun, P. (2021). Member Domain 3 (LRIG3) Activates Hypoxia-Inducible Factor-1 α /Vascular Endothelial Growth Factor (HIF-1 α /VEGF) Pathway to Inhibit the Growth of Bone Marrow Mesenchymal Stem Cells in Glioma. *Journal of Biomaterials and Tissue Engineering*, *11*(5), 1022–1027. https://doi.org/10.1166/jbt.2021.2629

Zhou, X., & Stephens, M. (2012). Genome-wide efficient mixed-model analysis for association studies. *Nature Genetics*, *44*(7), 821–824. https://doi.org/10.1038/ng.2310

ACKNOWLEDGMENTS

I am sincerely grateful for everyone who contributed their time, expertise, and friendship, making this achievement possible.

I first wish to thank my supervisors, Dr Mayukh Mondal and Dr François-Xavier Ricaut, for their support and mentorship throughout my PhD project. I also want to thank Dr Nicolas Brucato for his guidance. A huge thanks to Professor Mait Metspalu for giving me the opportunity to do my research in such a peaceful place. I can't forget to mention all the co-authors I enjoyed working with over the years.

I owe a massive thank you to Vasili, Michael, and Anders for always taking the time to provide super helpful feedback on my research projects. I couldn't have finished my PhD on time without the countless assistance from Kristiina and Monika. They were always there for me, checking in and keeping me sane when deadlines seemed impossible.

Now, let's talk about my fellow roommates. First and foremost, I want to give a huge thanks to Stefania with whom I embarked on this PhD journey. We've had so many unforgettable adventures together, even if we did get lost more times than we can count! I'm also incredibly grateful to all the other wonderful people inhabiting the "PhD room", an exotic place where we encountered freezing winter and boiling summer days. So, a big thank you to Bianca, Tina, Ludovica, Katri, Linda, Anne-Mai, Danat, Ivan, Kadri, Hovik, and everyone who graced that vibrant space. But brilliant people aren't found only in this wonderful room; I'd like to thank Ajai, Toni, Ivan, and Rodrigo for their contagious joy and entertaining conversation.

I want to give a special shout-out to the aDNA team for including me in so many fantastic activities. Many thanks to Anu, Helja, Lehti, Lena, Olga, Marcel, Meriam, Freddi and Kristiina for teaching me how real Estonian birthdays should be celebrated.

Jüri, your perfectly timed and amazing photographs always brought a smile to my face. Thanks for capturing those special moments and sharing your incredible stories. And a big thank you to Merilin, Mariza, and Kaisa for their endless patience whenever I was lost in paperwork.

I also want to thank Monika and Anne-Mai for their help with the Estonian translations, and Professor Reedik Mägi for kindly agreeing to be the internal reviewer.

I am very grateful to Professor Lluis Quintana-Murci and Dr Etienne Patin and the Unit of Human Evolutionary Genetics of the Institute Pasteur for hosting me in their lab; it was a blast to collaborate with so many great minds.

This PhD adventure would never have started without Professor Andrés Ruiz-Linares and Dr Pierre Faux, who gave me the love of doing good research. Working with such kind and brilliant researchers raised my expectation in research. I'll always be grateful for the opportunity they offer me with such a great master's project.

Thank you, Ciara, Barbara, and Mai. You are all such inspiring women. You brought so much fun (and good food!) to Estonia. Thank you for all the great times and for welcoming me to your home when I desperately needed a change of scenery.

Pour finir en beauté, un énorme merci Morgane de me soutenir depuis toutes ces années. Il n'y a pas de mots assez fort pour décrire le pilier que tu es pour moi. Pauline et Clément, merci d'êtres toujours disponible pour une bière et me faire me sentir comme à la maison à chacun de mes retours! Merci à Juliette, Claire, Manon et Julie d'être toujours présentes malgré les kilomètres.

Merci Charlotte pour ton soutien et tes recommandation de lectures, Benjamin pour ton esprit de contradiction rafraichissant et Clarisse pour ton inimitable joie de vivre et énergie.

Un grand merci à Papili et Papy Roger pour leur support sans faille. J'ai également une pensée pour Mamie Dani et Mamili, qui sont une immense source d'inspiration par leur force de caractère et bienveillance.

Enfin je remercie mes parents de m'avoir supportée dans tous mes projets depuis que je leur exposé mon plan de carrière il y a plus de 10 ans.

PUBLICATIONS

CURRICULUM VITAE

Name: Mathilde Frédérique E. André
Date of birth: November 9, 1995
Nationality: Belgian
Address: University of Tartu, Institute of Genomics, Riia 23b, 51010 Tartu, Estonia
E-mail: mathilde.andre@ut.ee

Education:

2019–2023 Doctoral studies, University of Tartu, Faculty of Science and Technology, Institute of Genomics, Chair of Evolutionary Biology

2017–2019 MSc, Evolutionary Anthropology ("Master Humanités Médicales"), University of Aix-Marseille (AMU)

2013–2017 BSc, Major in Biological Sciences with Minor in History of Art and Archaeology, Catholic University of Louvain (UCLouvain)

Professional employment:

2021–2023 Junior Research Fellow of Modern Population

Teaching:

2020–2021 Teaching assistant, University of Tartu, Bachelor in Science and Technology, Chair of Evolutionary Biology, Evolution and the Natural World course

International Courses and Conferences:

2023 Society for Molecular Biology and Evolution meeting (SMBE), Ferrara, Italy
Poster: Genomic Signatures of Local Adaptation to Extreme Environments in Papua New Guinean Highlanders and Lowlanders

2023 The Biology of Genomes, Cold Spring Laboratory (CSHL), USA
Poster: Positive selection in the genomes of two Papua New Guinean populations at distinct altitude levels

2022 Congress of the Indo-Pacific Prehistory Association (IPPA), Chiang Mai, Thaïland
Oral presentation: Phenotype Adaptation and Positive Selection to Altitude in Papua New Guinean Highlanders

2022 EMBO | EMBL Symposium: Reconstructing the human past: using ancient and modern genomics, Heidelberg, Germany
Flask talk and poster: Papuan New Guineans blood composition is affected by positive selection

2021 Australian Archeological Association conference, online
Oral presentation: Papua New Guineans Show Signs of Biological Adaptations to Altitude

2021 Wellcome Connecting Science conference– Human Evolution– From Fossils to Ancient and Modern Genomes, online
Poster: Natural Selection and phenotypic associations in Papuan Highlanders

2021 ASHG virtual meeting 2021, online
Poster: Positive selection to altitude and phenotypic associations in Papua New Guinea highlanders

2021 EMBO Course: Population genomics: Background, tools, and programming, online

2021 1st Virtual Conference for Women Archaeologists and Paleontologists, online
Oral presentation: Papua New Guineans show unique phenotypic traits at altitude

2021 Introduction to the Statistical Analysis of Genome-wide Association Studies, held by the University of Surrey, online

Awards:

2022 Dora Pluss Scholarship

Publications:

2023 **M André**, N Brucato, G Hudjasov, V Pankratov, D Yermakovich, R Kreevan, J Kariwiga, J Muke, A Boland, J-F Deleuze, V Meyer, N Evans, M P Cox, M Leavesley, M Dannemann, T Org, M Metspalu, M Mondal, F-X Ricaut, *Positive selection in the genomes of two Papua New Guinean populations at distinct altitude levels*, bioRxiv, [preprint], 2022 Dec, https://doi.org/10.1101/2022.12.15.520226

2022 N Brucato, **M André**, G Hudjashov, M Mondal, M P Cox, M Leavesley, F-X Ricaut, *Chronology of natural selection in Oceanian genomes*, IScience, 2022 Jun, 25(7), https://doi.org/10.1016/j.isci.2022.104583

2021 Y Chen, **M André**, K Adhikari, M Blin, B Bonfante, J Mendoza-Revilla, M Fuentes-Guajardo, S Palmal, J C Chacón-Duque, M Hurtado, V Villegas, V Granja, C Jaramillo, W Arias, R B Lozano, P Everardo-Martínez, J Gómez-Valdés, H Villamil-Ramírez, C C S de Cerqueira, T Hünemeier, V Ramallo, R Gonzalez-José, L Schüler-Faccini, M-C Bortolini, V Acuña-Alonzo, S Canizales-Quinteros, C Gallo, G Poletti, G Bedoya, F Rothhammer, D Balding, D J Tobin, S Wang, P Faux 2, A Ruiz-Linares; *A genome-wide association study identifies novel gene associations with facial skin wrinkling and mole count in Latin Americans*; British Journal of Dermatology; 2021 Nov; 18(5):988-998; https://doi.org/10.1111/bjd.20436

2021 N Brucato, **M André**, R Tsang, L Saag, J Kariwiga, K Sesuki, T Beni, W Pomat, J Muke, V Meyer, A Boland, J-F Deleuze, H Sudoyo, M Mondal, L Pagani, I Gallego Romero, M Metspalu, M P Cox, M Leavesley, F-X Ricaut; *Papua New Guinean Genomes Reveal the Complex Settlement of North Sahul*; Molecular Biology and Evolution; 2021 Oct; 38(11):5107–5121; doi:10.1093/molbev/msab238

2021 **M André**, N Brucato, S Plutniak, J Kariwiga, J Muke, A Morez, M Leavesley, M Mondal, F-X Ricaut; *Phenotypic differences between highlanders and lowlanders in Papua New Guinea*; PLOS ONE; 2021 Jul; 16(7): e0253921; doi:10.1371/journal.pone.0253921

2020 N Pedro, N Brucato, V Fernandes, **M André**, L Saag, W Pomat, C Besse, A Boland, J-F Deleuze, C Clarkson, H Sudoyo, M Metspalu, M Stoneking, M P Cox, M Leavesley, L Pereira, F-X Ricaut; *Papuan mitochondrial genomes and the settlement of Sahul*; Journal of Human Genetics, 2020 Jun; 65:875:887; doi: 10.1038/s10038-020-0781-3

ELULOOKIRJELDUS

Nimi: Mathilde Frédérique E. André
Sünniaeg: 9. november, 1995
Rahvus: Belglane
Aadress: Genoomika instituut, Tartu Ülikool, Riia 23b, 51010 Tartu, Eesti
E-post: mathilde.andre@ut.ee

Haridustee:

2019–2023 Doktoriõpingud, Genoomika instituut, Loodus- ja täppisteaduste valdkond, Tartu Ülikool
2017–2019 Magistrikraad (*MSc*), evolutsiooniline antropoloogia ("Master Humanités Médicales"), Aix-Marseille Ülikool (AMU), Prantsusmaa
2013–2017 Bakalaureusekraad (*BSc)*, bioloogia, kõrvaleriala kunstiajalugu ja arheoloogia, Louvaini Katoliiklik Ülikool (UCLouvain), Belgia

Erialane tööhõive:

2021–2023 Kaasaegse populatsioonigeneetika nooremteadur

Õpetamine:

2020–2021 Õppeassistent, Evolutsioonilise bioloogia õppetool, Molekulaar- ja Rakubioloogia instituut, Loodus- ja täppisteaduste valdkond, Tartu Ülikool. Bakalaureuseaine "Evolutsioon ja loodus"

Rahvusvahelised kursused ja konverentsid:

2023 Molekulaarbioloogia ja evolutsiooniühingu konverents, Ferrara, Itaalia.
Poster: Genoomsed mustrid ja piirkondlik kohastumus ekstreemsetele keskkonnatingimustele Paapua Uus-Guinea mäestiku- ja tasandikuelanikel.

2023 Genoomide bioloogia, Cold Spring Harbour Laboratory (CSHL), USA
Poster: Paapua Uus-Guinea eri kõrgustel elavatele populatsioonidele mõjuv positiivne valik .

2022 India ja Vaikse ookeani piirkonna esiajaloo ühingu kongress (IPPA), Chiang Mai, Taimaa
Suuline ettekanne: Kõrgusest tingitud fenotüübi kohastumine ja positiivne valik Paapua Uus-Guinea mäestikuelanikel.

2022 EMBO | EMBL Sümpoosium: Inimese mineviku rekonstrueerimine vana-DNA abil. Heidelberg, Saksamaa
Välkettekanne ja poster: Positiivse valiku mõju Paapua Uus-Guinea elanike verekoostisele.

2021 Austraalia arheoloogide ühingu konverents, veebipõhine.
Suuline ettekanne: Paapua-uusguinealased on kohanenud kõrgel elamiseks.

2021 Wellcome Connecting Science konverents:Inimese evolutsioon – fossiilidest vanade ja kaasaegsete genoomideni, veebipõhine.
Poster: Looduslik valik ja fenotüüpide seosed Paapua mäestikuelanikel.

2021 ASHG virtuaalkonverents 2021, veebipõhine
Poster: Positiivne valik mäestikes elamiseks ja fenotüübi seosed Paapua Uus-Guinea mäestikuelanikel.
2021 EMBO kursus: Populatsioonigenoomika: taust, tööriistad ja programmid. Veebipõhine.
2021 Esimene virtuaalkonverents naisarheoloogidele ja -paleontoloogidele. Veebipõhine.
Suuline ettekanne: Paapua Uus-Guinea mäestikuelanikeiseloomulikud tunnused
2021 Sissejuhatus ülegenoomsete assotsiatsiooniuuringute statistikameetoditesse. Surrey Ülikool, veebipõhine.

Teaduspreemiad ja tunnustused:
2022 Dora Pluss stipendium

Publikatsioonid: Loetletud inglisekeelse CV rubriigis publikatsioonid ('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.
387. **Linda Ongaro**. A genomic portrait of American populations. Tartu, 2021, 182 p.
388. **Kairi Raime**. The identification of plant DNA in metagenomic samples. Tartu, 2021, 108 p.
389. **Heli Einberg**. Non-linear and non-stationary relationships in the pelagic ecosystem of the Gulf of Riga (Baltic Sea). Tartu, 2021, 119 p.
390. **Mickaël Mathieu Pihain**. The evolutionary effect of phylogenetic neighbourhoods of trees on their resistance to herbivores and climatic stress. Tartu, 2022, 145 p.
391. **Annika Joy Meitern**. Impact of potassium ion content of xylem sap and of light conditions on the hydraulic properties of trees. Tartu, 2022, 132 p.
392. **Elise Joonas**. Evaluation of metal contaminant hazard on microalgae with environmentally relevant testing strategies. Tartu, 2022, 118 p.
393. **Kreete Lüll**. Investigating the relationships between human microbiome, host factors and female health. Tartu, 2022, 141 p.
394. **Triin Kaasiku**. A wader perspective to Boreal Baltic coastal grasslands: from habitat availability to breeding site selection and nest survival. Tartu, 2022, 141 p.
395. **Meeli Alber**. Impact of elevated atmospheric humidity on the structure of the water transport pathway in deciduous trees. Tartu, 2022, 170 p.
396. **Ludovica Molinaro**. Ancestry deconvolution of Estonian, European and Worldwide genomic layers: a human population genomics excavation. Tartu, 2022, 138 p.
397. **Tina Saupe**. The genetic history of the Mediterranean before the common era: a focus on the Italian Peninsula. Tartu, 2022, 165 p.
398. **Mari-Ann Lind**. Internal constraints on energy processing and their consequences: an integrative study of behaviour, ornaments and digestive health in greenfinches. Tartu, 2022, 137 p.
399. **Markus Valge**. Testing the predictions of life history theory on anthropometric data. Tartu, 2022, 171 p.
400. **Ants Tull**. Domesticated and wild mammals as reservoirs for zoonotic helminth parasites in Estonia. Tartu, 2022, 152 p.

401. **Saleh Rahimlouye Barabi**. Investigation of diazotrophic bacteria association with plants. Tartu, 2022, 137 p.
402. **Farzad Aslani**. Towards revealing the biogeography of belowground diversity. Tartu, 2022, 124 p.
403. **Nele Taba**. Diet, blood metabolites, and health. Tartu, 2022, 163 p.
404. **Katri Pärna**. Improving the personalized prediction of complex traits and diseases: application to type 2 diabetes. Tartu, 2022, 190 p.
405. **Silva Lilleorg**. Bacterial ribosome heterogeneity on the example of bL31 paralogs in *Escherichia coli.* Tartu, 2022, 189 p.
406. **Oliver Aasmets**. The importance of microbiome in human health. Tartu, 2022, 123 p.
407. **Henel Jürgens**. Exploring post-translational modifications of histones in RNA polymerase II-dependent transcription. Tartu, 2022, 147 p.
408. **Mari Tagel**. Finding novel factors affecting the mutation frequency: a case study of tRNA modification enzymes TruA and RluA. Tartu, 2022, 176 p.
409. **Marili Sell**. The impact of environmental change on ecophysiology of hemiboreal tree species – acclimation mechanisms in belowground. Tartu, 2022, 163 p.
410. **Kaarin Hein**. The hissing behaviour of Great Tit (*Parus major*) females reflects behavioural phenotype and breeding success in a wild population. Tartu, 2022, 96 p.
411. **Maret Gerz**. The distribution and role of mycorrhizal symbiosis in plant communities. Tartu, 2022, 206 p.
412. **Kristiina Nõomaa**. Role of invasive species in brackish benthic community structure and biomass changes. Tartu, 2023, 151 p.
413. **Anton Savchenko**. Taxonomic studies in Dacrymycetes: *Cerinomyces* and allied taxa. Tartu, 2023, 181 p.
414. **Ahto Agan**. Interactions between invasive pathogens and resident mycobiome in the foliage of trees. Tartu, 2023, 155 p.
415. **Diego Pires Ferraz Trindade**. Dark diversity dynamics linked to global change: taxonomic and functional perspective. Tartu, 2023, 134 p.
416. **Madli Jõks**. Biodiversity drivers in oceanic archipelagos and habitat fragments, explored by agent-based simulation models. Tartu, 2023, 116 p.
417. **Ciara Baines**. Adaptation to oncogenic pollution and natural cancer defences in the aquatic environment. Tartu, 2023, 164 p.
418. **Rain Inno**. Placental transcriptome and miRNome in normal and complicated pregnancies. Tartu, 2023, 145 p.
419. **Daniyal Gohar**. Diversity, genomics, and potential functions of fungus-inhabiting bacteria. Tartu, 2023, 138 p.
420. **Sirli Rosendahl**. Fitness effects of chromosomal toxin-antitoxin systems in *Pseudomonas putida.* Tartu, 2023, 154 p.