



HAL
open science

Origins and adaptation in humans : a case study of taste and lifestyle

Agnès Sjöstrand

► **To cite this version:**

Agnès Sjöstrand. Origins and adaptation in humans : a case study of taste and lifestyle. Human genetics. Université Pierre et Marie Curie - Paris VI; Uppsala universitet, 2015. English. NNT : 2015PA066724 . tel-01609900

HAL Id: tel-01609900

<https://theses.hal.science/tel-01609900v1>

Submitted on 4 Oct 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Université Pierre et Marie Curie Université d'Uppsala

ED227 : Sciences de la Nature et de l'Homme : Écologie et
Évolution

*UMR7206 Eco-Anthropologie et Ethnobiologie,
Evolutionary Biology Center (Uppsala University)
UMR5525 Laboratoire TIMC-IMAG.*

Origins and Adaptation in Humans *a Case Study of Taste and Lifestyle*

Par Agnès Sjöstrand

Thèse de doctorat de Génétique des populations

Dirigée par Pr. Evelyne Heyer, Pr. Mattias Jakobsson &
Dr. Michael Blum

Présentée et soutenue publiquement le 20 novembre 2015

Devant un jury composé de :

David Comas, professeur (rapporteur)

Luis Quintana-Murci, directeur de recherche (rapporteur)

Thierry Wirth, professeur (examinateur)

Denis Pierron, chargé de recherche (examinateur)

Åsa Johansson, chercheuse (examinatrice)

Evelyne Heyer, professeure (directrice de thèse)

Mattias Jakobsson, professeur (directeur de thèse)

Michaël Blum, directeur de recherche (co-directeur de thèse)



Except where otherwise noted, this work is licensed under
<http://creativecommons.org/licenses/by-nc-nd/3.0/>

Ubuntu (Nguni word)
Shosholozza (Ndebele word)

List of Papers

This thesis is based on the following papers, which are referred to in the text by their Roman numerals.

- I Sjödin, P., Sjöstrand, A. E., Jakobsson, M., Blum, M. G. B. (2012) Resequencing data provide no evidence for a human bottleneck in Africa during the penultimate glacial period. *Molecular biology and evolution*, 29(7): 1851-1860.
- II Sjöstrand, A. E., Sjödin, P., & Jakobsson, M. (2014) Private haplotypes can reveal local adaptation. *BMC genetics*, 15(1): 61.
- III Sjöstrand, A.E., Sjödin P., Naidoo T., Schlebusch C. M., Gattepaille L., Hollfelder N., Günther T., Jakobsson M., Patterns of local adaptation in Africans using genome data (*manuscript*).
- IV Sjöstrand, A.E., Sjödin, P., Heyer, E., Blum, M.G.B., Jakobsson, M., Taste and lifestyle: insights from SNP-chip data, (*manuscript*).

Reprints were made with permission from the respective publishers.

Contents

Introduction.....	13
Origins of <i>Homo sapiens</i>	13
Out of Africa	15
Population structure in Africa	17
Peopling of Central Asia	20
Lifestyle transitions in Human history	21
Adaptation in modern Humans.....	24
Taste and diet	26
Mechanisms of taste perception in humans.....	28
Background and methods.....	31
Genetic variation	31
Evolutionary forces and equilibrium models	33
The human genome and genetic diversity	34
Detecting local adaptation in populations	37
Inferring human demography	40
Association studies	43
Enrichment analysis applied to genetic data	44
Psycho-physiological tests for taste perception.....	46
Research Aims	49
Results and discussion	50
Origins of modern humans (Paper I).....	50
Methods for detecting local adaptation (Paper II).....	53
Detecting local adaptation in Humans (Paper II, III and IV)	55
Taste perception and lifestyle (Paper IV).....	58
Conclusions and future prospects	61
Svensk sammanfattning	63
Résumé en Français	66
Origines de l’Homme Moderne.....	66
Peuplement du globe et structure de population.....	67
Etude de l’adaptation locale	68
Goût et mode de vie	69
Conclusions générales	71
Acknowledgments.....	72
References.....	78

Abbreviations

TRC:	Taste receptor cell
bp:	base pairs
kb:	thousand base pairs
Mb:	Million base pairs
DNA:	Deoxyribonucleic acid
TAS1R:	taste 1 receptor
TAS2R:	taste 2 receptor
GPCR:	G-Protein Coupled Receptor
ABC:	Approximate Bayesian Computation
SNP:	Single Nucleotide Polymorphism
TMRCA:	Time to Most Recent Common Ancestor
MHC:	Major Histoincompatibility Complex
kya:	thousand years ago
mya:	million years ago
GWAS:	Genome-Wide Association Study

Introduction

Origins of *Homo sapiens*

The common ancestors of our species *Homo sapiens* and all the extinct related species from genus *Homo* are believed to originate in Africa, where we find the most ancient remains of *Homo habilis* and *Homo ergaster* (between 2.8 and 1.7 mya). During more than two million years, the *Homo* genus evolved to a high diversity of morphologies: *Homo erectus*, the first that left Africa around 1.9 million years ago, *Homo neanderthalensis* and *Homo heidelbergensis*, *Homo floresiensis*, and *Homo sapiens*.

Two models for the apparition of *Homo sapiens* have been proposed, the “Multiregional model” and the “Out-of-Africa model”. According to the Multiregional model, the transition from *Homo erectus* to *Homo sapiens* appeared in several continents with different modern human characteristics arising in different regions of the world. The Out-of-Africa model on the other hand, states that the transition to modern traits happened in Africa and thereafter spread across the world and outcompeted all other *Homo* forms.

In favor of the Out-of-Africa model, the first remains of anatomical modernity – any trait close to what is found in humans nowadays – dating from 300kya to 150kya are found in Eastern and Southern Africa (Stringer 2002; White et al. 2003; Trinkaus 2005; Barham and Mitchell 2008; Tattersall 2009). Analyses of genetic data are also in favor of the Out-of-Africa model, revealing a loss of genetic diversity with the distance to Africa (Cann et al. 1987; Ingman et al. 2000; Jobling and Tyler-Smith 2003; Ramachandran et al. 2005), among other aspects.

First anatomically modern fossils coincide with a glacial period referred to as MIS6 (from 190 to 130kya) (White et al. 2003; McDougall et al. 2005). During this period, the ice sheet was extended (Hetherington and Reid 2010) and the unfrozen areas consisted to a large extent of deserts, savannas and open dry forests whereas rain forest receded in comparison to interglacial periods (Van Andel and Tzedakis 1996; Dupont et al. 2000). Open grasslands, found during MIS6, would not protect early humans against predators (Lahr and Foley 1998). Because of environmental factors, it is thought that this glacial period has led to a bottleneck in human populations

and could be the cause of speciation at the origin of the *Homo sapiens* species.

There are several pieces of genetic evidence in favor of a bottleneck at the origin of modern humans. First, the most recent common ancestor to all human mitochondrial lineages has been dated to the onset of MIS6 around 190kya, with an uncertainty interval of 30kya (Gonder et al. 2007; Behar et al. 2008). Also, the times to the most recent common ancestors in autosomes is eight times longer than in mitochondrial DNA, which is higher than expected at equilibrium and possibly reconciled with a model including a bottleneck during MIS6 (Blum and Jakobsson 2011). Finally, humans present 1.5 to 4 times lower genetic diversity than apes, suggesting a strong reduction of the population size at the origin of modern humans (Kaessmann et al. 2001).

There are two proposed scenarios for explaining the influence of the climate on human populations during the MIS6 (Lahr and Foley 1998). The fragmentation hypothesis suggests that, as a consequence of the severe aridity during MIS6 in Africa (Castañeda et al. 2009), potential refugia suitable for humans were fragmented and therefore human populations consisted in small groups geographically isolated making any gene flow impossible. After the glacial period, the areas suitable for living would be wider and these groups would have been able to get in contact again. On the contrary, the founder scenario postulates that only one group among all the hominins found in Africa would have been the ancestor of all modern humans. Under that model, the proposed refugia during MIS6 are East Africa where the earliest well-dated anatomically modern humans have been described (McDougall et al. 2005), the south coast of South Africa with its unusual confluence of plant diversity, coastline richness, and moderate climate (Marean et al. 2007; Marean 2010) or North Africa where a fossil of early *Homo sapiens* has been dated to 160kya (Smith et al. 2007).

Today, the generally accepted model for the origin of humans, consistent with most of the evidence from genetics, archaeology and paleontology, would be that modern humans originated in Africa at a period of arid climate before spreading to the rest of the globe almost 100kya afterwards.

Out of Africa

Outside of Africa, fossil record date anatomically modern humans from 40 to 60kya in Eurasia & Oceania (Trinkaus 2005; Mellars 2006) and up to 100kya in the Levant (Armitage et al. 2011; Rose et al. 2011) with fossils of anatomically but not necessarily behaviorally modern humans, as the earliest evidence of culture is dated to 70kya in Southern Africa (Henshilwood et al. 2002). In similar ways, genetic data also suggest an Out-of-Africa migration event dating from 100-60kya with a bottleneck in all non-African populations around that time (Laval et al. 2010) which lead to a loss of genetic diversity and increasing of linkage disequilibrium with the distance to Africa (Jakobsson et al. 2008). Two routes of migration out of Africa have been proposed, a southern route via the Bab-el-Mandeb strait and a northern route via the Sinai Peninsula into the Levant.

The Northern route (Figure 1A) is supported by fossils of anatomically modern humans in Israel dating between 90 and 130kya (Qafzeh & Skhul) (Stringer et al. 1989). Remains of *Homo sapiens* were later replaced by Neanderthal remains, and this migration of *Homo sapiens* out of Africa was therefore thought to have failed (Shea 2003; Grün et al. 2005). Nevertheless, it is possible that a later migration occurred via the northern route. Indeed, a recent study (Pagani et al. 2015), using genome data from Egyptians and Ethiopians shows that the Northern route is more coherent with genetic data, and maybe the northern route should not be discarded yet.

The Southern route (Figure 1B) is supported by the ancient remains found in Oceania dating from 50kya, older than remains found in Eurasia (Roberts et al. 1990). Some genetic evidence also support the Southern route such as the presence of very deeply rooted mitochondrial haplotypes in South Asia absent from populations from the Levant (Quintana-Murci et al. 1999; Macaulay et al. 2005), and an Aboriginal Australian genome (Rasmussen et al. 2011). In this model, modern humans migrated out of Africa around 61-65 kya to the Arabic peninsula. After the migration into the Arabic peninsula, several routes (coastal and more inland) are suggested to account for the fossils observed in Australia (50kya), Eastern China (40kya), Lake Baikal (39kya), the Middle East (49kya) and Europe (46-41kya).

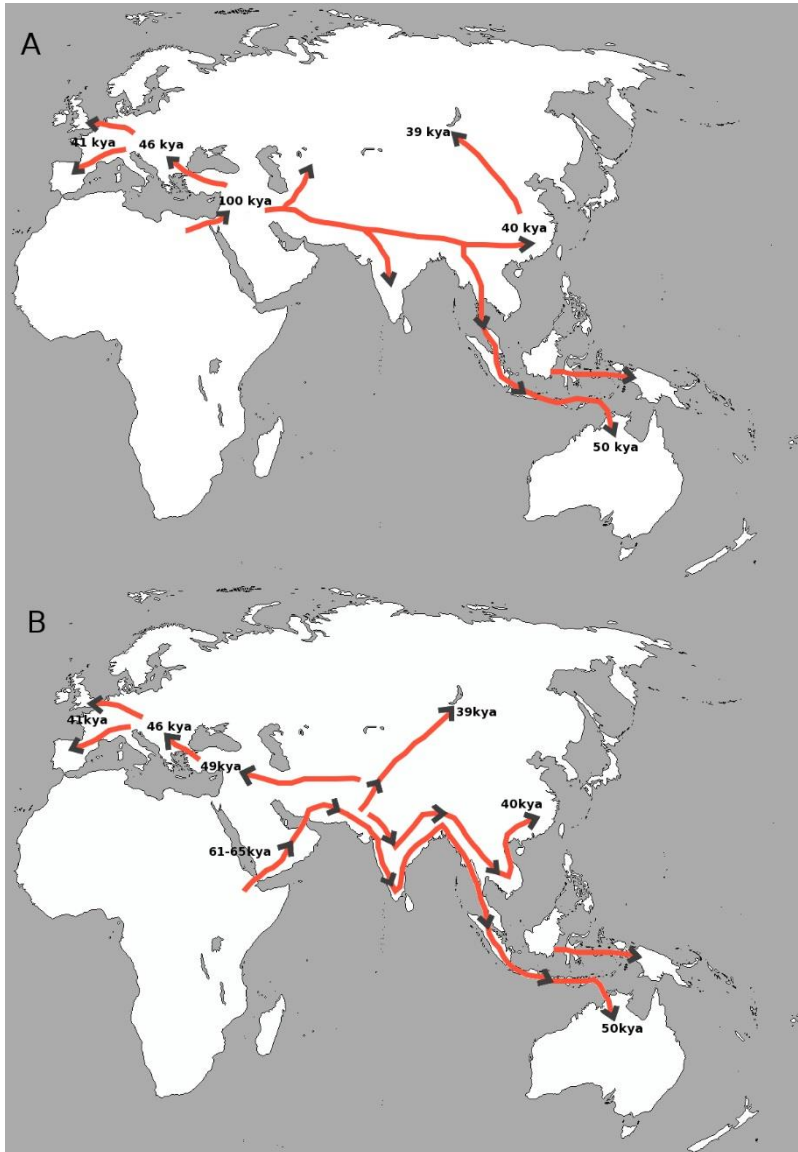


Figure 1: Maps of human migration routes out of Africa to the old world and Australia. A: Northern route model B: Southern route model. The dates are the estimated ages of the fossils found in the various regions and have been reported by in Jobling et al. (2013).

Shared patterns of genetic variation in non-Africans in contrast to the high divergence of genetic variation within Africans argue against migration Out-of-Africa from multiple sampling of different source populations (Hellenthal et al. 2008; Behar et al. 2012).

Population structure in Africa

It is thought that our species *Homo sapiens* have emerged in Africa with the first human remains presenting modern anatomy dating up to 190kya while they are not found before 100kya in the Levant. This means that anatomically modern humans have lived almost twice as long time in Africa as outside of Africa. Genetic data show that the diversity is highest in Africa (Cann et al. 1987; Behar et al. 2008; Jakobsson et al. 2008; Li et al. 2008; Lachance et al. 2012), but also indicate that ancestral human populations in Africa were structured before any anatomically modern human population migrated out of Africa (Shimada et al. 2007; Yotova et al. 2007; Behar et al. 2008; Schlebusch et al. 2012).

Today, Africa is extremely diverse in terms of culture and linguistics with more than 2000 ethnolinguistic groups identified (ethnologue.com). African languages can be divided in four major language families, Niger-Kordofanian, Afroasiatic, Nilo-Saharan, and Khoisan (see figure 2 for precise geographic main repartition). Niger-Kordofanian languages are broadly spread from western Africa up to South-East and down to South Africa. The Niger-Kordofanian language family contains among others the Bantu languages. Non-Bantu Niger-Kordofanian languages are spoken in West and Central Africa, while Bantu languages are spread everywhere south of the equatorial latitudes. Afroasiatic languages are spoken in the Sahara desert, North Africa and East Africa up to the horn of Africa (Somalia). As the name suggests, Afroasiatic languages are also spoken outside of Africa in the Arabic peninsula and Near East. Nilo-Saharan languages are spoken in Central and East Africa, in the upper parts of the Chari and Nile rivers including Nubia. Last, the Khoisan language family contains all the languages that use click-consonants. They are spoken in the Kalahari desert by Khoe-San people as well as Hadza and Sandawe from East Africa, although the linguistic relatedness is debated. Khoisan languages have been previously suggested to be modern representatives of languages spoken by early anatomically modern humans (Blench 2006; Mitchell 2010), although language similarities and relationships are thought to only remain for maximum 10,000 years (Güldemann and Stoneking 2008).

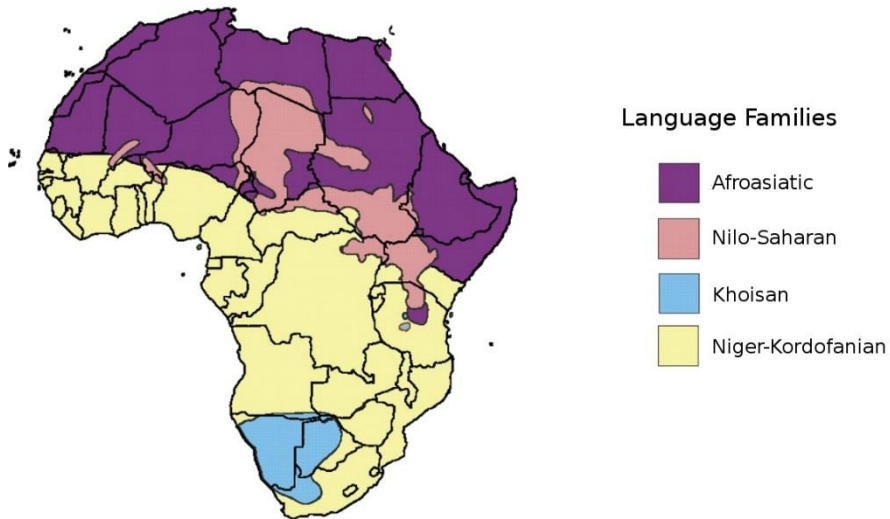


Figure 2: Map of geographic repartition of language macro families in Africa (reproduced from Scheinfeldt et al. 2010)

In Africa, linguistics as well as culture and lifestyle are partly correlated with geography and genetic structure (Tishkoff et al. 2009; Schlebusch et al. 2012). If we consider genetic relationships between populations as a tree (figure 3), the deepest split is between Khoe-San and the rest of African groups and dates to more than 100kya. Northern and Southern Khoe-San groups diverge from each other around 35kya. This is consistent with previous observations that Khoe-San people present high levels of diversity both within and between groups (Vigilant et al. 1991; Jorde et al. 2000; Tishkoff et al. 2009). Southern Khoe-San (Nama, Karretjie and ≠Khomani, see figure 3) are characterized by hunter-gatherer and pastoralist lifestyles, probably adopted recently, while Northern Khoe-San groups (!Xun and Ju/'hoansi, figure 3) are hunter-gatherers that supposedly kept the original lifestyle of proto Khoe-San groups (Barnard 1992).

The Central Africans represented by the pygmy hunter-gatherers from West and East Africa diverge around 60kya (Patin et al. 2009) from other West and East African groups. The divergence between Eastern and Western pygmies has been estimated around 20kya (Patin et al. 2009; Verdu et al. 2009). This Central African branch does not correspond to any language unity as Pygmy populations probably adopted languages from the neighboring populations with whom they interact. Nevertheless, some similarities in vocabulary referring to specific words related to the lifestyles suggest that there was a common proto-pygmy language (Letouzey 1976; Bahuchet 1993; Bahuchet 2012).

The next divergence time between groups is between East Africans and West Africans with a timing around 35kya. Within East African groups, the divergence between Maasai pastoralists on one hand and Hadza and Sandawe hunter-gatherers on the other hand is about 10 kya. Hadza present a hunter-gatherer lifestyle, while Sandawe shifted to a farmer lifestyle under colonial period. Hadza and Sandawe also both speak a click language (Sands 1998). The East African branch is not homogeneous in terms of languages but groups pastoralists and hunter-gatherers together.

The last branch referred to as the western African branch correlates with Niger-Kordofanian languages. The relatedness between West Africans is largely influenced by the Bantu expansion which refers to the large demographic and migratory expansion dating around 5kya from western Central Africa (current Nigeria and Cameroon) to eastern and southern Africa coincident with the emergence of agriculture.

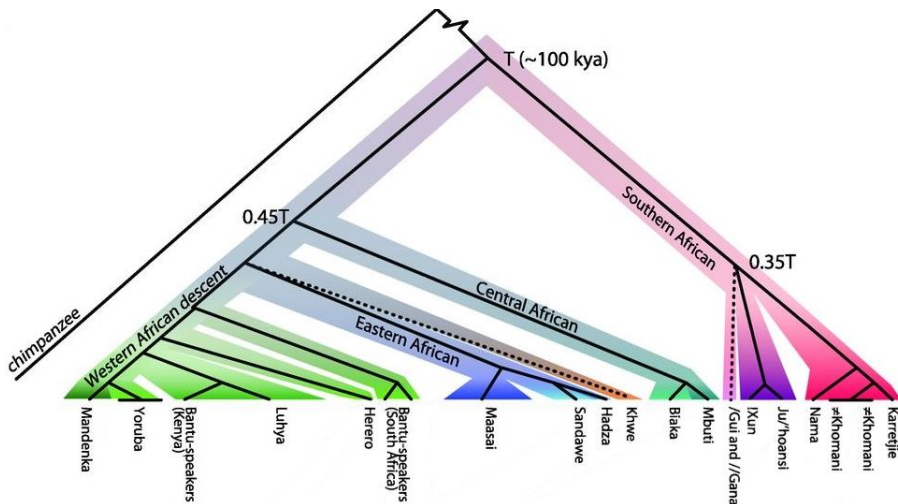


Figure 3: Tree representation of population structure observed in Africa from genotype data (adapted from Schlebusch et al. 2012). (We don't consider the branching of Khwe and /Gui and //Gana in the text as it is uncertain)

As shown highly simplified here, African groups present deep divergences and very high genetic diversity probably due to the long history of modern humans in Africa. Also, African groups live in very diverse environments and present very different lifestyles among which are hunting-gathering, farming and pastoralism. The study of adaptation patterns in African groups is therefore important as the variety of environments and lifestyles probably caused different selective constraints.

Peopling of Central Asia

Due to its location, Central Asia must have played an important role during the peopling of Asia and Europe by anatomically modern humans. Nevertheless, its peopling is poorly known (Nei and Roychoudhury 1993; Comas et al. 1998; Karafet et al. 2001; Wells et al. 2001; Cordaux et al. 2004; Macaulay et al. 2005). Three models have been proposed: Central Asia could be a source region from which populations have migrated to Europe and East Asia, a well population – the late meeting region of populations from Europe and East Asia – or finally the result of an east-to-west migration from East Asia to Europe (Chaix et al. 2008). Statistical analysis of genetic data (microsatellites and sequences of autosomal regions) suggest several waves of migrations in Central Asia. An ancient migration during the Paleolithic from East to West across Eurasia. A more recent both from West and East to Central Asia around 8.5kya and later a migration from the East 2-3kya (Palstra et al. 2015).

The first historical records of Central Asia describe the Aryan invasions of the actual Turkmenistan toward Iran around 3.3kya, and the building of the Persian Empire around 2.7kya. These nomadic peoples called the Scythians were speaking Indo-Iranian languages. Later, many expansion waves came from the East around 2.2kya with the Huns. The Silk Road, which started around 2.2kya, linking East Asia to the Mediterranean basin, contributed to Central Asia being a region of great exchange and movement of people.

From 1.8-1.7kya, numerous Turkic expansion waves provoked linguistic changes in Central Asia. This event put the milestone to constant invasions from empires from West or East such as muslim invasions from Persia (between 1.4 and 1.1kya) and the Mongol Empire of Genghis Khan spreading from China to Europe (around 900ya).

All these waves of migrations have caused a high ethnic, linguistic and lifestyle diversity in Central Asia today. People speaking Indo-Iranian languages, such as Tajiks, are supposed to descend from the first Iranian people described in 3.3kya but their ancestry may even trace back to the Neolithic. Groups speaking Turkic languages, such as the Kazaks, probably settled in the region after the eastern invasions starting 1.7kya.

Lifestyle transitions in the human history

From the first *Homo* groups to present-day humans, diet has radically changed. Changes in diet have probably led to adaptation among hominins. Three main events might have strongly influenced the diet: the transition to meat and animal products consumption, the use of fire and the emergence of agriculture.

The consensus about the diet of hominins predating the *Homo* genus is that they had a diet based on fruits and leaves. Robust *Australopithecus* groups might have been vegetarian while the more gracile ones would include some meat in their diet (Kay 1985). Nevertheless, isotope measures suggest that some of the robust groups might also have eaten animals in small amounts as it is thought for *Australopithecus africanus* (Sponheimer and Lee-Thorp 1999).

The “Food for thought” theory first suggested by Aiello and Wheeler (1995) assumes that the brain is a costly organ and therefore the diet of *Homo* groups, characterized by an increase in brain volume, must have been omnivorous with a certain extent of meat since vegetable foods are costly to digest and detoxify. Fossil records suggest that one of the first *Homo* individuals from Oldowan in Ethiopia were either scavengers or hunters. Remains of accumulated animal bones (but not whole skeletons) showing cut marks of stone tools are found in the site but whether they were killed by hominins or by other predators is uncertain. Some almost complete animal skeletons suggest that hominins, if not hunters, were efficient foragers that managed to get primary access to the animal carcasses (Bunn et al. 1986; Blumenschine et al. 1987; Domínguez-Rodrigo 2002; Domínguez-Rodrigo and Pickering 2003).

Anatomically modern human developed better tools for hunting and therefore would have a higher consumption of meat than earlier *Homo* groups although some evidence of scavenging activity remain as well (Binford 1984). Their diet was omnivorous from a broad spectrum economy from hunting (Mitchell 2002; Klein 2009), fishing (Mitchell 2002) and gathering tubers.

The ability to use fire has deeply modified human diet and metabolism by facilitating digestion of foods and helping in detoxification (e. g. parasites and toxins). The first widely accepted evidence of the use of fire dates to 800kya in Israel (Goren-Inbar et al. 2004), 670 to 400kya in China (Weiner et al. 1998) and 600 to 400kya in Europe (Roebroeks and van Kolfschoten 1994; Gowlett 2006; Valoch 2013). The older evidence for use of fire are few because the sites observed are mostly open air sites. A greater

abundance of cave sites starting from around 130kya with evidence of fire usage confirm that, from this period on, all *Homo* groups could make fire when needed. Nutrition of *Homo* groups might have changed together with the control of fire starting around 450kya or even earlier, although the frequency at which fire was used is difficult for assess before 130kya.

The emergence of agriculture and domestication, consisting in the cultivation of soil, production of crops and raising of livestock, modified the diet and demography of many human populations. Agriculture, frequently associated with a sedentary lifestyle, appears late in the archeological record but was not developed instantaneously. For example, it has been found that hunter-gatherer groups managed the plants around their settlements. The environmental changes played a crucial role (Allaby et al. 2008) for the emergence of agriculture, which coincided with an increase and stabilization in temperature starting around 14kya.

Humans developed agriculture independently in different regions of the world at various times (Fuller et al. 2011). The first signs of agriculture are from the Fertile Crescent about 10kya, followed by Mesoamerica and the lowland South America around 9-8kya, east China (8-7kya), about 5kya for eastern woodlands of North America, Sahel/West Africa and Ethiopia, and finally spanning from 7 to 4kya in New Guinea (Denham et al. 2003; Jobling et al. 2013) (see figure 4 for locations and species domesticated).

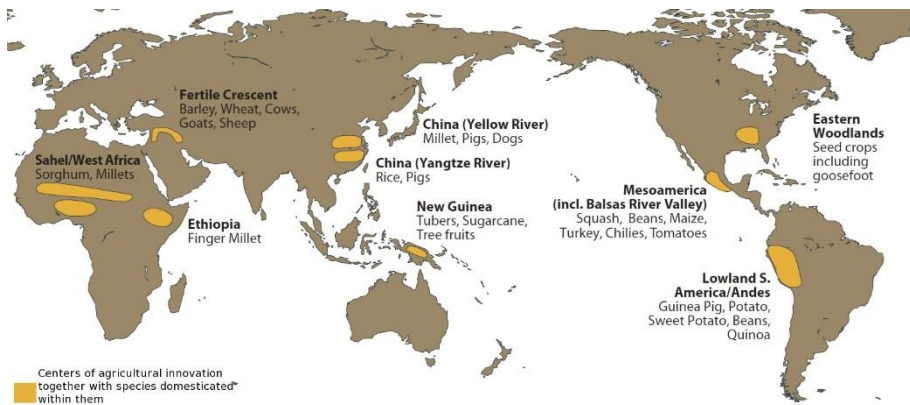


Figure 4: Map of the centers of agricultural innovation, some species domesticated in each region are given as well (adapted from Jobling et al. 2013)

The most important domesticated species are the widely spread and easy to find cereals: Wheat and barley in Near East, millet and rice in China, maize in the Americas, sorghum in Africa and sugarcane in New Guinea. What constraints the locations of origins of domestication, was the domestication of crops and animals that had restricted areas of repartition.

Agriculture had a dramatic impact on the diet of the populations that adopted it. The plant and animal resources that are used by agriculturalists are much less diverse than the resources exploited by hunter-gatherer societies (Harris and Hillman 2014). Although agriculture fed more people than hunting and gathering for an equal area, malnutrition, anemia and lack in nutrients and vitamins were initially frequent (Pinhasi and Stock 2011).

Agriculture not only modified diet and lifestyle but also population growth, which led to higher disease exposure (zoonosis due to close relationship to animals, infections and parasites from increased human and animal wastes, such as plague and cholera) and fast spreading (Armelagos and Cohen 1984).

Pastoralism is another globally spread lifestyle in the world. It is a strategy to raise livestock on pastures and by doing that, to benefit also from less productive land where agriculture is not suitable. It is usually a mobile lifestyle that adapts to seasons.

The origin of pastoralism is unclear. One hypothesis is that pastoralism derived from mixed farming: the increased productivity of irrigated agriculture led to less access to water and land for the livestock. Bordering areas of land remained in use for animal breeding and large distances had to be covered by herds to collect sufficient forage. Specialization occurred as a result of the increasing importance of both intensive agriculture and pastoralism. According to this hypothesis, both agriculture and pastoralism developed together, with constant interactions (Lees and Bates 1974). Another theory is that pastoralism evolved from hunting and gathering. Following herds for hunting indeed provided a good knowledge of wild goats and sheep mobility and needs. According to this theory, undomesticated animals were chosen to become more controllable for the proto-pastoralist nomadic hunter-gatherer groups by taming and domesticating them (Marshall 1990). In this thesis, I will present the case of Central Asia.

Three hypotheses have been proposed for the emergence of pastoralism in Central Asia. Two suggest that pastoralism come from agropastoralist cultures, either from the Djeitun culture found around 8kya (Renfrew 1996), or later 6.6-4.8kya from the Kourgan culture of the Volga region (Ukraine), which spread from the Altai mountains to Bulgaria (Anthony and Brown 1991; Gimbutas and Gimbutas 1991). The third hypothesis, is in favor of a direct transition from a hunting-gathering society into pastoralism by a cultural transition of hunter-gatherers to a Kourgan-related culture. This is supported by the observation in this region of hunter-gathering cultures (Khlobystina 1973; Shilov 1975) contemporary to exclusively pastoralist

cultures (Levine 1999; Jacobson-Tepfer 2008) without sign of agriculture. This suggests that the hunter-gathering cultures adopted pastoralism from already specialized pastoralist groups.

All of these lifestyles supposedly inherited by these transitions during human evolution, still exist among contemporary human populations and probably shaped their adaptive history.

Adaptation in modern humans

Since their appearance in Africa, anatomically modern humans have peopled almost all terrestrial environments on Earth from the richest to the most extremes: from tropical to polar, rain forests to tundra and deserts and up to high altitudes. Each environment presents both resources and constraints to which human populations have adapted.

Human populations present phenotypic variations from the most conspicuous to invisible ones. It is now generally understood that phenotypic variation in humans often have a genetic basis although it is often complex and interacting with the environment, some of the phenotypic variation is adaptive. The observation of body shapes from closely related species or subpopulations of a species from different climates show that animals from cold environments tend to have larger and rounder bodies (Bergmann's rule) and shorter limbs (Allen's rule) to improve thermal efficiency. This tendency can be observed in human populations as well with Inuits from cold environments having a rather short stature and stocky body while the Maasai or other groups from hot and dry environments have a long stature and long limbs (Ruff 1994).

Skin color is a highly variable trait in humans and correlates with the solar exposure of human populations: darker skin is found in tropical latitudes while lighter skin appeared partly independently in northern latitudes of East Asia and Europe. Several genes have been found to be affecting skin, hair and eye pigmentation and have also been the targets of adaptation (Lamason et al. 2005; Voight et al. 2006; Miller et al. 2007; Norton et al. 2007; Williamson et al. 2007; Coop et al. 2009; Pickrell et al. 2009). A potential explanation for selection on skin color is the protection against UV in lower latitudes, as well as a better sensibility to UV for better assimilation of vitamin D in higher latitudes.

A striking example of convergent adaptation to the environment is the case of adaptation to high altitude populations living in the Andes, Himalayas and East African plateau that possess alleles preventing against complication

during pregnancy in newborns as well as providing better oxygen fixation in blood in hypoxic conditions (Moore 2001; Beall 2007; Brutsaert 2007; Alkorta-Aranburu et al. 2012).

Many examples of local adaptation in humans are explained by adaptation to diseases and diet (Barreiro et al. 2008). A famous example is the selection on alleles that are protective against malaria in regions where malaria is endemic (Hamblin and Di Rienzo 2000; Tishkoff et al. 2001).

Depending on their environment and their lifestyles, human populations have various energy expenditures. Climatic conditions, such as temperature and humidity, as well as the mobility of the lifestyle (nomadic, hunter-gatherer, sedentary farmer) influence the energy spent by the body. Different environments do not present the same resources: the biodiversities vary greatly with the distance to the equator as well as variation along a year-cycle. These factors constitute constraints for humans both for climatic and food reasons. A simple thing as the drinking water can be at the basis of local adaptation, a recent example for that is the selection on haplotypes of a gene providing better excretion of arsenic in populations of the Andes living in regions where the drinking water is rich in arsenic (Schlebusch et al. 2015). Another famous and well-studied example is the lactase persistence developing independently in populations consuming dairy products in Europe, Middle East and East Africa (Enattah et al. 2002; Tishkoff et al. 2007; Schlebusch et al. 2012). Lifestyle and diet might also have led to selection on protective alleles against diabetes both in pastoralist and farmer populations of Central Asia (Ségurel et al. 2013), and better meat digestion in pastoralists (Ségurel et al. 2010). As an effect of agriculture and high starch consumption, there is evidence of selection on copy number variation of the amylase gene, facilitating the digestion of starch, in populations with high starch diet (Perry et al. 2007).

Although there are spectacular examples of phenotypic variations caused by local adaptation in humans for which the genetic basis is known, the large majority of human phenotypic variation is still little understood. This is probably due to the complexity of biological traits, making it unlikely that selection could happen on only one or few genes. Instead, it is more likely that selection acts on several genes and on already existing genetic variation in populations, making it difficult to detect (Pritchard et al. 2010; Pritchard and Di Rienzo 2010). New methods for detecting selection or enrichment analysis of Gene-Ontologies may provide insights of adaptation in humans.

Taste and diet

In its common meaning, taste is the general sensation provided in the mouth by food. This broad definition stands for a complex combination of olfactory, somatosensory and gustatory perceptions. What strictly defines taste is all the sensations generated by the stimulation of the gustatory system by water-soluble compounds dissolved in the saliva when eating. There are five major tastes: sweet, bitter, salty, sour and umami (savory). Taste perception informs on the quality and nutritious content of potential food and is therefore essential in food choices, nutrition and survival.

Each basic taste has its importance in the diet. Sweetness, produced mostly by carbohydrates, tells us about the potential energetic aspect, it generates an innate pleasure as shown by studies of facial expression of newborns stimulated by sweet solutions (Steiner 1974). Bitterness, often generated by toxins such as alkaloids, warns against potential poisons, but also about potential medicine. Umami (Japanese word meaning “savory”) is produced by glutamate and some nucleic acids and therefore suggests high energy value (proteins) as in meat and cheese (Breslin and Spector 2008). Salty taste is produced by alkali ions, important for terrestrial animals that lose minerals through sweating and excretion. Salty taste is therefore important in regulating the mineral balance of the body (Lindemann 1996). Strong sourness is aversive and potentially revealing either unripe fruits or spoiled food, whereas mild sourness is perceived as interesting and many beverages are sour (Lindemann 1996). As an example, vitamin C, present in many fruits and essential for humans, tastes sour (Schiffman and Dackis 1975). Taste is therefore considered as a way to prevent from eating harmful food as well as a manner to look for beneficial food fulfilling the organism’s needs.

Several studies on taste receptors in various species show a coevolution of taste perception with diet. The global tendency is that, regarding the diet (frugivore, insectivore, nectarivore, carnivore, etc.), species tend to lose the tastes that are not present in their diet. We can list some interesting examples. Both domestic and wild cats as well as lions, all carnivorous, lost the ability to taste sugar after a pseudogenization of their sweet receptor. Species from the Carnivora order but presenting another diet than exclusively carnivore, have the ability to taste some sugars (Li et al. 2009). On the contrary, a high sugar diet can produce a recovery of sugar perception that has been lost in a taxa, this is the case of some nectarivore hummingbirds that perceive sweet via the originally umami receptor that has been repurposed in these species to a carbohydrate receptor (Baldwin et al. 2014). The number of bitter taste receptors in genomes also vary tremendously among vertebrates and coincide with diet of species (Dong et

al. 2009). Many other interesting examples have been subjects of studies (bats, whales etc.), and made available due to genetic sequence data.

At one stage of human evolution, the use of fire potentially caused a relaxation on bitter and sour perception because cooking helps removing noxious and poisonous substances. But a study on bitter taste receptors suggests that these receptors evolved similarly in humans as in great apes with little selective constraint (Fischer et al. 2005). Since then, humans peopled almost all terrestrial environments and adopted various lifestyles (hunting, gathering, herding, agriculture etc.) in close relationship with the characteristics of the environment they lived in. Depending on the environment in which one lives, the food resources vary as well as the potential dangers. Therefore, tastes could have evolved differentially between populations because of natural selection. Moreover, within a same environment, depending on their lifestyle (nomad, sedentary) populations present different energetic needs and explore their environment differentially.

Anthropological studies comparing populations of diverse lifestyles and environments have reported variation in taste perception across human populations. Gieli and Aka Pygmies from forest environment are less sensitive to sugars than Mvae and Yassa populations from the savanna. Fruits in a forest environment are more abundant and sweeter than in the savanna where they can even be poisonous. In savanna conditions, being more sensitive to sweet would provide a selective advantage because of a better aptitude in exploring the environment in search for highly energetic foods. Inuits from Greenland, have a high fat and high protein diet from hunting sea mammals and therefore have a high concentration of urea in the blood. To eliminate the urea, they drink high amounts of water extracted from freshwater ice and this for a great deal of the year. The choice of the ice has to be careful in order not to take in salt from the sea water. Consumption of slightly salty water in such quantities could be harmful. Interestingly, Inuits are more sensitive to salt than any other populations studied supposedly because of selection toward a greater sensitivity to salt. So depending on the environment human populations have evolved in, their tastes seem to have adapted probably at a genetic level (Hladik and Simmen 1993).

Mechanisms of taste perception in humans

Among the five senses humans can experience – vision, olfaction, hearing, taste and touching –, taste is essential for nutrition and survival.

Because of the connection between the oral and the nasal cavities, smell is also stimulated when eating. The combination in the brain of olfactory and gustatory information generate the flavor of the food (Rolls and Baylis 1994; Small and Prescott 2005). For this reason, during a cold, flavor is considerably diminished, while tastes are still perceived. In addition, the somatosensory system – detecting sensations of temperature, pain and texture – is also involved when eating mint or chili for example. Although generally included in the broader definition of taste, these sensations are perceived by different systems. In this thesis, I will use the strict definition of taste which involves the gustatory system for the perception of sweet, bitter, umami, salty and sour.

Taste sensation is mediated by taste receptor cells (TRCs) located in taste buds, themselves organized in papillae (see figure 5). Taste buds are embedded in the stratified tongue epithelium and are open to the oral cavity via a small opening in the tongue epithelium, called taste pore. Taste buds are pseudostratified and consist in several cell types: mature polarized elongated cells of type I, II & III as well as basal cells (Roper 2006). Basal cells are nonpolarized, presumably undifferentiated, whose function is supposed to be the regulation of the turnover of taste receptor cells.

Type I cells are presumed to have a role in terminating synaptic transmission and restricting the spread of transmitters and may also transduce salt taste perception (Vandenbeuch et al. 2008). Type II cells can perceive sweet, umami and bitter, each cell expresses only receptors dedicated to one taste. Type III cells, or presynaptic cells, are characterized by clear synaptic contacts with gustatory nerve fibers and therefore are thought to be responsible for information transmission to the nervous system. Type III cells also respond to sour taste stimuli and could be the receptor cells for sour taste (Chandrashekar et al. 2006; Huang et al. 2006). It is currently accepted that each cell type is tuned to detect one taste each.

Four types of papillae can be distinguished by their shape and structure: fungiform, foliate, circumvallate and filiform (Figure 5). The latter are found all over the tongue but do not contain any taste bud and therefore are not thought to be directly involved in taste perception but to have a mechanical function. The three other type of papillae do contain taste buds: fungiform papillae are located in the anterior two-third of the dorsal surface of the tongue and contain one or two taste buds, foliate papillae are on the lateral

edge of the tongue and consist in up to hundreds of taste buds, last, circumvallate papillae are found at the extreme back of the tongue and contain up to thousands of taste buds. Several nerves innervate the taste papillae. Fungiform papillae and anterior foliate papillae are innervated by the facial nerve, posterior foliate papillae and circumvallate papillae are innervated by the glossopharyngeal nerve.

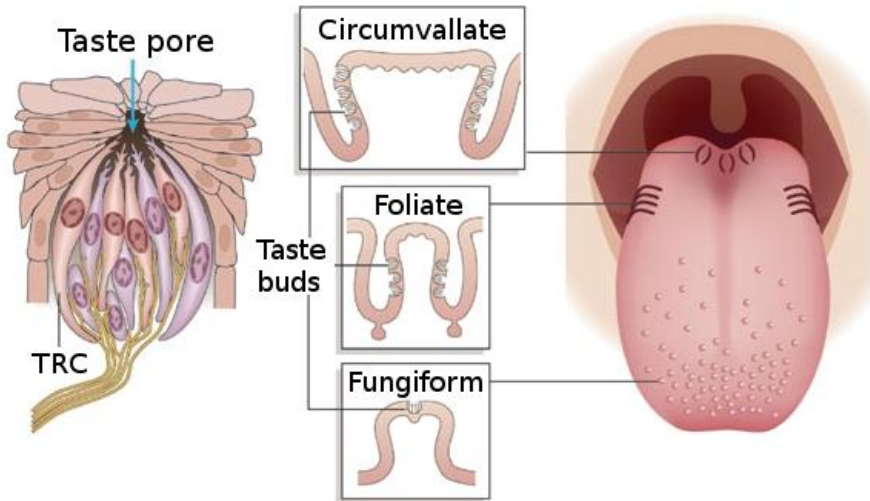


Figure 5: Location and structure of taste receptor cells, taste buds and taste papillae in the oral cavity (taken from Chandrashekar et al. 2006).

Each basic taste is mediated by a specific mechanism of signal transduction initiated by specialized taste receptors. Perception of bitter, sweet and umami tastes, the best understood, are initiated by G-protein coupled receptors (GPCR) from TAS1R and TAS2R gene families (taste 1 receptor and taste 2 receptor families). Signal transduction for bitter, sweet and umami involve gustducin, a heterodimer of G-proteins, which induces calcium release permitting the release of neurotransmitter.

Humans have 25 genes coding for TAS2R bitter receptors (Adler et al. 2000; Chandrashekar et al. 2000), located on chromosome 5, 7 and 12. The complexity of bitterness perception is permitted by the different responsiveness of each receptors to bitter compounds. Some bitter compounds are detected by a large number of TAS2Rs while some TAS2Rs can be specific to one of few chemical compounds (Meyerhof et al. 2010). Sweet and umami taste are detected by only three TAS1R receptors functioning as heterodimers: TAS1R2 and TAS1R3 bind to create sweet receptors while TAS1R1 and TAS1R3 create umami receptors (Nelson et al. 2001; Zhao et al. 2003). Humans can distinguish many sugars and umami-

tasting compounds although only three genes code for the receptors of these tastes, the mechanism for the distinction between these compounds is still unknown.

Salty and sour tastes are less understood but the best candidate for salt perception are four genes forming an amilorid-sensitive ion channel with four subunits: alpha, beta, delta and gamma (SCNN1)(Heck et al. 1984; Avenet and Lindemann 1988). The best evidence so far for sour receptor is for the PKD2L1 gene (Huang et al. 2006). Sour and salt receptors directly induce cell-depolarization and calcium release (Meyerhof 2008). Several studies show the effect of other genes on taste perception, such as the effect of RTP/REEP genes in bitter (Behrens et al. 2006) and sweet (Ilegems et al. 2010) perception. Hormones also can have an effect on taste perception, such as angiotensin on sweet and salty tastes (Shigemura et al. 2013) and leptin on sweet (Yoshida et al. 2013).

Background and methods

This thesis focuses both on demographic inference of human history and adaptation in humans. The knowledge and methods necessary to understand the results of the papers are introduced in this section.

Genetic variation

DNA is the molecular basis of genetic information and is built up as a sequence consisting of four nucleotides. The nucleotides are themselves composed of a phosphate, a deoxyribose sugar and a nitrogenous base. There are two types of nitrogenous bases, the purines: Adenine (A) and Guanine (G) and the pyrimidines: Thymine (T) and Cytosine (C).

At each generation, parental DNA is replicated and eventually transmitted to the offspring. However, several types of mutations can occur during the replication of the DNA. In this thesis I will only consider one type, single nucleotide polymorphisms. Single nucleotide polymorphisms occur when a nucleotide in the original DNA sequence is replaced by another nucleotide in the copy. There are two types of single nucleotide polymorphisms: transitions which are mutations of a purine to a purine or a pyrimidine to a pyrimidine, and transversions which are mutations of a purine to a pyrimidine or a pyrimidine to a purine. The two variants of a locus after a mutation has occurred are called alleles.

After a mutation has occurred, and if the new allele is transmitted to following generations, the mutated locus will be variable. The term polymorphism is used when at least two alleles are present in a population and it is then called a Single Nucleotide Polymorphism (SNP). The most recent allele of the two, is called the derived allele, while the original is the ancestral allele.

One of the best understood and studied regions of the DNA are the coding genes: the sequences that are transcribed and translated into proteins, the main actors within organisms. In order to translate DNA into proteins, DNA is transcribed into RNA which is then read in groups of three nucleotides at a time, the codons. Each codon is translated into a corresponding amino-acid

determined by the genetic code (table 1). The new amino-acid is joined to the previous ones to form the protein. The genetic code also gives the start and stop signal for translation. Several codons can produce the same amino-acid. When a mutation happens in a genic sequence, its position in a codon plays a key role. A mutation is synonymous if the two alleles code for the same amino-acid (CTT and CTG for example, see table 1). In such cases, the function of the gene is most often unaffected. In some cases, a mutation can have a dramatic effect if it results in premature stop-codons (substitution from TCA to TAA, see table 1) or protein-damaging missense (substitution from ATC to ACC, see table 1) alleles.

Table 1: Table of the common nuclear genetic code. The colors refer to the chemical properties of the amino-acids yellow: nonpolar, green: polar, blue: basic, rose: acidic, and grey: stop codon.

1st base	2nd base								3rd base
	T		C		A		G		
T	TTT	Phenylalanine	TCT	Serine	TAT	Tyrosine	TGT	Cysteine	T
	TTT		TCC		TAC		TGC		C
	TTA		TCA		TAA		TGA		A
	TTG		TCG		TAG		TGG		G
C	CTT	Leucine	CCT	Proline	CAT	Histidine	CGT	Arginine	T
	CTC		CCC		CAC		CGC		C
	CTA		CCA		CAA		CGA		A
	CTG		CCG		CAG		CGG		G
A	ATT	Isoleucine	ACT	Threonine	AAT	Asparagine	AGT	Serine	T
	ATC		ACC		AAC		AGC		C
	ATA		ACA		AAA		AGA		A
	ATG	Methionine	ACG	AAG	Lysine	AGG	Arginine	G	
G	GTT	Valine	GCT	Alanine	GAT	Aspartic acid	GGT	Glycine	T
	GTC		GCC		GAC		GGC		C
	GTA		GCA		GAA	GGA	A		
	GTG		GCG		GAG	GGG	G		

The rate at which mutations happen depends on many factors and varies across the genome as well as between species. In humans for example the genome averaged mutation rate has been estimated to be between 10^{-8} and $2.5 \cdot 10^{-8}$ /bp/generation (Nachman and Crowell 2000; Roach et al. 2010; Scally and Durbin 2012). This is in fact a very small rate, therefore the probability that two independent mutations happen at a same site is low. For this reason, we can approximate the mutation process with the infinite-site model (Nei and Kumar 2000): if two variants are observed at a site, it is assumed to be the result of only one mutation event. If two chromosomes present the same allele at a locus, they are “identical by state” on this locus. Because of the infinite-site assumption, identical by state is considered to be “identical by descent” for the derived allele. In other words, two

chromosomes sharing a derived allele at a locus inherited it from a common ancestor.

In species with sexual reproduction, diploid individuals produce haploid gametes, and each copy of the chromosomes are separated among the gametes. During this process, recombination between the two copies of chromosomes occurs and shuffles the chromosomes. At each generation, recombination generates chromosomes that are mosaics of the genetic material from the previous generations. Recombination increases diversity by generating new combinations of genetic material as well as, over time, breaking down the dependence of sites located on the same chromosome.

Mutation and recombination are the motors of genetic diversity in natural populations. The genetic diversity of a population is thought to be directly associated with its ability to adapt to environmental constraints. By this logic, the more diverse a population is, the more likely it will harbor alleles that can improve the fitness of their carriers in a given environment (Hartl et al. 1997).

Evolutionary forces and equilibrium models

The term population refers to a group of individuals who have a high probability to reproduce together. On operational term, this means a group of individuals that share a certain criteria such as geography, culture, lifestyle, phenotype, ancestry. In humans, it can also be determined by self-identification. For this reason, populations are not immutable and usually interconnected with each other. Biological differences between populations often reveal what is called “population structure”, which refers to a global genetic differentiation between populations due to non-random mating of individuals.

I previously presented mutation and recombination processes as the basis of genetic diversity, but these are not the only motors of evolution. The survival of gametes is a stochastic process, and some alleles might not be transmitted to the next generation, producing genetic drift. Also, populations are not isolated from each other, and individuals migrate and exchange genetic material across populations potentially introducing new genetic diversity. In addition to stochasticity and gene flow, some alleles may increase their carriers' probability to reproduce or survive and therefore increase their chance to transmit their genetic material to the next generation. All these evolutionary forces play a role in how populations change through time, and make the study of populations' allelic diversity complex.

In a sexually reproducing diploid species, the autosomal chromosomes are combined in pairs (one maternal and one paternal). At each locus, the maternal and paternal alleles are jointly referred to as a genotype.

Given the allele frequencies, the Hardy-Weinberg equilibrium gives the frequency of genotypes in an ideal diploid population with sexual reproduction. Such a population fulfills the following properties: no mutations and no selection occur, all individuals mate randomly, the generations do not overlap, every individuals have offspring in the same amount and the population size is infinitely large. Under these conditions, allele frequencies remain the same over generations. Each alleles of a locus will be distributed in the population among homozygous individuals (two copies of the same allele) and heterozygous individuals (one copy of each allele). For each variable locus presenting alleles “A” and “a” with respective frequency p and q , the observed frequencies or the genotypes are: AA: p^2 , aa: q^2 , aA: $2pq$. The heterozygosity of a population, gives an idea of its diversity and represents the expected fraction of heterozygous individuals in the population.

As I presented at the beginning of this section, no natural population fulfills the Hardy-Weinberg model, but it permits to measure how far from the model populations are, and in that way we can evaluate the effect of evolutionary forces.

The human genome and genetic diversity

The human nuclear genome is distributed among 22 pairs of autosomes, common to all humans and one pair of sex chromosomes, women carrying two copies of the X chromosome and men one copy of the X and one copy of the Y chromosome. Each nucleus contains one copy of the nuclear genome. The mitochondrial genome (~16.5kb) is found in mitochondria which exist in many copies in each cell. The Y chromosome is transmitted from father to sons while mitochondria are found in everyone but only transmitted from the mother to children.

In 2001, the International Human Genome Consortium released the major parts of the human genome (Lander et al. 2001). This has been a milestone for understanding human genetics, gene functions as well as human history. This reference human genome has been obtained from several individuals and its assembly is constantly improved. The haploid genome size is approximately 3.2 gigabases. The human genome differs at some loci among individuals and it is estimated that on average there is a SNP every 1250 base pair.

The structure of the human genome has been extensively studied but is still not completely understood. We can distinguish two types of sequences, the coding part (less than 2% of the genome, containing information for the generation of proteins) and the non-coding part. The estimated number of coding genes by the ENCODE consortium is between 20.000 to 25.000 (Birney et al. 2007). Human genes are highly variable in their structure and size but they usually consist in introns (non-coding) and exons (translated into proteins), with a promoter sequence, as well as enhancer sequences for the transcription. Out of these, only exons belong to the coding part of the genome.

The non-coding part of the genome can be divided into several categories: pseudogenes, genes for non-coding RNA, introns and untranslated regions of mRNA, regulatory regions (up to 40% of the genome (Stamatoyannopoulos 2012)), repetitive DNA sequences and mobile genetic elements (transposons and relics).

Together with the human genome project, many online tools have been developed for browsing the genome and describing gene functions. The UCSC (Kent et al. 2002) and Ensembl (Hubbard et al. 2002) genome browsers allow users to browse the genome for regions of interest. Other databases such as Gene Ontology (Ashburner et al. 2000; Gene Ontology Consortium 2015) and KEGG (Kyoto Encyclopedia of Genes and Genomes) (Kanehisa and Goto 2000) aim at classifying genes in functional sets. The PANTHER (Nikolsky and Bryant 2009) database classifies protein-coding genes into 29 functions that I display in figure 6. This classification gives an idea of the major functions and their relative importance in the genome in terms of the percentage of genes from the total number of genes in the database contained in each category. The first category is nucleic acid binding, revealing the importance of DNA handling, second comes receptors followed by transcription factors, several enzymes and signaling molecules. The “defense/immunity protein” category comes at the 11th place.

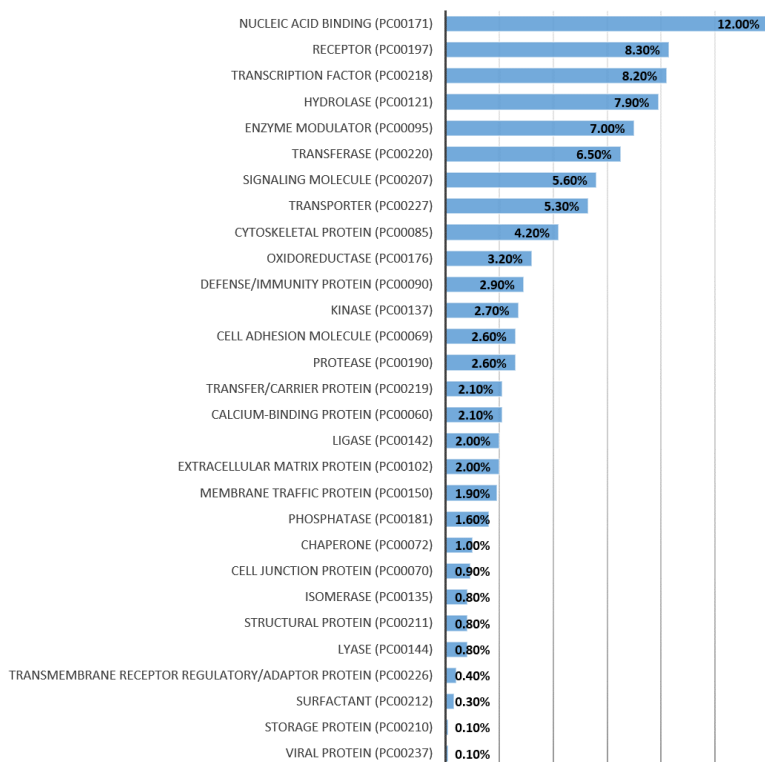


Figure 6: functional categories of human coding genes from PANTHER (Nikolsky and Bryant 2009). The percentage represents the number of gene in each category compared to the total amount of genes in the database.

Fourteen years after the first publication of the human genome, many tools are available to access the genetic information of individuals and study the genetic diversity in populations. The capture of single base variant diversity can be easily obtained by SNP array technologies with more than 5 million variants. The access to almost complete sequences of coding genes can be performed using whole exome sequencing technology, targeting the coding sequences in the genome. Finally, the high coverage sequences of complete genomes has become more frequent and vast quantities of data is constantly being generated (Consortium 1000 Genomes Project 2010; Drmanac et al. 2010; Lachance et al. 2012; Fu et al. 2013).

This technological revolution has influenced many areas of science and health care. In population genetics, it has permitted the study of human diversity, inferring population history and detecting regions of local adaptation. These new technologies also make it possible to perform genome-wide association studies, linking phenotypes with genetic variants to improve the genetic understanding of diseases and other phenotypes.

Detecting local adaptation in populations

The survival of an individual is partly dependent on how well adapted its phenotype is to the environment. Beneficial phenotypes will increase the chance of an individual for survival and transmitting its genetic material. Deleterious phenotypes on the contrary will decrease the chance of survival and mating. Alleles at the basis of beneficial phenotypes tend to increase in frequency in the population while alleles causing deleterious phenotypes usually decrease in frequency. These phenomena are the product of natural selection. There are several ways to measure the impact of selection on the genetic diversity in order to study the adaptive history of a population.

Selection is one of the evolutionary force that will disturb the Hardy-Weinberg equilibrium described earlier, since individuals carrying a favorable variant tend to have more offspring compared to other individuals carrying the non-favorable variant. Measuring deviations from the expected under the Hardy-Weinberg equilibrium can be a way to detect selection. F-statistics (Wright 1943; Wright 1950) are measures of heterozygosity in a population as a distance from the heterozygosity expected under Hardy-Weinberg equilibrium. For example, F_{ST} measures the fraction of heterozygosity that is between/among populations compared to the expected heterozygosity for the pooled population. Nei's definition of F_{ST} is (Nei 1977):

$$F_{ST} = \frac{H_T - H_S}{H_T}$$

H_T is the expected heterozygosity in the whole population if it follows Hardy-Weinberg, and H_S is the expected heterozygosity in subpopulations. If $H_S = H_T$, the F_{ST} is null meaning that both populations are equally diverse, whereas when $H_S < H_T$ the F_{ST} is positive showing a decrease in heterozygosity in the subpopulation that might be due to selection.

F_{ST} reveals allele frequency differences between populations, in other words, population structure. As selection impacts the frequency of beneficial variants in the population where it's acting, it can increase the population allele frequency differences of selected variants. At the whole genome scale (considering all variants together), F_{ST} is a measure of genetic differentiation between populations or structure.

Variants with a higher F_{ST} than the genome average might be due to the effect of selection in one population, since selection might dramatically affect the allele frequency of the selected variant in the population concerned. In that logic, the outlier variants with the highest F_{ST} values over the genome, can be considered as potential targets of local adaptation.

When positive selection is acting on an allele, its frequency tends to increase in the population over time. If an allele increases rapidly in frequency, depending on the recombination rate, the frequency of the alleles next to the selected allele may also be affected. This process is called genetic hitchhiking or selective sweep. Selective sweeps leave particular patterns in the genome: unusually long haplotypes, low genetic diversity and high level of linkage disequilibrium between variants. Several statistics have been developed for detecting such signals focusing on detecting combinations of the specific patterns of selective sweeps. In this thesis, I used three statistics based on haplotypes: iHS, XP-EHH and MFPH (paper II).

iHS and XP-EHH are both based on extended haplotype homozygosity (EHH). EHH for a locus l is the sum of the probabilities that two chromosomes which have the same allele at this locus, are identical by descent at other loci going further and further away from l (Sabeti et al. 2006). The iHS statistic (Voight et al. 2006), is an integration of EHH, taking into account the derived/ancestral states of alleles. When iHS is negative, the selected allele is the derived allele, whereas when it is positive, the ancestral allele is selected. XP-EHH (Sabeti et al. 2007) is an extension of EHH that contrasts two population. When XP-EHH is positive, it reflects potential selection in one population, while when it is negative it reflects selection in the other population.

MFPH, which I develop in paper II, is also based on linkage disequilibrium but is looking for private haplotypes, found in only one population. Private haplotypes can be a result of population structure (figure 7), but may also reflect selective events. Population structure affects all loci equally while selection affect specific loci. The private haplotypes with the highest frequencies genome-wide, are therefore likely to represent loci that have been affected not only by population structure but also by selection. This phenomenon is illustrated in figure 7.

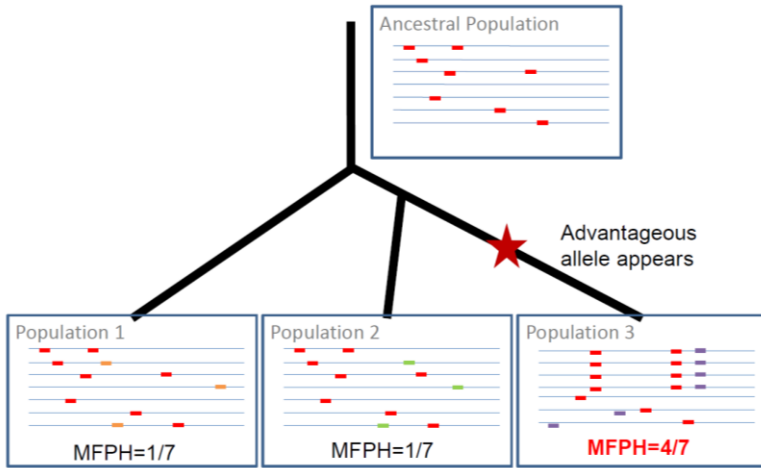


Figure 7: Principle behind MFPH. Population 1 and 2 evolved neutrally while selection occurs in an allele that appeared in population 3.

Since traits are usually complex, it is unlikely that selection on a trait is only linked to one variant of one gene. Moreover new mutations are relatively rare, and it is therefore unlikely that selection will bring them to fixation in a short time-span (unless the strength of selection is great). For these two reasons, it is more likely that selection affects several genes involved in the same trait, on pre-existing variation, this phenomenon is called polygenic selection (illustrated on figure 8) (Pritchard et al. 2010). In polygenic selection, variants existing in low frequency in the population before the selection pressure increase in frequency. New variants occurring during the selection pressure would also increase in frequency and create selective sweep patterns around them.

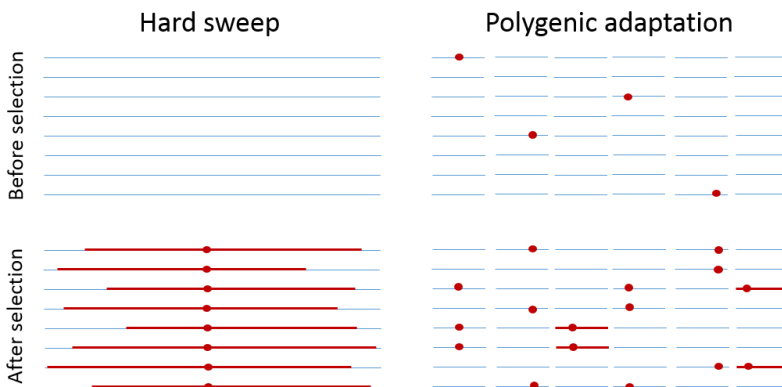


Figure 8: Illustration of hard sweep vs. polygenic sweep in adaptation. The blue lines represent haplotypes, and the red lines represent regions that are identical by descent (IBD). The red circles indicate alleles that are favored following an environmental change. The figure is adapted from Pritchard et al. (2010).

Inferring human demography

Inferring past human demography is of key importance for understanding the evolution and history of our species. Evolutionary events leave footprints in the genome that can be utilized for demographic inference. Several methods have been developed to estimate demography from genetic data.

Genetic drift is a consequence of a finite population size and is a key concept in demographic inference. Importantly it is inversely proportional to population size but also depends on population structure and breeding structure among other factors. Because population size is an important determinant of genetic drift, but not the only one, the concept of effective population size has been introduced. This concept translates observed genetic drift into a population size that would result in the observed genetic drift. Genetic drift affects genetic diversity in populations and can therefore be studied if one makes assumptions of the mutational and recombinational processes.

A widely used mathematical tool in demographic inference is the coalescent. Coalescent theory studies the effect of genetic drift backward in time on the genealogy of a sample of individuals. In simple terms, the coalescent relates the distribution of mutations in a sample of individuals to demography.

In this thesis, I used a Bayesian approach to investigate bottlenecks in early modern humans. Bayesian statistical inference is a method to infer parameters of interests using Bayes' theorem. Considering the genetic data D and an evolutionary parameter θ , Bayes' formula can be written as follows:

$$P(\theta|D) \propto P(\theta) \cdot P(D|\theta)$$

where $P(\theta|D)$ –the posterior distribution– is the conditional distribution of a parameter θ given the data D , $P(\theta)$ –the prior distribution of the parameter– and $P(D|\theta)$ the probability of the data given the parameter (also called the likelihood: $L(\theta) = P(D|\theta)$). In other words, the posterior distribution of a parameter is proportional to the product of the prior distribution of the parameter and the likelihood.

In human population genetics, because of the scale and complexity of the data, it is often difficult to evaluate the likelihood (Csilléry et al. 2010). It has therefore been suggested to use summary statistics to describe the data instead of using the whole data, which often has a too high dimensionality (Tavaré et al. 1997; Pritchard et al. 1999; Beaumont et al. 2002).

This method is referred to as Approximate Bayesian Computation (ABC) and is generally performed as follows:

- The data is summarized using statistics that are sensitive to values of the parameters of interest (time of divergence, effective population size or migration events for example).
- Models are designed to reflect knowledge of the studied population and scenarios that could explain the observed data. The parameters of interest are randomly sampled from a prior distribution. These models with the randomly chosen parameter values are then used to simulate data using a software such as *ms* (Hudson 2002).
- The summary statistics are then computed on the simulated data. Only simulated datasets showing summary statistics that are close to the summary statistics calculated on the real data are kept (rejection algorithm).

There are two aims in performing ABC: choosing the best model and estimating the values of the parameters of interest. When several models are considered, the probability of the observed data being a product of each models, the posterior probabilities of each model, can be used to selected the most likely model (Pritchard et al. 1999; Beaumont 2008).

For the estimation of the parameters, the distribution of the parameter values kept in the simulated data is called the posterior distribution of the parameter. Posterior distributions of parameters can be corrected using heteroscedastic regression adjustment (Blum and François 2010; Csilléry et al. 2010).

In this thesis, we tested models for a putative bottleneck in early anatomically modern humans, to test the hypothesis on the origin of modern humans presented in the introduction of this thesis. Three models were considered, a model without any bottleneck, a bottleneck happening in one population, or a bottleneck happening in fragmented populations. The two latter models have been suggested to be part of human history and possibly linked to the emergence of anatomically modern humans (Lahr and Foley 1998).

Because of the time scale of the event studied, the choice of summary statistics favored statistics that have power to detect relatively old events (Sabeti et al. 2002). The mean and variance of the number of SNPs, Tajima's D (Tajima 1989), Fu & Li's D* (Fu and Li 1993), Fay & Wu's H (Fay and Wu 2000) per 10kb windows were chosen as summary statistics.

Tajima's D (Tajima 1989) is defined as the difference between two estimators of $\theta = 4N_e\mu$ divided by the standard variation of this difference:

$$D = \frac{\hat{\theta}_\pi - \hat{\theta}_w}{\sqrt{\hat{V}(\theta_\pi - \theta_w)}},$$

where $\hat{\theta}_\pi = \hat{k} = \frac{1}{\binom{n}{2}} \sum_{i \leq j} k_{ij}$ and, $\hat{\theta}_w = \frac{S}{\sum_{i=1}^{n-1} \frac{1}{i}}$. $\hat{\theta}_\pi$ is the average number of pairwise differences (k_{ij}) between chromosomes in a sample of size n . $\hat{\theta}_w$ is the number of segregating sites S in a sample of n chromosomes divided by $\sum_{i=1}^{n-1} \frac{1}{i}$. If the population has had a constant size and no population structure, the expected value of D is zero. $\hat{\theta}_\pi$ is sensitive to the frequency of SNPs in the population while $\hat{\theta}_w$ is based on the number of segregating sites without considering their frequencies. If there is an excess of low frequency alleles, D is negative. If there is an excess of medium frequency alleles, D is positive.

Fay and Wu's H (Fay and Wu 2000) is also defined as the difference between two estimators of θ :

$$H = \hat{\theta}_\pi - \hat{\theta}_H.$$

$\hat{\theta}_H$ is defined as $\hat{\theta}_H = \sum_{i=1}^{n-1} \frac{2S_i i^2}{n(n-1)}$ where S_i is the number of derived variants found i -times in a sample of n chromosomes. $\hat{\theta}_H$ is mainly influenced by high frequency variants so that H will be sensitive to an excess or deficit of medium and high frequency alleles.

Fu & Li's D* (Fu and Li 1993) may be expressed as follows:

$$D^* = \frac{\hat{\theta}_w - \hat{\theta}_\eta}{\sqrt{\hat{V}(\theta_w - \theta_\eta)}},$$

where $\hat{\theta}_w = \frac{S}{\sum_{i=1}^{n-1} \frac{1}{i}}$ and $\hat{\theta}_\eta = \frac{n-1}{n} \eta_s$, η_s being the number of singletons, variants that only occur on a single chromosome in the sample. In this formula, $\hat{\theta}_\eta$ is sensitive to low frequency variants and therefore D* will be indicative of the relative amount of low frequency variants in the population.

Finally, Hudson's estimate of recombination rate C, estimated from the variance of pairwise sequence differences (Hudson, 1987), was included as a summary statistics in the ABC-setup.

Association studies

Association studies aim at understanding the genetic basis of phenotypic variation among individuals. Phenotypic variation can be explained by genetic or environmental factors. To study the genetic basis of phenotypic variation, it is important to have clear definitions of the studied traits. In studies aiming at determining the genetic basis of bimodal traits often used for diseases, the individuals are categorized into cases or controls. Other traits can be quantitative as they can take any value, such as height. In this thesis, I studied taste perception using categorical traits: the taste sensitivity is a quantitative trait but the study set-up cuts it into several categories. Before performing an association study it is essential to check whether a trait is heritable, i.e., transmitted from parents to offspring. Traits are usually influenced by the environment, height for example is heritable, but will also be determined by, for example diet or disease, which can also be correlated between parents and offspring. For this reason, it is complex to disentangle genetic and environmental effects in phenotypic variation. To do that, family and twin studies are usually performed.

A trait can also be correlated with population structure. In this case, variants that are also correlated with population structure might appear associated with phenotypic variation as a by-product of population structure (Hirschhorn and Daly 2005). Among these variants, it is essential to distinguish the variants that are effectively involved in the phenotypic variation from the ones that only are correlated with the population structure. For this reason, it is of key importance to take population structure into account when performing association studies.

With the advances of genotyping and sequencing technologies, genome-wide association studies (GWAS) are nowadays possible. The idea behind GWAS is to scan the genome for alleles associated with a trait without a priori assumptions of the genetic basis of the trait. The first GWAS was performed in 2005 on macular degeneration (Klein et al. 2005), and huge databases for disease association have been generated on up to 200,000 patients (International Consortium for Blood Pressure Genome-Wide Association Studies 2011). GWAS have improved our understanding of many traits, but the genetic basis of traits remains difficult to understand. Statistical analysis of large scale data implies multiple testing and linkage between variants as well as population structure can produce false positives. Several programs have been developed to perform GWAS based on various statistical methods. In this thesis, PLINK and GEMMA were used.

PLINK (Purcell et al. 2007) performs standard GWAS and large-scale data management. It allows the user to perform different types of associations of

phenotypes such as case/control or quantitative traits with large-scale genotype data. It also allows to add covariates in the association study, such as population structure or sex. In this thesis, linear regression between phenotype and genotype data was used, as the taste phenotype tests can take several values.

GEMMA (Zhou and Stephens 2012) performs a linear mixed model between genotype and phenotype data aiming at accounting for population stratification and relatedness in association studies. The observed phenotype is considered as a combination of several factors: fixed effects (age, sex, etc.), the marker effect, a random effect due to structure/relatedness and a random effect due to stochasticity.

As it permits to better account for population structure, I used GEMMA when using data from structured populations. Nevertheless, GEMMA cannot be used for low sample sizes. In these cases, I used PLINK.

Enrichment analysis applied to genetic data

Today, human genetic studies can produce large amounts of data that can be complex to make sense of. For example, the number of candidate loci identified in genome scans or association studies is typically large and difficult to assess. Moreover, the large amount of genetic variants in the data, generate considerable multiple testing issues.

The motivation behind enrichment analysis is to test whether it is possible to see patterns in high dimensional data. For example, scans of selection or association can be interpreted in a functional way (Huang et al. 2009). The basic idea is to test whether a gene or a set of genes contains a greater fraction of extreme values than expected by chance. A set of genes is usually a list of genes put together because of their involvement in a functional or metabolic pathway. Another advantage of enrichment analysis is to give more weight to not very strong signals. For example, p-values at 0.01 are not significant in a genome-wide scan considering millions of variants, but they can reflect small effects that are scattered over a whole set of genes. In that case, their distribution among gene-sets becomes informative.

Many databases have been developed to group the knowledge about genes. UCSC genome browser, Ensembl and PANTHER have already been mentioned. The Gene Ontology database (Gene Ontology Consortium 2015) and the KEGG encyclopedia (Kanehisa and Goto 2000) can be used as gene-set databases for enrichment analysis. The reason for this is that they group

genes together with regard to their function or involvement in biological processes.

The Gene-Ontology project aims at integrating the available knowledge of human genetics to get a more global view of mechanisms and functions, as well as giving a consensus vocabulary for genes and their functions. The term ontology refers to a representation of something for which there is knowledge. The Gene Ontology database consists of gene ontology terms (GO-terms) that are classes of genes sharing similarity in different domains. There are three main domains of GO-terms: biological process, molecular function and cellular component.

The KEGG database started together with the human genome sequencing project in 1995. KEGG contains three databases: PATHWAY, GENES and LIGAND. Similar to the Gene Ontology database, the PATHWAY database, used in this thesis, can represent higher order functions in terms of the network of interacting molecules. The pathways are divided into seven groups: metabolism, genetic information processing, environmental information processing, cellular processes, organismal systems, human diseases and drug development.

The general idea of an enrichment analysis is to look for patterns in the candidate variants or genes given by a specific analysis. Several programs have been developed to perform enrichment analysis using gene-sets. Their approach can be based on different statistical approaches such as chi-square, Fisher's exact test, Binomial probability or Hypergeometric distribution (Huang et al. 2009). Here, I present two programs to perform enrichment analysis, David (Huang et al. 2007) and Gowinda (Kofler and Schlotterer 2012).

David provides a gene-based online enrichment tool. It considers a list of gene IDs and checks for enrichment based on a Fisher's exact test algorithm in various pathways or functional groups. An advantage of David is that it gives insight from several databases such as KEGG pathway as well as links with diseases. A drawback of David is however that it does not account for the gene-length. Longer genes are more likely to contain candidate variants and are therefore more likely to end up in the gene-list considered for the enrichment analysis. Therefore, the way the candidate gene-list is established is of key importance when using David.

Gowinda is an unbiased SNP-based enrichment program. It considers a list of positions, links them to the genes they are located in, and tests for enrichment of gene-ontology terms using a permutation test. Different

options are available to account for linkage disequilibrium between loci and gene size.

Enrichment analyses are useful both in association studies and selection studies. Indeed, phenotypic variation of a trait might reveal the effect of several loci linked to various genes supposedly belonging to the same biological process or presenting similar molecular functions. Each locus only explains a small part of the phenotypic variance and not turn out significantly associated because it should be considered together with other loci linked to genes of a same gene-set. In a similar manner, in selection scans, some variants may only be moderately advantageous because selection affects several genes in e.g. the same pathway (polygenic selection). An enrichment analysis could potentially reveal such complex signals.

Psycho-physiological tests for taste perception

Psycho-physiological tests are non-invasive ways to study the response of a person to stimuli in order to better understand cognitive processes such as taste perception.

There are different concepts in taste perception: “taste sensation” is a taste stimuli, “taste acuity” (or taste sensitivity) is the ability of a person to detect a taste, “taste intensity” is how intense a person judges a taste sensation, and “hedonic sensation” is how a person experiences (liking/disliking) a taste sensation.

Because the overall taste experience is actually a combination of gustatory, olfactory and visual stimuli, the study of taste perception in its strict sense is usually reduced to the dissolution of non-odorant and invisible compounds in water. In this thesis, taste acuity was studied, as it is believed to be the least influenced by judgement and potentially closest to reflecting the determinism of taste perception. Taste acuity is evaluated using measurement of a taste threshold: the minimum concentration at which a taste is perceived. To do this, tests are developed in which a participant is asked to taste solutions of single compounds dissolved in water with increasing concentration (staircase method).

In this thesis, a field-adapted test was performed as follows:

- Participants are informed that they might encounter tastes like water, salty, sour, sweet, bitter and umami (described as savory, meaty, or bouillon taste).
- Participants are semi-blindly (the tester knows the order) given to taste solutions with increasing concentrations of a pure compound dissolved in water. After tasting each solution, they express the taste they perceived and are asked to rinse their mouth with the water used for the test.
- Once participants clearly determine the actual taste of the solution for two consecutive concentration, the threshold is set as the least concentrated of the two.

This test is adapted to field conditions, mainly for time constraints. There are some caveats to this method. The first is that the test is not repeated although answers to the test of a same person might vary when the same test is duplicated. The second is the water used for the dissolutions. For taste psycho-physiological tests, the usage of the least mineral-rich water is recommended in order not to influence the answer of the participants. Under field conditions, the access to ideal water is often not possible. In these cases, we used filtered-water from local sources (in small villages with little access to shops) or bottles (in cities), for all individuals of the same sample group. Although the water we used might have a higher mineral content than the recommended waters, they usually have the neutrality of the every-day water consumption.

One issue when performing tests in groups with multiple languages is the vocabulary to describe tastes. Furthermore, the translator can also impact the taste. I list in table 2 and 3 the main words used for basic taste description in the mother tongues and language spoken by the participants. It is to notice for example the use of the same word to describe two different tastes, for example “ngbembembembe” in the Baka language refers both to sour and salty, or “lokoloko” used for sweet can also mean “good salty taste”. Similarly, in Kazak the word “ащы” can mean bitter, salty or sour. Translators were always present during the tests and further precision for the described taste was sometimes necessary to avoid ambiguity caused by e.g. word-use in specific languages.

A final language issue is related to umami, a Japanese notion that in many languages does not have any clear definition or word. Umami literally means “savory” in English. It is generated by some amino acids and nucleotides. In Uzbekistan it can be designed as “Chinese salt” because of the rather common use of imported monosodium glutamate from China for cooking.

Generally, it is possible to remind the umami taste with the term “bouillon cube” as it is a widely used ingredient rich in glutamate.

Table 2: Vocabulary for basic tastes in Cameroonese languages and correspondence with English. The words are given in phonetics, for Baka I give an approximate pronunciation in common alphabet in quotation marks. (Brisson 1984; Beavon and Beavon 1996).

English	French	Baka	Nzime
Sweet	sucré	lɔkɔlɔkɔ “lokoloko”	esweh
Bitter	amer	sòsò “sousou”	ejwie
Salty	salé	ngbɛngbɛ̀ngbɛ “ngbemgbemgbe”; lɔkɔlɔkɔ “lokoloko”	ekpa
Sour	acide	ngbɛngbɛ̀ngbɛ “ngbemgbemgbe”	ebyah

Table 3: Vocabulary for basic taste in Tajik Uzbek and Kazak, correspondence with English. Words are given in Russian alphabet with their phonetic pronunciation in parenthesis.

English	Russian	Tajik	Uzbek	Kazak
Sweet	сладкий (slɒdkəj)	ширин (ʃirin)	ширин (ʃirin)	тәтті (tətti)
Bitter	горький (gɔrˈkəj)	талх (talx)	тахир (taxir); аччик (aʃʃiq)	ашы (aʃi)
Salty	солёны (solˈonnəj)	шўр (ʃur)	тузли (tuzli)	ашы (aʃi); тұзды (tuzdi)
Sour	кислый (kisləj)	турш (turʃ)	нордон (nordon); аччик (aʃʃiq)	ашы (aʃi); қышкыл (qifqil)

Research Aims

The main goal of this thesis was to study early human demographic history and local adaptation. Understanding how humans adapted to their environments not only tells us about our history but can help us to better understand susceptibility to diseases in a globalized world. One aspect of adaptation to the environment is the adaptation to diet, which is tightly linked to lifestyle. For this reason, a more specific goal of this thesis was to understand the evolution of taste perception in concert with lifestyle, since taste has a key role in alimentation. More specifically, the aims were to:

- I. Infer early human demography between 100kya and 200kya which has been suggested to be a key period for the emergence of anatomically modern humans.
- II. Develop a tool for detecting local adaptation in populations by studying private haplotypes.
- III. Identify genomic regions potentially involved in local adaptation of various human populations.
- IV. Study taste perception in populations differing on lifestyles in order to understand the genetic basis of taste perception and its evolution in human populations.

Results and discussion

Origins of modern humans (Paper I)

Our species is believed to have emerged some 100-200kya in Africa. This period coincides with a glacial era referred to as MIS6 (190-130kya), which has been suggested to be important in the forming of the genetic diversity of modern humans. During this time, the environment was considered to be less favorable for early humans and therefore the human population size was severely reduced during this period. Models such as the “fragmentation” model and the “founder” model, have been used to describe human demography during MIS6 (figure 9). According to the fragmentation model, small groups were isolated in refugia without contact to each other the end of the glacial period, when contact was re-established between these populations. The founder model posits a single population that underwent a severe bottleneck during MIS6. In **paper I**, we tested which of three models (figure 9), best fit empirical genetic variation in three African populations. We compared three models for the human population size during MIS6: the founder model, the fragmentation models and a model with constant population size.

We simulated data under all three models to fit data from three African populations. These were the San from southern Africa, the Biaka from Central Africa and the Mandenka from West Africa. None of these three populations were affected by the well-documented Out-Of-Africa bottleneck. The data for this study was that of previously published sequence data of 61 autosomal loci far from genes and supposedly neutral were used (Wall et al. 2008, Hammer et al. 2010). Each of the 61 loci is 20kb long and composed of three fragments of 2kb separated by 7kb. This data constituted the observed data.

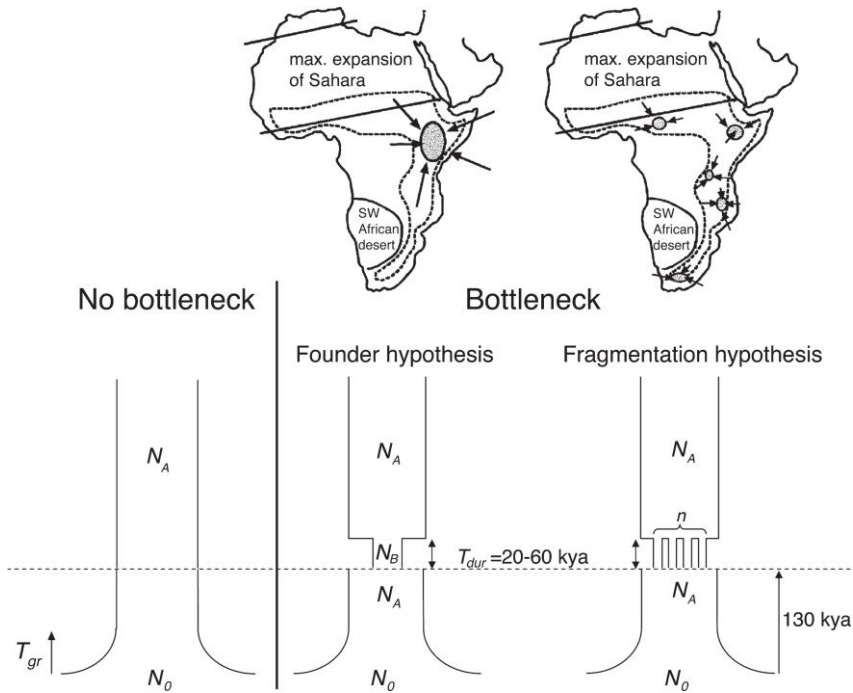


Figure 9: Schematic overview of the three investigated demographic models. The end of MIS6, approximately 130 kya, is marked by a dashed line. The map of Africa where potential human refugia are displayed is an excerpt from Lahr and Foley (1998). We denote by N_A and N_B the effective population sizes before and during the bottleneck, by N_0 the present-day effective population size, and by n the number of subpopulations during the bottleneck.

We simulated 100,000 datasets from each of the three models and kept the 0.5% of simulated datasets from each model for which the summary statistics were the closest to the observed summary statistics. By fitting a multinomial logistic regression between the models and the simulated summary statistics, we calculated the probability of each model accounting for the observed summary statistics. The no-bottleneck model had the highest posterior probability (at least 79%) in all three populations.

Because the event we test happened at least 130kya, we checked whether the data contained enough information for the summary statistics to detect such an ancient bottleneck. We replaced the observed dataset with a simulated one and checked how often we would assign it to the model it was generated from. Simulated data from the no-bottleneck model were correctly assigned to the no-bottleneck in 79% of the cases. However, data simulated with a weak bottleneck (under the “founder” or the “fragmentation” models) were often wrongly assigned to the no-bottleneck model (figure 10). This is not surprising considering that a bottleneck model with a bottleneck intensity of

0 is in fact identical to the no bottleneck model. This reveals that the data should permit us to detect a bottleneck with at least a 3-fold reduction of the population.

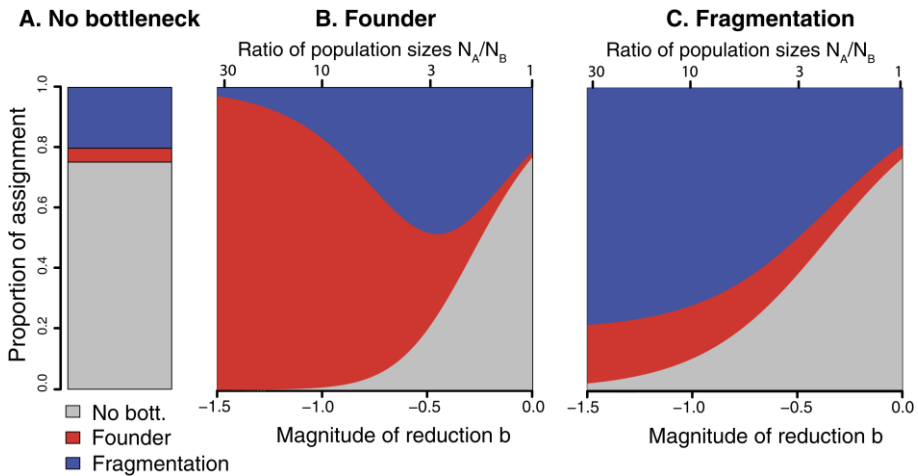


Figure 10: Proportions of simulated data assigned to each model when simulations were performed under the (A) no bottleneck, (B) founder, and (C) fragmentation model. For each model, we performed model selection using 100,000 simulations per model. For each model, the proportion of correctly assigned simulations was estimated with a total of 5,000 simulations that were considered as target data. For the two bottleneck models, we used a smoothing procedure (multinomial logistic regression) to display the proportions of simulations assigned to each model as a function of the bottleneck parameter b , with $N_B = N_A \times 10^b$.

Thus, if there was a bottleneck at the origin of our species, it was most likely less severe than a 3-fold decrease in population size. More and more genome databases from African populations are released (Consortium 1000 Genomes Project 2010; Lachance et al. 2012; Kim et al. 2014; Gurdasani et al. 2015), which could provide increased statistical power for further investigation of the demographic details during MIS6.

PSMC (Li and Durbin 2011) and MSMC (Schiffels and Durbin 2014) are two recent methods for inferring population size over time using sequential Markovian coalescent models that are becoming popular to infer demography of populations. The original paper of PSMC, shows an increased population size between 60kya and 200kya (Li and Durbin 2011). Also, the recent use of PSMC on genome data to infer demography on Khoe-San, other Sub-Saharan and non-African population (Kim et al. 2014) identifies two bottlenecks: the most recent around 30-120kya affecting all populations except Khoe-San and a more ancient from 1mya and 0,5mya. According to their results, the period between 200 and 100kya (corresponding to the MIS6) is a period of population expansion. Although

this expansion is consistent with the fragmentation model, which could inflate estimates of effective population size during MIS6, our results suggest that a fragmentation model is unlikely to be the only explanation for the large population size during MIS6 that they estimate.

There seems to be little consensus from the various methods mentioned regarding the inference of ancient demographic events. It would therefore be interesting to apply ABC methods to genome-wide data, and to compare it to results obtained from other methods such as PSMC and MSMC.

Methods for detecting local adaptation (Paper II)

Another aim of this thesis was to study patterns of local adaptation in human populations. In **paper II**, we developed a statistical tool to search for local adaptation based on private haplotypes. A “haplotype” is defined as a specific combination of alleles at the variable sites in a predefined window. In general, the fact that a haplotype is much more common in one population than in other populations may be the result of population specific selection in the region of this variant. Such a haplotype is also likely to be “private” (found in one population only). Moreover, the stronger the selection, the higher the expected frequency of the selected haplotype. Given samples from different populations, “private” means that a variant is only present in one of the sampled populations. We consider the Maximum Frequency of Private Haplotypes (MFPH) as a statistic to detect local adaptation.

To investigate the ability of MFPH to detect local adaptation, we simulated data from a simple population split model with one mutation – selected or neutral – appearing in one population after the split (figure 11A). Under this model we used several selection coefficients, migration rates, sampling times and recombination rates in order to study how these parameters affect MFPH. This simulation study revealed that MFPH can detect selective sweeps (figure 11B).

By performing a MFPH scan on publicly available genotypic data (HapMap III), we were able to detect already known signals of selection such as selection around the lactase gene (*LCT*) in populations with a diet rich in dairy products (Maasai and Europeans) (Tishkoff et al. 2007; Schlebusch et al. 2012) as well as selection around *EDAR* and *ADH1B* genes in East Asians (Li et al. 2008; Tan et al. 2013).

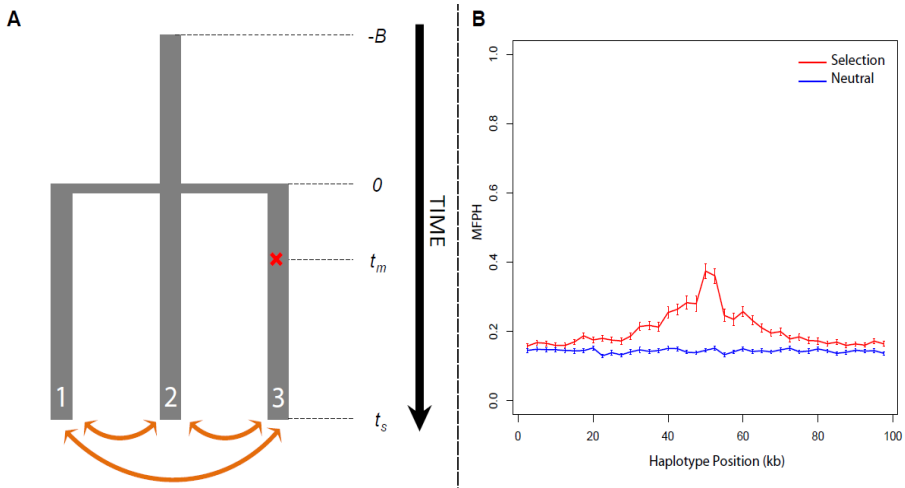


Figure 11: Scheme of the model used for simulations. A: An ancestral population of 500 diploid individuals reaches Wright-Fisher equilibrium during B generations and then splits into three populations of 500 random-mating diploid individuals each. At time t_m a mutation occurs in the middle of the sequence in population 3. B: At time t_s (50 generations), 15 individuals are sampled from all populations and MFPH is computed in population 3 over 100 simulations, in blue: the mutation introduced is neutral, red: the mutation introduced is selected for. In this particular case, the parameters are $\rho=0.005$, $G=150$, $m=1$, $\theta=0.001$, $t_m=100$, $N=500$. Window-size=5kb, step=2.5kb

MFPH is a window-defined statistic, and one can choose how the windows are defined depending on the data and the interest. A window can be defined as a fixed number of base-pairs, a fixed number of centimorgans when a genetic map is available or as a fixed number of SNPs. Regardless of the definition of the window, the size of the window will influence the frequency of the haplotypes and their probability to be private and therefore it influences the value of MFPH. The larger the window, the more likely a haplotype will be private, but it will also have a lower expected frequency.

Thus, MFPH is an easily computed statistic that can detect local adaptation both in simulated and real data. Further studies are needed to investigate how MFPH behaves in more complex demographic scenarios such as bottleneck models. That bottlenecks will have a dramatic effect on MFPH is illustrated by the fact that the MFPH scan using HapMap III data showed that MFPH have more variance and reached higher values in East Asians than in East Africans, a result which is most likely linked to the out-of-Africa bottleneck that affected East Asians but not East Africans.

An instance where MFPH could outperform other statistics is for detecting selection on standing variation. It is likely that many selective events (especially local adaptation) may act on variation already present in the populations. Such selective events have been revealed to be very difficult to detect using statistics designed for traditional selective sweep detection. Nevertheless, even if an adaptive event happens locally in one subpopulation on a variant shared among several populations, it could still create private haplotypes. One site might be polymorphic in several populations but the haplotypes might be different across populations. For this reason, it will be interesting to investigate the properties of MFPH under various scenarios of selection on standing variation.

Detecting local adaptation in Humans (Paper II, III and IV)

Human populations have peopled and adapted to almost all terrestrial environments. The understanding of the genetic basis of local adaptation in human populations is a key topic in human population genetics. Although local adaptation is intensively studied, it is still poorly understood. This is probably due to methodological reasons, since most of the methods aim at detecting hard sweeps, but also to the so far uneven sampling of human populations around the world.

In **paper II**, we specifically looked for regions in the HapMap III data where a specific population show a signal of high frequency private haplotype using MFPH, and where other statistics do not show a strong signal (table 4). Interestingly, on chromosome 3 we found a region potentially selected in Maasai between 50.6Mb and 51.3 Mb previously shown to be linked to height (Jarvis et al. 2012). Also potentially selected in Maasai, the region containing the gene *IMPG2* coding for an interphotoreceptor proteoglycan that has been associated with retinopathy (Simo-Servat et al. 2013) and potentially have been selected in Maasai in response to a high solar exposure probably damaging to the eyes.

Table 4: The regions with the highest MFPH value across the Hapmap III data after excluding chromosome 2 (LCT and EDAR are located on chromosome 2).

Population	Region	Genes
CEU	Chr10:74,416,452-75,102,866	<i>MCU, OIT3, PLA2G12B, NUDT13, ECD, DNAJC9, MRPS16, TTC18</i>
CEU	Chr6:145,190,620-145,554,235	
MKK	Chr3:100,979,041-101,375,515	<i>IMPG2, SENP7, PCNP</i>
MKK	Chr3:50,617,979-51,354,540	<i>HEMK1, CISH, MAPKAPK3, DOCK3</i>
JPT + CHB	Chr15:62,232,223-62,888,060	<i>TLN2, VPS13C, C2CD4A, C2CD4B</i>

Using genome data from various African populations, we performed selection scans using iHS (Voight et al. 2006) and XP-EHH (Sabeti et al. 2007) (**paper III**). Population grouping at several levels, permitted us to study local adaptation at various scales, from regional to continental scales. We considered signals of adaptation in a diverse set of Central Africans, Northern San, Southern San and Khoe-San.

In Northern San, who are characterized by a hunter-gatherer lifestyle, we find a signal of adaptation around sequences potentially protective against the anthrax bacteria *Bacillus anthracis*, which is frequent in west southern Africa. We also find selection on genes related to fat storage especially in women-related fat storage, glycogen degradation and dietary absorption of iron. Selection on these genes are potentially related to adaptations to a hunter-gatherer diet that may be characterized with uneven food availability.

In Southern San, presenting both pastoralist and hunter-gatherer lifestyles, we find signals of adaptation in two milk-related region, probably linked to the pastoralist diet highly dependent on milk consumption of some of the Southern San groups, (Lombard and Parsons 2015). We also find selection upstream of a gene upregulated during viral infection and on histatins, salivary peptides with antibacterial and antifungal functions.

When considering signals selected in Khoe-San together compared to other Africans, we find selection around a gene involved in spermatogenesis and also linked to adult height. Another region, containing an ion channel linked to cognitive performance and several diseases, presents with a signal for selection. Interestingly, two of the diseases associated with this gene are characterized by light skin and short stature, also characteristic of Khoe-San groups compared to other Africans. It could be that this gene had been selected and co-incidentally had effect on skin pigmentation and height.

In sub-Saharan Africans (west, central and east) compared to Khoe-San, we find a signal in XP-EHH upstream of *MINPP1* coding for an enzyme hydrolyzing phytic acid. Phytic acid is found in high concentration in crops as a storage of phosphate for plants but cannot be assimilated by humans. High consumption of phytic acid can be disadvantageous to humans as it chelates minerals and vitamins. Its degradation could therefore be advantageous in farming populations who consume high amounts of cereals in a less diverse diet than hunting and gathering societies.

Taken together, these analyses showed that diet and diseases are key drivers of adaptation in African populations.

Another way to look for adaptation is to perform enrichment analysis on the output of various selection scans. Enrichment analysis on functional set of genes using selection data can provide an idea of polygenic selection. To do that, it is possible to use functional sets of genes under the form of Gene Ontology terms (GO-terms).

In **paper IV**, I performed Gene Ontology enrichment analyses on selection patterns in populations with different lifestyles. I studied Baka hunter-gatherers and Nzime farmers from Cameroon. I also studied two Uzbek groups, one with Kazak (traditionally nomad-herders) origins and one from the Tajik-speaking area with farming lifestyle. The results showed that some of the top selected Gene Ontology terms (GO-terms) indicated by iHS, XP-EHH and F_{ST} , were potentially related to lifestyle.

The GO-term “striated muscle development” is enriched in F_{ST} outliers among F_{ST} signals between Bakas and Nzimes. The phenotypes associated with this GO-term are not known, but it could be related to height. For example, a recent study showed that Baka children have an increased muscular development relative to children from other populations (Rozzi et al. 2015). This is potentially linked to the average short stature of the Bakas in adulthood.

In Nzime farmers, GO-terms involved in response to vitamin and immune response to viruses are enriched in iHS outliers. These enrichment signals are potentially linked to the farming lifestyle. The transition to farming likely caused health issues, such as lacks of certain nutrients and an increased occurrence of epidemics (Armelagos and Cohen 1984). A GO enrichment analysis of selective events specific to Nzime as revealed by XP-EHH between Nzime and Baka is also consistent with adaptation to a farming subsistence. The GO-terms enriched in this analysis were “interspecies interaction between organisms” and “response to fungus”, which are clearly in line with selection for coping with problems due to the transition to farming.

When comparing populations from farming and nomad-herding origins, the F_{ST} enrichment pattern showed a higher divergence than expected in the “negative regulation of lipoprotein lipase activity” between Tajiks and Kazaks. This is potentially linked to lifestyle as Tajiks and Kazaks do not have the same amount of animal products in their diet as well as possible differences in energy expenditure. Also potentially linked to the nomad-herder lifestyle in the Kazaks are the GO-terms “response to immobilization stress”, “interspecies interaction between organism” and “viral process” enriched in XP-EHH outliers in the Kazaks when comparing to Tajiks.

Summarizing these different observations, I find that selection scans show regions of the genome that are possibly targets of selection which often make anthropological sense. Furthermore, enrichment analysis can bring more sense to large-scale outputs in addition to searching in literature databases. Nevertheless, they should only be seen as suggestions for further investigation. Functional studies, such as those using phenotypic data, are one potential strategy for a better understanding of adaptation.

Taste perception and lifestyle (Paper IV)

Taste is important for dietary choices, as it helps to select beneficial foods and avoid harmful ones. Lifestyle and environment have a strong impact on the availability of different foods and for this reason, taste might have been under various selection pressures in human populations. In **paper IV**, I present the results of taste sensitivity phenotypes together with genotype data from four populations differing in lifestyle and environmental diversity from Cameroon and Uzbekistan. I contrasted Baka hunter-gatherers and Nzime farmers living in the Cameroonian rainforest as well as nomad-herders of Kazak origin (referred to as Kazaks) and farmers of the Tajik-speaking area of Uzbekistan (referred to as Tajiks).

I performed various association studies on taste phenotypes and confirmed previous knowledge about taste determinism of sensitivity to bitter-tasting compounds 6-n-propylthiouracil (PROP) and Quinine. We confirmed the association between haplotypes of the SNPs rs10246939 (Ile296Val), rs1726866 (Val262Ala) and rs713598 (Ala49Pro) in the *TAS2R38* gene with variation in PROP perception. The signal was stronger in Uzbekistan than in Cameroon, and this is probably due to a higher genetic diversity in Cameroonian populations (Campbell et al. 2011; Robino et al. 2014). In both areas, the PAV (Pro-Ala-Val) haplotype is dominant and confers hypersensitivity to PROP. We also detected associations between several *TAS2R* genes previously shown to be associated with sensitivity to Quinine and also in vitro responsive to Quinine stimuli (Meyerhof et al. 2010; Reed et al. 2010).

Our results also suggest links between Gene Ontology terms (GO-term) and taste perception that were not previously reported (see table 5). GO-terms for rhombomere and kidney development appear to be associated with Fructose tasting. These GO-terms either contain genes previously shown to be involved in taste perception and food intake, or are linked with the development of the facial nerve necessary for the transport of taste stimuli to the brain. Moreover, the GO-term “stereocilium” is associated with the overall sensitivity (calculated across all taste phenotypes). Stereocilium is a

microvilli structure potentially similar to the membrane structure of taste receptor cells. Apart from previously identified genes involved in bitter tasting, the GO-term “olfactory receptor activity” is associated with Quinine perception. Although no previous link has been made between olfactory receptors and bitter taste perception, some of these receptors are expressed in the oral cavity and could be responsive to taste stimuli as well. GO-terms related to the neurotransmitter norepinephrine are associated with salt perception. Finally, GO-terms related to localization to mitochondrion are associated with sucrose tasting.

Table 5: GO-terms with FDR<10% for enrichment in taste perception associated candidate variants (p-values<10⁻³) based on meta-analysis and for SPAs (single population associations) and RAs (regional associations). The overall sensitivity phenotype is calculated as the mean of the scaled phenotypes for perception of quinine, fructose, sucrose, citric acid, monosodium glutamate and salt.

Meta-Analysis on:	Taste	GO-term	FDR	Description
SPA	Fructose	GO:0021546	0.03178	rhombomere development
SPA	Fructose	GO:0090183	0.08547	regulation of kidney development
SPA	PROP	GO:0006858	0.038322	extracellular transport
SPA	Quinine	GO:0004984	0.05467	olfactory receptor activity
SPA	Quinine	GO:0050909	0.05467	sensory perception of taste
SPA	Sucrose	GO:0006626	0.03196	protein targeting to mitochondrion
SPA	Sucrose	GO:0070585	0.049495	protein localization to mitochondrion
SPA	Sucrose	GO:0072655	0.06765	establishment of protein localization to mitochondrion
RA	NaCl	GO:0042415	0.03688	norepinephrine metabolic process
RA	NaCl	GO:0042421	0.03759	norepinephrine biosynthetic process
RA	Overall sensitivity	GO:0032420	0.03672	stereocilium

Other genes from the mentioned gene ontologies have not previously been associated with any variation in taste perception yet. Previous studies have focused on taste receptor genes when studying variation in taste perception among humans, but our results suggest that variations in receptors are not the only explanation for phenotypic variation in taste perception. Additional GWAS studies on taste perception are therefore important to further elucidate the genetics of taste perception.

Analyzing the variability of phenotypes reveals that Bakas and Nzimes differ on their overall sensitivity, Fructose and Quinine perception. For these three phenotypes, Bakas are less sensitive on average. Tajiks and Kazaks differ on

two tastes: Fructose to which Tajiks are more sensitive, and Salt to which Kazaks are more sensitive. The results from these enrichment analyses highlight that genes related to taste perception may indeed have co-evolved with lifestyle. More specifically, the observed overall sensitivity difference between Bakas and Nzimes can be related to divergence in the taste signal transduction pathway while single-taste differences between Tajiks and Kazaks rather relate to divergence of corresponding receptors. Moreover, the GO-term “tongue morphogenesis” is enriched in F_{ST} outliers among F_{ST} comparison between Bakas and Nzimes which could explain the difference in overall sensitivity between these two populations.

The gene *TAS2R19* is an iHS-outlier in Nzimes. Variation in this gene presents the strongest association to variation in Quinine perception (Reed et al. 2010). Quinine is a natural anti-malarial drug and being able to detect it in food may be beneficial in areas where malaria is endemic. Another gene, *TAS2R42* is an iHS outlier in Bakas. This gene has been associated with phantogeusia, a bitter taste sensation in the absence of stimulation by food. For XP-EHH contrasting Bakas and Nzimes, we also find *TAS2R*-genes coding for receptors responsive either to Quinine or Chloroquine among both positive and negative XP-EHH outliers suggesting a convergent evolution on genes involved in detection in anti-malarial substances in both populations.

In Central Asia, the GO-term “detection of chemical stimulus involved in sensory perception of taste” (GO:0050912 FDR=4%) is enriched in XP-EHH outliers in the Kazaks when compared to Tajiks. This suggests the presence of more recent local adaptation of taste perception in Kazaks than in Tajiks.

Overall, our results suggest that taste perception is a key factor in adaptation to lifestyle and environment as it seems to have evolved with lifestyle in various populations. Also, variations in taste perception seem to be linked to more genes than those coding for taste receptors only.

Conclusions and future prospects

By analyzing various types of genetic data, I have been able in this thesis to draw several conclusions about the origins and adaptation in humans. First, there was probably no strong bottleneck for humans between 190 and 130kya, which has been suggested previously to explain the low diversity of humans and to coincide with the emergence with anatomically modern humans. Second, the study of selection patterns in various human populations suggests that diet and pathogens have been driving adaptation in African populations, which is likely to be true for all humans. Last, by using phenotypic data from populations differing in lifestyle, I could show that taste genetically evolved in concert with lifestyle, but also that more genes are involved in the variations of taste perception than the number that is usually studied.

Currently, there are many exciting challenges in the field of human genetics. Our genome is one of the best understood (if not the best understood) and we are among the species for which great reference sequence data exist and abundant information (both trait and genetic) from many individuals is readily available. A lot of data has recently been generated and much more will be generated in the near future. Also, the extent to which various human groups are being genetically documented is improving permitting to draw the big picture of human history from the genetic perspective. For a long time, the theory was outperforming the amount of data available, but today, that is not the case anymore. Because of this effect, one of the current challenges is to improve our ability to analyze such huge datasets in a reasonable amount of time.

Studying adaptation in humans is challenging, although a lot has been achieved by the outlier detection approaches. The mechanisms of adaptation are in most cases complex and there is a need to develop tools for detecting such signals. Enrichment analysis is one option, and still needs to be improved both by the better description of the gene ontologies and gene pathways, and improving the enrichment methods themselves. Indeed enrichment analysis on such variable targets as human genes is complex since there is no real unit to compare genes with each other in a statistical test. It is important to have in mind that improving gene ontologies is easily said, but is not an easy task in practice. It involves the assimilation of many

molecular genetics studies and gene annotation studies to find out the function of genes and their interactions. Another way to increase our knowledge in terms of adaptation is to give more focus to developing and improving the methods for detecting selection on standing variation.

The tremendous development in technology for obtaining genetic data will help us decipher the genetic basis for traits, but what is of key importance is the phenotypic variation. It is essential to well characterize the phenotypes of interest, which is likely a limiting factor in many large association studies of complex traits. Even in the case of large collaborations, where sometimes hundreds of thousands of participants are investigated, the phenotype potentially remains an issue because of confounding factors, such as environment, social, or human subjective categorization. In the case of disease association study for example, the assignment of case or control to participants can hide complex mechanisms, since a single symptom or disease can have several unrelated causes.

In the case of taste perception, the results of this thesis suggest that the determinism of taste perception involves more genes than the taste receptor genes. It would therefore be important to improve the phenotype assessment and consider additional taste-related phenotypes. One of the phenotypes that would be extremely interesting to understand better is the density of fungiform papillae on the tongue. It is a phenotype that should be straightforward to measure and it will likely not be influenced by many confounding factors. It has been shown to influence the sensitivity to all tastes, but the genetic basis for this trait is poorly understood.

To conclude, molecular genetics is rapidly advancing and human genetics is in the forefront. Yet, much of human biology is still poorly understood, but the next step in its understanding will at least not be hindered by lack of genomic information. Thanks to these advances in genomics, the next few decades will be exciting for human genetics and human evolution.

Svensk sammanfattning

I den här avhandlingen undersöker jag först en specifik fråga angående vår arts ursprung. Jag uvecklar sedan en metod för att upptäcka positivt selekterade gener, metoden används för att studera genom-mönster som kan visa på anpassning till lokala miljöer hos vissa individer och grupper. Slutligen fokuserar jag på hur skillnader i livsstil har påverkat upplevelsen av smak i grupper med olika livsstil.

Den moderna människan har sitt ursprung i Afrika för kanske 150 000 till 200 000 år sedan. Den här tiden sammanfaller med en istid, för mellan 190 000 och 130 000 år sedan, då även lämpliga habitat för människor minskade kraftigt. Många observationer tyder på en kraftig minskning av populationsstorlek hos tidiga människor vilket skulle kunna spela en viktig roll för den moderna människans uppkomst. Jag använde en Bayesiansk metodik tillsammans med simuleringar av populationsgenetiska modeller för att utvärdera sannolikheten att det fanns en populationsflaskhals för 130 000 till 190 000 år sedan. Jag analyserade genetiska sekvensdata från tre afrikanska populationer, som inte är påverkade av den kraftiga flaskhalsen i samband med migrationen från Afrika för ca 70 000 år sedan, och fann att genetiska data stöder en modell utan flaskhals för 130 000 till 190 000 år sedan.

Efter nästan 100 000 år på den afrikanska kontinenten började människor bosätta sig också utanför Afrika för ca 70 000 år sedan. Idag har människor anpassat sig till nästan alla landmiljöer på planeten. Flera exempel på lokal anpassning är kända från populationer runt om i världen. En av de mest studerade är anpassningen till att kunna bryta ner mjölksocker även i vuxen ålder, en egenskap vanlig i populationer där mejeriprodukter ingår i kosten. Anpassningen har skett oberoende i Östafrika, Mellanöstern och Europa. En annan flitigt studerad anpassning är den till hög höjd och kopplat till lågt lufttryck. Anpassning till hög höjd har också skett flera gånger hos människor som lever i Himalaya, på den Östafrikanska platån och i Anderna. Folk från dessa extrema områden har ofta alleler som skyddar dem mot graviditetskomplikationer samt ger dem en högre syresättning i blodet. Fler exempel på lokal anpassning i människor finns, men de utgör fortfarande bara en liten del jämfört med all anpassning människan som art har

genomgått. En sätt för att bättre förstå adaptation i människor, är att förbättra metoderna för att upptäcka potentiella adaptiva händelser hos människor.

För att studera signalerna i genomet från positiv selektion utvecklade jag en metod baserad på haplotyp-strukturer. Jag använde först simulerade data för att visa att metoden kan detektera selektion, därefter testade jag den på genotypdata för att se om metoden hittar kända regioner i genomet som har varit påverkade av selektion. Jag visar att vi kan detektera en signal runt den genomiska region som är involverad i nedbrytning av mjölksocker hos vuxna både i östafrikaner och européer, även genen som påverkar hårstruktur ger utslag med min metod hos östasiatiska grupper. Slutligen fann jag tidigare icke beskrivna kandidat-gen regioner för anpassning till lokala miljöer. Särskilt en gen-region som påvisats var associerad med längd i andra populationer visade tydliga tecken på selektion hos det förhållandevis långa Maasai-folket i Kenya.

En bidragande orsak till att vi inte vet så mycket om anpassning hos människor, är att inte alla populationer har varit studerade med samma intresse i allmänhet och för afrikanska populationer i synnerhet trots att mänskligheten har sitt ursprung i Afrika. Därför undersökte jag genom-signaler på selektion i afrikanska populationer och visar att patogener och kost har varit viktiga drivkrafter för selektion hos människan. Till exempel fann jag en kandidat-gen region hos jägare-samlare folk i södra Afrika som förmodligen skyddar mot mjältbrand, vilket är en vanligt förekommande sjukdom i sydvästra Afrika.

En aspekt som tros ha spelat en stor roll i evolutionen av homininer är dieten, vilket också mina resultat från analys av genomiska signaler av adaptation tyder på. Dieten hos människor idag skiljer sig mycket jämfört med den hos tidiga homininer, och tre steg tros ha spelat en särskilt viktig roll i utvecklingen av kosten hos homininer: en drastisk ökning av kött i kosten, användning av eld och (långt senare) jordbruket.

Nästan alla *Australopithecus* grupper, bland de äldsta hominin grupperna, hade en vegetarisk diet. Mängden av kött i kosten har ökat i *Homo* grupper i jämförelse med *Australopithecus*. Användningen av eld började för minst 800 000 år sedan, med kontrollerad användning för runt 450 000 år sen och ökande användning för ca 130 000 år sedan. Användning av eld har förändrat både diet och matsmältning i homininer. Betydligt senare har övergången till jordbruk och pastoralism haft en tydlig påverkan på vissa gruppers diet. Människor i jordbrukssamhällen har traditionellt en mindre varierad kost och äter mindre kött än jägare-samlare och herdesamhällen har traditionellt en diet med tydliga inslag av mjölkprodukter och kött.

Ett viktigt sinne för att bedöma kosten är smak. Smaksinnet gör det möjligt att uppfatta den kemiska sammansättningen av föda, att upptäcka gifter (bittert), identifiera skämd eller omogen föda (surt) men också identifiera energirika näringsämnen (sött, umami), mineral (salt) och vitaminer (t.ex. vitamin C smakar surt). Det finns också belägg för att smaksinnet har evolverat i samklang med den lokala miljön och med livsstil. Jag har studerat smakupplevelse i populationer med olika livsstil (jägare-samlare, bönder, och nomadiska herdefolk). De studerade populationerna kommer från två regioner: Centralafrika och Centralasien. I Centralafrika har jag jobbat med jägare-samlare och deras grannpopulation som är jordbrukare. I Centralasien jobbade jag med herderfolk och traditionella bondefolk.

Baserat på fenotypiska observationer visar jag först att smak har evolverat tillsammans med livsstil. Jägare-samlare är i allmänhet mindre känsliga för smaker än vad deras jordbrukande grannar är. Detta kan förklaras som en anpassning till att utnyttja den större del av födoresurser i en väldigt variationsrik miljö. Bönder å andra sidan, har kanske anpassats (i medeltal) till att vara mer känsliga för smaker för att kunna välja mer energirika grödor och undvika skämd mat. Skillnaderna i smakupplevelse är mindre mellan herdefolk och deras jordbrukande grannar i Centralasien. Bönder är mer känsliga för sött, vilket kan relateras till en intensiv selektion av frukter i Centralasien, medan herdefolk är mer känsliga för salt, vilket kanske hänger samman med en proteinrik kost och användningen av salt för att konservera kött. Dessa fenotypiska skillnader kan förklaras med genetisk avvikelse i smak-involverad gener. Med en association analys mellan fenotypisk variation och genotyp data, visar jag också att variation i smakupplevelse inte bara innefattar variation hos smakreceptor-gener utan även andra typer av gener.

Jag har under mitt doktorprojekt, med hjälp av analyser av genetiska och fenotypiska data, ökat kunskapen om den moderna människan. Dels kunskapen om den tidiga demografien men också vilka gener som har påverkats av selektion. I framtiden kommer storskaliga genetisk data att bli ännu enklare att generera. Troligtvis kommer den begränsande faktorn för att studera människans evolution att vara metoderna för att analysera stora datamängder och karaktärisering av fenotypisk variation. De närmsta årtiondena kommer att vara mycket spännande för att studera människans evolution och genetik och flera utmaningar återstår att angripa.

Résumé en Français

Cette thèse présente plusieurs approches visant à étudier l'évolution de l'Homme. Nous nous sommes, en particulier, intéressés à ses origines et aux signaux génétiques révélant des processus d'adaptation à l'environnement. Plus précisément, par l'utilisation de scans génomiques de signatures d'événements de sélection, nous avons identifié diverses régions potentiellement sélectionnées dans des populations différant par leur mode de vie et leur environnement. L'adaptation au régime alimentaire a été plus amplement traitée, avec l'étude de perceptions gustatives dans des populations différant par leur mode de subsistance.

Origines de l'Homme Moderne

L'Homme serait apparu en Afrique il y a environ 200,000 ans. Les premiers fossiles d'hommes anatomiquement modernes coïncident avec une période glaciaire désignée par MIS6 (entre 190,000 et 130,000 ans) (White et al. 2003; McDougall et al. 2005). Durant cette période, les paysages hors de la calotte glaciaire étaient en grande partie des déserts, des savanes et des forêts sèches, les forêts tropicales étant moins importantes qu'actuellement (Van Andel and Tzedakis 1996; Dupont et al. 2000). Les environnements ouverts protégeaient peu les populations humaines des prédateurs et pour cette raison il a été suggéré que la période glaciaire MIS6 a été une période de réduction importante de la taille des populations humaines ce qui aurait conduit à la spéciation de l'espèce *Homo sapiens*.

Deux scénarios ont été proposés pour expliquer l'impact du climat sur les populations humaines durant le MIS6 (Lahr and Foley 1998). D'un côté l'hypothèse de fragmentation suggère qu'en raison d'une forte aridité durant le MIS6 en Afrique, les environnements propices aux humains étaient fragmentés et pour cette raison les populations humaines consistaient en petits groupes isolés géographiquement sans échange de matériel génétique. A la fin de l'ère glaciaire, ces populations auraient été à nouveau en contact du fait de l'élargissement des environnements habitables par l'Homme. D'un autre côté, le scénario de l'effet fondateur propose qu'un seul groupe humain parmi tous les hominins de l'époque soit à l'origine de tous les hommes modernes actuels.

En utilisant une approche bayésienne sur des données de séquence (Article I), nous avons testé les probabilités relatives de ces deux scénarios ainsi que d'un modèle sans réduction de population. Nos résultats montrent que les données génétiques ne suggèrent pas de réduction de population de plus de trois fois la taille initiale de la population à cette période. Nous proposons donc un scénario selon lequel l'ère glaciaire n'affecte pas ou peu la taille des populations humaines durant le MIS6.

Peuplement du globe et structure de population

Après environ 100,000 ans d'évolution sur le continent africain, l'Homme a colonisé le reste du monde. Les restes fossiles indiquent en effet que l'Homme a atteint au plus tard la péninsule arabique il y a 61,000 ans, l'Europe il y a 41-46,000 ans, l'Asie de l'Est il y a 40,000 ans et l'Australie il y a 50,000 ans (Pour une revue, voir Jobling et al. 2013). Plusieurs routes de migrations sont encore aujourd'hui discutées pour modéliser les migrations humaines ancestrales.

C'est sur le continent africain que l'Homme a passé le plus de temps. L'origine africaine de l'Homme se traduit aujourd'hui par une grande diversité culturelle et linguistique en Afrique. Cette diversité s'observe aussi au niveau génétique. Les populations africaines présentent les plus forts taux de diversité génétique ainsi qu'une forte structuration entre groupes (Campbell and Tishkoff 2010). Le Néolithique en Afrique a fortement influencé la répartition des populations africaines avec l'expansion Bantoue liée à la révolution agricole. Des lignées humaines plus divergentes sont encore représentées aujourd'hui notamment par les populations Pygmées d'Afrique Centrale et Khoe-San d'Afrique Australe. Les groupes Pygmées, initialement décrits pour leur petite stature, présentant pour la plus part un mode de vie chasseur-cueilleur. Plusieurs études ont montré leur origine commune ancienne, malgré une grande diversité de culture et de langue (Patin et al. 2009; Verdu et al. 2009). En Afrique Australe, les peuples Khoe et San parlant tous des langues à clics et des modes de vie nomades, pastoral ou chasseur-cueilleur, ont aussi une origine commune ancienne (Schlebusch et al. 2012).

Une autre région considérée dans cette thèse, l'Asie Centrale, présente une grande diversité culturelle et linguistique. Les premiers peuplements de l'Asie Centrale sont peu clairs, malgré la position clé de l'Asie Centrale dans les premières migrations hors d'Afrique vers l'Est. Néanmoins, les peuples actuels d'Asie Centrale ont des origines fortement influencées par des migrations venues de l'Est et de l'Ouest. Deux groupes linguistiques sont observés en Asie Centrale : indo-iranien et turcique. Les groupes parlants

des langues turciques seraient descendants d'empires mongols nomades venant de l'Est, alors que les locuteurs de langues indo-iraniennes, seraient descendant des Scythes venus de Perse. En Asie Centrale il y a une grande diversité culturelle et de modes de subsistance entre les populations. Ainsi, les peuples de langue turcique sont traditionnellement éleveurs nomades, tandis que les locuteurs de langues perses, sont traditionnellement agriculteurs sédentaires.

L'étude des événements de sélection qui a pu affecter les populations Africaines et d'Asie Centrale permet de dresser un tableau de ce qui a pu forger l'adaptation des grands groupes à leur environnement et leur mode de vie.

Etude de l'adaptation locale

Une partie de cette thèse a été consacrée au développement d'une méthode utilisant les structures haplotypiques (Article II) pour la détection d'adaptations locales. Un haplotype est une combinaison de variants sur une même séquence d'ADN. On dit qu'un haplotype est privé à une population, lorsqu'il n'est présent que dans celle-ci. Les haplotypes privés d'une population sont dans la majeure partie apparus après la divergence entre cette population et les autres. Lorsqu'un haplotype privé atteint une fréquence assez forte, cela révèle potentiellement un événement de sélection dirigé sur cet haplotype. Pour cette raison, la fréquence maximale des haplotypes privés (MFPH) en une région du génome peut révéler de l'adaptation locale dans une population.

En utilisant des données simulées ainsi que des données réelles de la base de données HapMap III, nous avons pu montrer que la statistique MFPH permet en effet de détecter des signaux de sélection connus. Cette statistique permet aussi d'identifier des régions potentiellement sélectionnées et non encore décrites. C'est le cas notamment chez les Maasai, où MFPH permet de détecter un fort signal dans une région associée à la stature (Jarvis et al. 2012).

Nous avons aussi procédé à l'analyse de données de génomes entiers de populations Africaines dans le but d'étudier les patrons d'adaptation dans les populations Africaines. Nous avons dans ce cas utilisé les statistiques iHS and XP-EHH (Voight et al. 2006; Sabeti et al. 2007) destinées à la détection d'adaptation récente. Globalement, les résultats montrent l'importante influence du régime alimentaire et des pathogènes dans les phénomènes d'adaptation locale dans la plupart des populations étudiées (Article III).

Goût et mode de vie

Le régime alimentaire des individus du genre *Homo* a beaucoup varié au cours du temps. La plus part des résultats concernant les groupes d'Australopithèques révèlent qu'ils étaient majoritairement herbivores et frugivores, même si certains des groupes *Australopithecus* les plus graciles consommaient probablement de la viande crue (Kay 1985; Sponheimer et Lee-Thorp 1999). Les premiers individus des groupes *Homo*, seraient omnivores avec un apport en viande probablement issue de carcasses et de la chasse (Bunn et al. 1986; Blumenschine et al. 1987; Domínguez-Rodrigo 2002; Domínguez-Rodrigo et Pickering 2003). Avec le temps, les fossiles du genre *Homo* sont associés avec une amélioration des outils traduisant une plus forte part de viande dans le régime alimentaire grâce à une amélioration des techniques de chasse (Binford 1984). La cuisson des aliments a provoqué un grand changement dans l'alimentation, facilitant leur digestion et assimilation. L'usage du feu a également contribué à la protection contre les pathogènes et à l'élimination de certaines toxines.

La transition à l'agriculture, apparue relativement récemment et indépendamment dans plusieurs régions (Jobling et al. 2013), a aussi beaucoup modifié le mode de vie des populations qui la pratiquent. L'agriculture est associée à de la sédentarisation, une augmentation de la population ainsi qu'une importante réduction de la diversité du régime alimentaire. La transition vers l'agriculture s'est associée de carences et de maladies facilitées par l'augmentation de la population (Armelagos et Cohen 1984). Un autre mode de vie apparu récemment est le pastoralisme, spécialisé dans l'élevage et souvent nomade. Les origines du pastoralisme sont incertaines. Le pastoralisme se serait soit développé à partir d'un mode de vie chasseur-cueilleur, soit à partir d'une société agro-pastorale intensifiée conduisant à la spécialisation de la culture d'une part et de l'élevage d'autre part. Le régime alimentaire lié à un mode de vie pastoral s'accompagne d'une plus forte consommation de viande et/ou de produits laitiers.

Le goût est un sens essentiel dans le régime alimentaire, car il permet de discriminer les aliments potentiellement toxiques et évaluer les propriétés nutritives des aliments comestibles. Les cinq goûts perçus par l'Homme, l'amer, le sucré, l'umami, le salé et l'acide, révèlent chacun divers aspects des aliments. Ainsi la perception de l'amertume est corrélée avec la présence de toxines dans les aliments, tandis que le sucré ou l'umami traduisent souvent une forte valeur énergétique liée à la présence de sucres ou de protéines. Le salé permet la détection de minéraux dans les aliments, essentiels à la balance des liquides corporels. L'acidité peut être perçue à la fois négativement lorsqu'elle traduit des aliments périmés ou bénéfique dans

le cas de boissons. En effet certaines vitamines et nutriments ont un goût acide, comme c'est le cas de la vitamine C.

Les mécanismes génétiques et physiologiques de la perception des goûts sont assez bien documentés. Les goûts sont perçus au niveau de la cavité buccale principalement sur la langue au niveau des papilles gustatives. Les papilles gustatives sont composées de plusieurs cellules réceptrices exprimant des récepteurs sensibles aux composés chimiques des aliments. Les gènes codants pour les récepteurs gustatifs des différents goûts ainsi que les gènes impliqués dans la cascade de signalisation sont bien décrits.

Ceci permet de conduire des études d'association pour déterminer les loci à la base des variations de perceptions gustatives. La sensibilité aux goûts peut s'évaluer par des tests psycho-physiologiques consistant en la présentation aux participants de concentrations croissantes de solutions aqueuses contenant les composés simples. Le seuil de perception d'un goût est caractérisé par la concentration la plus faible à partir de laquelle un individu est apte à caractériser le goût de la substance diluée dans l'eau.

Nous avons comparés pour leur sensibilité aux cinq goûts quatre populations différant par leur mode de subsistance dans deux régions : L'Afrique Centrale et l'Asie Centrale. En Afrique Centrale, au Cameroun, nous avons comparé les perceptions gustatives des chasseur-cueilleurs Bakas avec celle des agriculteurs Nzimés. En Asie Centrale, nous avons comparé en Ouzbékistan des participants originaires du Kazakhstan, traditionnellement éleveur-nomades et des participants originaires de la partie Tadjikophone du pays, traditionnellement agriculteurs (Article IV).

Les phénotypes obtenus révèlent que les chasseur-cueilleurs Bakas sont en moyenne moins sensibles aux goûts que les agriculteurs Nzimés. Cette différence est en parallèle avec une divergence génétique de gènes impliqués dans la signalisation de plusieurs goûts à la fois ainsi qu'une divergence dans les gènes de la morphogénèse de la langue.

En Asie Centrale, les éleveur-nomades et les agriculteurs diffèrent moins sur les perceptions gustatives. Les participants d'origine Kazakhe sont plus sensibles au salé et les participants de la région Tadjikophone, plus sensibles au fructose. Ces différences sont en parallèle avec la divergence génétique des récepteurs impliqués dans la perception du salé et du sucré.

Les résultats d'analyse d'association révèlent que la perception des goûts implique probablement plus de gènes qu'uniquement les récepteurs et gènes de la cascade de signalisation. Ainsi nous trouvons qu'en plus des gènes connus, d'autres structures telles que le développement nerveux, des

neurotransmetteurs ou les structures membranaires de microvilli sont associés avec la perception de certains goûts.

Enfin, les analyses d'enrichissement des signaux de sélection obtenus par scan à F_{ST} , iHS et XP-EHH, montrent plusieurs signaux d'enrichissement en accord avec le mode de subsistance des diverses populations.

Conclusions générales

Les résultats obtenus durant cette thèse permettent d'affiner nos connaissances sur les processus démographiques et adaptatifs qu'ont connus les populations humaines. Nos résultats indiquent qu'il n'y a pas eu de réduction majeure de taille de la population des premiers hommes anatomiquement modernes. Nous montrons également que les pathogènes et le régime alimentaire semblent être des moteurs de l'adaptation dans de nombreuses populations.

Acknowledgments

This Ph.D. project has been a long journey for me, and the road has often been stiff. If I made through all the way, it was because I've been extremely lucky to have amazing colleagues, family and friends! Let me warn you now, this section will be long... Indeed, being part of three labs, also triples the number of persons to thank, and this is a great pleasure for me to do so!

I shall first thank the three persons without whom this would never have happened, my supervisors Mattias Jakobsson, Evelyne Heyer and Michael Blum. You are all both wonderful persons and incredible researchers. I have been really proud of being able to work with you. You provided me the best scientific opportunities I could have imagined and each of you taught me so much scientifically and personally. It was a pleasure for me to work under your supervision. Apart from being the great scientists everyone knows you for, I'd like to thank the persons you are, which made the difference for me during these four years.

Mattias, as a Ph.D student I especially enjoyed your incredible calmness, clarity and rigor that made me feel better so many times when in doubts both regarding work and other things. You have the ability to put clarity and logic in someone's most stressed and confused speech or text. This is really stunning! It makes you an amazing supervisor and team leader. I will never forget that once you said "coding is like meditation, you put on some music and you just go on", you might be the only person on Earth that says that!

Evelyne, I love how inspiring you are and would like to thank you for finding this fascinating Ph.D. project, as well as for the extremely interesting environment you've built in your team. Also, thank you for making me able of doing two wonderful field trips, especially the one with you in Uzbekistan. It has been unforgettable and wonderful experiences. You've always been positive and enthusiastic about my work – at the exact contrary of me – and putting me back on the right way, this really helped me reconsider my data in my most skeptical periods.

Michael, you are a challenging and critical scientist and you provided me a lot of precious reflection and methods that I needed. This was especially kind of you since most of my project was not at the same theoretical level as your work, and I am deeply grateful for that. You are easy to talk to about anything and makes one feel comfortable, that makes the work environment really convivial (in such a geek world that is BCM). Besides from work,

thanks for always having interesting/exciting, stories/gossip, to tell at lunch breaks during warming up your frozen dishes!

Next I want to deeply thank my closest colleague and friend, Per, my fourth supervisor! I really enjoyed working with you, although you put every single thing into question... but your extremely critical mind makes you a high standard scientist, so I can't complain! Hopefully you are also a really fun (and weird) person sharing so many jokes (and stupid videos), making the work environment very... "special"! Thanks for all the effort you put into helping me for writing this thesis, writing the papers and performing the analysis, I definitely would not have made it without you. Thanks also for supporting my "other career", which permitted me to free my mind from work! I hope you and Agnieszka will enjoy my masterpieces for a long time. Who knows, maybe one day they could make you rich... But you will never dare to sale them, will you?! Just one last thing about you: I like telling stories, so I'd like to highlight for the people who weren't there in the beginnings of the Jakobsson Lab, Per is actually the person that said "Coding makes everybody swear!". Obviously, Per and Mattias don't listen to the same type of music, that might explain a few things...

Many thanks to Carina, you've been always so kind in helping, answering my questions and sharing the enormous knowledge you have. Kindness like yours is rare! I really enjoyed working with you and being able to be part of the fascinating research you do. I am sorry I have often been so slow, and not always at the level you deserve.

Less work involved, but as many deep thanks to Lucie, my good colleague, French buddy and especially my good friend! You are always there to talk great science, you are so clever and rigorous! You are also a really fun and wise woman, I enjoy that we can talk about anything, it helped a lot during this thesis! Last, thanks for being so proud of me, I'm so proud of you too!

Now it comes the moment to thank all the people who kindly helped proof-reading and correcting this thesis: Per, Lucie, Alexandra, Nina, Emma, Martin, Daniel and of course my supervisors! You provided a precious help, thank you! Thanks also to all of you that offered their help so kindly and worried about my mental health during the last moments... I really appreciated it!

Field work has been an intense and unforgettable part of this project, and that is why I would like to thank the persons that contributed to make it such a wonderful experience! I will start by all the persons in Paris that helped me before the big jump, by helping me in building up the experiment, questionnaires and transmitting me the precious knowledge and experience they have. So thanks to Serge Bahuchet, for the great discussion about the field and precious advices. Thanks to Claude-Marcel Hladik, Marie-Odile Monneuse, Patrick Pasquet and Shelly Masi for kindly helping me in learning about taste, building up the experiment, transmitting me literature

and so many more interesting discussions! Also thank to Noémie, Priscille and Paul, for sharing with me their experiences about the field in Cameroon. Paul, I would like to thank you especially for a single sentence, you said: “Anyway, whenever you don’t feel comfortable anymore on the field, just take a flight back”, it contributed to eradicate my stress and made everything going so well.

Thanks to people that were with me on the field in Cameroon, especially Alain Froment and Fernando Ramirez-Rozzi, you have been amazing with me in making sure everything would go smooth for me and having interesting discussions. Thanks to the members of the IRD-Yaoundé and the nuns for providing all the practical help to perform the field work in Cameroon. Deep thanks to my translators and friends Thierry, Lily and Ngala, you did an amazing work. Talking and spending time with you outside of the experiments was so enriching and delightful. I wish I’ll come back one day to sit down and chat with you! Thanks to the lovely movie makers Laurent and Steve, you made me an actress! Oh my God! Last, thanks to all the persons that showed hospitality and friendship to me, thanks for all the delightful chitchats!

Thanks to the great field-team in Uzbekistan, special thanks to the big boss Tatyana Hegay who made it possible, you are such a strong and intelligent woman! Thanks to “Tatyana’s slaves” Nargis, Farik, Shukrat, Gala and Ania, it was so great to work and spend time with you. Thanks for taking so good care of me, it was way too much! Thanks also to my little translator Nilufar in Bukhara. I’d like to come in person to Tashkent and celebrate the end of this project with you, uzbek style! Thanks to the spectacular Philippe, I love talking with (but especially listening to) you about linguistics, politics etc., you have so much knowledge and transmit it in such a fun and nice way, I wish I’ll still be able to learn from you. Many thanks to Eric, you are so motivated, fun and full of energy! I enjoyed the many interesting discussions we had, and you really made the field atmosphere a lot of fun.

Of course, thanks to all the participants of the study, who have given their time and interest to the study. I hope you know that I’m grateful!

Thanks to all the great members of the Evolutionary Biology department in Uppsala. First, thanks to Hans Ellegren for making the lab being a great environment. Thanks to the amazing alumni and current members of the Jakobsson Lab. So a huge grouped-thanks to the current members Carina, Per, Helena, Emma, Torsten, Olaf, Federico, Nina, TJ, Gwenna, Hanna, Luciana, Mario, Alex, Arielle, Ayca and Magdalena, although I grouped you, let me say that each of you make the work environment exciting and fun. Of course, for the same reasons, thanks to the past members Sen, Pontus and Hiba! And for those of you who are Ph.D. students: luck and success in your own projects!!

Thanks to the nice office mates (the ones not previously mentioned) Robb, Sergio, Constance and Eva, it's been super great sitting with you!

Many thanks to my lovely colleague, friend and flat mate Federico! You are a really great and supportive friend! Thanks for all the scientific help, the long discussions, the support and the fun we had!

Warm thanks also to Bart, Matthias, Verena, Alex Suh, Paulina, Vera, Claire, Jochen, Niclas, Ludo, Severin, Venkat, Hanna Johannesson, Cécile, Jelmer and so many others in the lab! I really enjoy all the nice time we spent at EBC!

Many thanks to Martin Lascoux for kindly accepting to co-supervise me, proof-reading my thesis and especially being such a delightful person to talk to from science to food. Thanks to the rest of the French members from the Plant ecology department, Sylvain Glémin, Cécile Meunier and Michel Thomann for great discussions, mostly in French...!

The Eco-anthropology and Ethnology department in Paris is a very special Lab full of really exciting people. First, thanks to Serge Bahuchet, the director of the department for maintaining such a nice work environment. I enjoy how close to the students you remain although you have so much responsibilities and work, I think you are right to do so. Thanks to the Evolutionary Anthropology, current and alumni, members: Paul, Laure, Noémie, Rihlat, Fred, Bruno, Flora, Nina, Julie, Elise, Goki, Cristophe, Romain, Raphaëlle, Jean-Tristan, Friso, Samuel, Priscille, Myriam, Sophie, Philipp and Franz. You all make this team fascinating! Lab work is definitely not an area in which I feel confident, hopefully Myriam and Sophie provided such a nice help with this. So, a huge thank for reinsuring and helping me when I was freaking out about any single tube or drop of water! A big thank also to Taoues, Farida and Sylvie. Just as lab work, administration is by default a nightmare for me and you cooled me down by being always so efficient with such a good mood! It was so nice for me in the stress of preparing field works! All the best to the Eco-anthropology and Ethnobiology lab in the brand new Musée de l'Homme. This lab is simply amazing!

Last, many thanks to the TIMC-IMAG Lab in Grenoble, especially the BCM team members. I have proudly been the only girl in the BCM team for a while, and although it's been way nicer with you Laure, it was most of the time fun. Thanks to Olivier François, for providing many good advices and asking difficult questions. Thanks to Laure for being such a nice colleague and great friend to talk to. Thanks to Nicolas D-F and Eric Frichot for the coffee room-service, but mostly for the precious help with my analysis! You have stunning skills! It was really nice working with you! Thanks also to Sean, Eric Fanchon, Nicolas TM, Daniel, Nicolas M, Thomas, Om and Laurent.

Big thanks also to Arnaud, my great office neighbor, it was always much better when your door was open! Thanks for the support, the nice advices,

the hosting, the coffee at the Absolu and... the chocolate!! Special thanks to Laure again for putting together a great basket team, and thanks to all its members for all the nice games! Grenoble, cannot be mentioned without climbing, so thanks to all the people that I've been lucky to climb with! It was a lot of fun!

Now I would like to thank all my close friends that supported me during the whole time and endured quite a lot of depressed conversions...

Amandine, même si tu as été loin la plus part du temps, tu as toujours été là pour m'écouter, me soutenir et me donner de bon conseils ! De même, merci aussi à Thu-linh, pour croire autant en moi malgré la distance.

Un grand merci à Alice, Dalia et Marion pour avoir toujours été là pour moi, une équipe de choc dans les moments les plus durs ! Merci aussi tout spécialement pour les supers soirées, week-ends et vacances passées ensemble !! Vous ne pouvez pas savoir le bien que ça m'a fait durant cette thèse ! Vivement le prochain voyage ! Bon, on va où ?! Et aussi merci à toi Alice pour cette super année de coloc ! C'était précieux !

Merci à tous mes amis de l'ENS Lyon, surtout à "mes ptits amis!" : Clémentine, Amélie, Fabian, Mathias, Repi, Marion et Morgane. Merci pour tous ces bons moments. Bravo pour ceux qui ont déjà fini la thèse ! Courage pour ceux qui y sont encore, ça va le faire à l'aise ! Et Bravo à Amélie qui dans tout ça nous écoute tous et nous soutient ! Heureusement y'a vraiment pas que la thèse dans la vie, alors merci pour tous ces bons moments, ces supers voyages, les fêtes, les longues discussions et tant d'autres choses !! Vous êtes géniaux !

Merci à mes super colocs Inès, Laurence et Agnès, c'était le bon vieux temps ! Inès, j'ai particulièrement apprécié nos super week-ends quand tu étais en Suède ! C'était trop bien, surtout celui pour l'anniversaire de Laurence !! C'est vraiment super de parler avec toi de tout, et comme il s'agit ici de la thèse surtout merci pour l'écoute et les supers conseils ! Vivement le prochain week-end colocs !

Je vous aime les amis!!!

Tusen tack till Daniel, du har lyssnat på mig, gav mig mycket stöd och lagat mig underbart mat medan jag jobbade! Tack för att du har vart där för att chatta varge gång jag skulle vara tråkad, vilket är rätt mycket... Kanske hade du inte räknat med det?! Tack för att göra mig mindre tråkad med alla dina R-problemer. Om det inte är redan gjort, snart kommer du att vara mycket bättre än jag på det! Och sist, tack en gång till för hjälpen med den svensk sammanfattning och alla andra insikter! Du är en stark och klok människa, med en kritisk och rigorös sinne, du kommer att klara din egen Ph.D. utmärkt! Jag är stolt över dig!

Merci à tous mes amis de Grenoble : Alix, Johanna, Baptiste B, Baptiste J, Carole, Sarah, Laurent, Guillaume, Damien, Maxime, Alex, Pamela et Youri. Vous me manquez, vivement une petite sortie dans les montagnes !

Un profond merci à Arnaud qui m'a soutenue et supportée pendant si longtemps, toi seul sait à quel point la tâche a pu être dure !! Je ne pense pas que j'y serais arrivée sans ton soutien. Je ne pense pas non plus avoir été assez reconnaissante sur le moment, et c'est pour ça que je voudrais l'exprimer ici : merci du fond du cœur !! Plein de bonheur et de réussite dans ta vie !

On en vient à la famille... Le meilleur pour la fin! Merci tout d'abord à mes oncles et tantes qui ont été là durant cette thèse : Rémy et Hélène pour m'avoir accueillie chez eux avec la plus grande gentillesse et su m'écouter longuement. Merci à mon super parrain Paul, toujours présent pour aider.

Merci à mes trois super grandes sœurs ! Marianne, tu as toujours été à l'écoute et pleine d'empathie pour moi, comme tu es toi-même passée par là. Ne te fais plus de mouron : regarde j'y suis arrivée !! Merci à Louise, tu es toujours là à donner tout l'aide que tu peux dans toutes les situations compliquées dans lesquelles je vais me fourrer, c'est absolument incroyable ! Enfin Charlotte, j'aime comme tu partages ton expérience et donnes de bon conseils d'ainée que tu as le sens pour donner avec la plus grande délicatesse, j'ai toujours été ta plus grande fan ! Vous êtes toutes les trois des femmes géniales au grand cœur et je vous aime tellement !!

Mes sœurs ne seraient sûrement pas aussi géniales si elles n'avaient pas été élevées par deux personnes aussi adorables que mes parents ! Donc enfin le plus grand des mercis à mes parents Odile et Johannes. Papa, tu m'as transmis le goût pour la curiosité et la réflexion scientifique sans quoi je n'en serais pas là. C'est probablement parce que tu as eu la patience de répondre au mieux à toutes mes questions étant petite, la plus part n'aillant pas de sens ! Tu es aussi une personne au grand cœur qui n'hésite jamais à tout donner pour soutenir ses filles. Maman, merci pour m'avoir transmis au moins un peu de ta capacité plus sensible de réflexion et ta fine analyse. Tu sais énormément de choses que tu aimes partager en parlant beaucoup, et c'est un vrai plaisir de t'écouter et prendre part à la conversation. Tu es bien sûr la preuve vivante que l'amour d'une mère est inégalable, et je ne pourrai jamais assez te remercier pour ça !! MERCI!! Je vous aime de tout mon coeur.

My God this is it! I'm sure I forgot someone, so if it is you, remember how grateful I am to you in my heart! So: one last **huge group-thanks to e-v-e-r-y-o-n-e!!** I wish you all a lot of success, love and happiness in your lives.

References

- Adler E, Hoon MA, Mueller KL, Chandrashekar J, Ryba NJP, Zuker CS. 2000. A Novel Family of Mammalian Taste Receptors. *Cell* 100:693–702.
- Aiello LC, Wheeler P. 1995. The expensive-tissue hypothesis: the brain and the digestive system in human and primate evolution. *Current anthropology*:199–221.
- Alkorta-Aranburu G, Beall CM, Witonsky DB, Gebremedhin A, Pritchard JK, Di Rienzo A. 2012. The genetic architecture of adaptations to high altitude in Ethiopia.
- Allaby RG, Fuller DQ, Brown TA. 2008. The genetic expectations of a protracted model for the origins of domesticated crops. *PNAS* 105:13982–13986.
- Van Andel TH, Tzedakis PC. 1996. Palaeolithic landscapes of Europe and environs, 150,000-25,000 years ago: an overview. *Quaternary Science Reviews* 15:481–500.
- Anthony DW, Brown DR. 1991. The origins of horseback riding. *Antiquity* 65:22–38.
- Armelagos GJ, Cohen MN. 1984. *Paleopathology at the Origins of Agriculture*. Academic Press
- Armitage SJ, Jasim SA, Marks AE, Parker AG, Usik VI, Uerpmann H-P. 2011. The Southern Route “Out of Africa”: Evidence for an Early Expansion of Modern Humans into Arabia. *Science* 331:453–456.
- Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, et al. 2000. Gene Ontology: tool for the unification of biology. *Nat Genet* 25:25–29.
- Avenet P, Lindemann B. 1988. Amiloride-blockable sodium currents in isolated taste receptor cells. *The Journal of membrane biology* 105:245–255.
- Bahuchet S. 1993. History of the inhabitants of the central African rain forest: perspectives from comparative linguistics. *Tropical forests, people and food. Biocultural interactions and applications to development*:37–54.
- Bahuchet S. 2012. Changing Language, Remaining Pygmy. *Human Biology* 84:11–43.
- Baldwin MW, Toda Y, Nakagita T, O’Connell MJ, Klasing KC, Misaka T, Edwards SV, Liberles SD. 2014. Evolution of sweet taste perception in hummingbirds by transformation of the ancestral umami receptor. *Science* 345:929–933.
- Barham L, Mitchell P. 2008. *The first Africans: African archaeology from the earliest toolmakers to most recent foragers*. Cambridge University Press
- Barnard A. 1992. *Hunters and herders of southern Africa: a comparative ethnography of the Khoisan peoples*. Cambridge University Press
- Barreiro LB, Laval G, Quach H, Patin E, Quintana-Murci L. 2008. Natural selection has driven population differentiation in modern humans. *Nat Genet* 40:340–345.
- Beall CM. 2007. Two routes to functional adaptation: Tibetan and Andean high-altitude natives. *PNAS* 104:8655–8660.

- Beaumont MA. 2008. Joint determination of topology, divergence time, and immigration in population trees. Renfrew C Matsumura S, Forster P, editor, *Simulation, Genetics and Human Prehistory*, McDonald Institute Monographs:134–1541.
- Beaumont MA, Zhang W, Balding DJ. 2002. Approximate Bayesian computation in population genetics. *Genetics* 162:2025–2035.
- Beavon K, Beavon M. 1996. *Lexique kônzime-français*.
- Behar DM, van Oven M, Rosset S, Metspalu M, Loogväli E-L, Silva NM, Kivisild T, Torroni A, Villems R. 2012. A “Copernican” reassessment of the human mitochondrial DNA tree from its root. *The American Journal of Human Genetics* 90:675–684.
- Behar DM, Villems R, Soodyall H, Blue-Smith J, Pereira L, Metspalu E, Scozzari R, Makkani H, Tzur S, Comas D. 2008. The dawn of human matrilineal diversity. *The American Journal of Human Genetics* 82:1130–1140.
- Behrens M, Bartelt J, Reichling C, Winnig M, Kuhn C, Meyerhof W. 2006. Members of RTP and REEP Gene Families Influence Functional Bitter Taste Receptor Expression. *J. Biol. Chem.* 281:20650–20659.
- Binford LR. 1984. Faunal remains from Klasies River mouth.
- Birney E, Stamatoyannopoulos JA, Dutta A, Guigó R, Gingeras TR, Margulies EH, Weng Z, Snyder M, Dermitzakis ET, Thurman RE. 2007. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* 447:799–816.
- Blench R. 2006. Archaeology, language, and the African past. Rowman Altamira
- Blumenschine RJ, Bunn HT, Geist V, Ikawa-Smith F, Marean CW, Payne AG, Tooby J, van der Merwe NJ. 1987. Characteristics of an early hominid scavenging niche [and comments and reply]. *Current anthropology*:383–407.
- Blum MGB, Jakobsson M. 2011. Deep Divergences of Human Gene Trees and Models of Human Origins. *Mol Biol Evol* 28:889–898.
- Blum MG, François O. 2010. Non-linear regression models for Approximate Bayesian Computation. *Statistics and Computing* 20:63–73.
- Breslin PAS, Spector AC. 2008. Mammalian taste perception. *Current Biology* 18:R148–R155.
- Brisson R. 1984. *Lexique français-baka*. R. Brisson
- Brutsaert TD. 2007. Population genetic aspects and phenotypic plasticity of ventilatory responses in high altitude natives. *Respiratory Physiology & Neurobiology* 158:151–160.
- Bunn HT, Kroll EM, Ambrose SH, Behrensmeier AK, Binford LR, Blumenschine RJ, Klein RG, McHenry HM, O’Brien CJ, Wymer JJ. 1986. Systematic Butchery by Plio/Pleistocene Hominids at Olduvai Gorge, Tanzania [and Comments and Reply]. *Current Anthropology*:431–452.
- Campbell MC, Ranciaro A, Froment A, Hirbo J, Omar S, Bodo J-M, Nyambo T, Lema G, Zinshteyn D, Drayna D, et al. 2011. Evolution of Functionally Diverse Alleles Associated with PTC Bitter Taste Sensitivity in Africa. *Mol Biol Evol*:msr293.
- Campbell MC, Tishkoff SA. 2010. The Evolution of Human Genetic and Phenotypic Variation in Africa. *Current Biology* 20:R166–R173.
- Cann RL, Stoneking M, Wilson AC. 1987. Mitochondrial DNA and human evolution. *Nature* 325:31–36.
- Castañeda IS, Mulitza S, Schefuß E, Santos RAL dos, Damsté JSS, Schouten S. 2009. Wet phases in the Sahara/Sahel region and human migration patterns in North Africa. *Proceedings of the National Academy of Sciences* 106:20159–20163.

- Chaix R, Austerlitz F, Hegay T, Quintana-Murci L, Heyer E. 2008. Genetic traces of east-to-west human expansion waves in Eurasia. *American journal of physical anthropology* 136:309–317.
- Chandrashekar J, Hoon MA, Ryba NJP, Zuker CS. 2006. The receptors and cells for mammalian taste. *Nature* 444:288–294.
- Chandrashekar J, Mueller KL, Hoon MA, Adler E, Feng L, Guo W, Zuker CS, Ryba NJP. 2000. T2Rs Function as Bitter Taste Receptors. *Cell* 100:703–711.
- Comas D, Calafell F, Mateu E, Pérez-Lezaun A, Bosch E, Martínez-Arias R, Clarimon J, Facchini F, Fiori G, Luiselli D. 1998. Trading genes along the silk road: mtDNA sequences and the origin of central Asian populations. *The American Journal of Human Genetics* 63:1824–1838.
- Consortium 1000 Genomes Project. 2010. A map of human genome variation from population-scale sequencing. *Nature* 467:1061–1073.
- Coop G, Pickrell JK, Novembre J, Kudaravalli S, Li J, Absher D, Myers RM, Cavalli-Sforza LL, Feldman MW, Pritchard JK. 2009. The role of geography in human adaptation. *PLoS Genet* 5:e1000500.
- Cordaux R, Weiss G, Saha N, Stoneking M. 2004. The Northeast Indian Passageway: A Barrier or Corridor for Human Migrations? *Mol Biol Evol* 21:1525–1533.
- Csilléry K, Blum MGB, Gaggiotti OE, François O. 2010. Approximate Bayesian Computation (ABC) in practice. *Trends in Ecology & Evolution* 25:410–418.
- Denham TP, Haberle SG, Lentfer C, Fullagar R, Field J, Therin M, Porch N, Winsborough B. 2003. Origins of Agriculture at Kuk Swamp in the Highlands of New Guinea. *Science* 301:189–193.
- Domínguez-Rodrigo M. 2002. Hunting and scavenging by early humans: the state of the debate. *Journal of World Prehistory* 16:1–54.
- Domínguez-Rodrigo M, Pickering TR. 2003. Early hominid hunting and scavenging: a zooarcheological review. *Evolutionary Anthropology: Issues, News, and Reviews* 12:275–282.
- Dong D, Jones G, Zhang S. 2009. Dynamic evolution of bitter taste receptor genes in vertebrates. *BMC Evolutionary Biology* 9:12.
- Drmanac R, Sparks AB, Callow MJ, Halpern AL, Burns NL, Kermani BG, Carnevali P, Nazarenko I, Nilsen GB, Yeung G, et al. 2010. Human Genome Sequencing Using Unchained Base Reads on Self-Assembling DNA Nanoarrays. *Science* 327:78–81.
- Dupont LM, Jahns S, Marret F, Ning S. 2000. Vegetation change in equatorial West Africa: time-slices for the last 150 ka. *Palaeogeography, Palaeoclimatology, Palaeoecology* 155:95–122.
- Enattah NS, Sahi T, Savilahti E, Terwilliger JD, Peltonen L, Järvelä I. 2002. Identification of a variant associated with adult-type hypolactasia. *Nat. Genet.* 30:233–237.
- Fay JC, Wu CI. 2000. Hitchhiking under positive Darwinian selection. *Genetics* 155:1405–1413.
- Fischer A, Gilad Y, Man O, Pääbo S. 2005. Evolution of Bitter Taste Receptors in Humans and Apes. *Mol Biol Evol* 22:432–436.
- Fuller DQ, Willcox G, Allaby RG. 2011. Cultivation and domestication had multiple origins: arguments against the core area hypothesis for the origins of agriculture in the Near East. *World Archaeology* 43:628–652.
- Fu Q, Meyer M, Gao X, Stenzel U, Burbano HA, Kelso J, Pääbo S. 2013. DNA analysis of an early modern human from Tianyuan Cave, China. *PNAS* 110:2223–2227.

- Fu Y-X, Li W-H. 1993. Statistical tests of neutrality of mutations. *Genetics* 133:693–709.
- Gene Ontology Consortium. 2015. Gene ontology consortium: going forward. *Nucleic acids research* 43:D1049–D1056.
- Gimbutas M, Gimbutas MA. 1991. *The civilization of the goddess: the world of Old Europe*. Harper San Francisco
- Gonder MK, Mortensen HM, Reed FA, de Sousa A, Tishkoff SA. 2007. Whole-mtDNA genome sequence analysis of ancient African lineages. *Molecular biology and evolution* 24:757–768.
- Goren-Inbar N, Alperson N, Kislev ME, Simchoni O, Melamed Y, Ben-Nun A, Werker E. 2004. Evidence of hominin control of fire at Gesher Benot Yaaqov, Israel. *Science* 304:725–727.
- Gowlett JA. 2006. The early settlement of northern Europe: fire history in the context of climate change and the social brain. *Comptes Rendus Palevol* 5:299–310.
- Grün R, Stringer C, McDermott F, Nathan R, Porat N, Robertson S, Taylor L, Mortimer G, Eggins S, McCulloch M. 2005. U-series and ESR analyses of bones and teeth relating to the human burials from Skhul. *Journal of Human Evolution* 49:316–334.
- Güldemann T, Stoneking M. 2008. A Historical Appraisal of Clicks: A Linguistic and Genetic Population Perspective. *Annual Review of Anthropology* 37:93–109.
- Gurdasani D, Carstensen T, Tekola-Ayele F, Pagani L, Tachmazidou I, Hatzikotoulas K, Karthikeyan S, Iles L, Pollard MO, Choudhury A, et al. 2015. The African Genome Variation Project shapes medical genetics in Africa. *Nature* 517:327–332.
- Hamblin MT, Di Rienzo A. 2000. Detection of the Signature of Natural Selection in Humans: Evidence from the Duffy Blood Group Locus. *The American Journal of Human Genetics* 66:1669–1679.
- Harris DR, Hillman GC. 2014. *Foraging and Farming: The Evolution of Plant Exploitation*. Routledge
- Hartl DL, Clark AG, Clark AG. 1997. *Principles of population genetics*. Sinauer associates Sunderland
- Heck GL, Mierson S, DeSimone JA. 1984. Salt taste transduction occurs through an amiloride-sensitive sodium transport pathway. *Science* 223:403–405.
- Hellenthal G, Auton A, Falush D. 2008. Inferring Human Colonization History Using a Copying Model. *PLoS Genet* 4:e1000078.
- Henshilwood CS, Errico F d', Yates R, Jacobs Z, Tribolo C, Duller GA, Mercier N, Sealy JC, Valladas H, Watts I. 2002. Emergence of modern human behavior: Middle Stone Age engravings from South Africa. *Science* 295:1278–1280.
- Hetherington R, Reid RG. 2010. *The climate connection: climate change and modern human evolution*. Cambridge University Press
- Hirschhorn JN, Daly MJ. 2005. Genome-wide association studies for common diseases and complex traits. *Nature Reviews Genetics* 6:95–108.
- Hladik CM, Simmen B. 1993. Perception gustative et adaptation à l'environnement nutritionnel des Primates non-humains et des populations humaines. *bmsap* 5:343–354.
- Huang AL, Chen X, Hoon MA, Chandrashekar J, Guo W, Tränkner D, Ryba NJP, Zuker CS. 2006. The cells and logic for mammalian sour taste detection. *Nature* 442:934–938.

- Huang DW, Sherman BT, Lempicki RA. 2009. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucl. Acids Res.* 37:1–13.
- Huang DW, Sherman BT, Tan Q, Kir J, Liu D, Bryant D, Guo Y, Stephens R, Baseler MW, Lane HC, et al. 2007. DAVID Bioinformatics Resources: expanded annotation database and novel algorithms to better extract biology from large gene lists. *Nucl. Acids Res.* 35:W169–W175.
- Hubbard T, Barker D, Birney E, Cameron G, Chen Y, Clark L, Cox T, Cuff J, Curwen V, Down T. 2002. The Ensembl genome database project. *Nucleic acids research* 30:38–41.
- Hudson RR. 2002. Generating samples under a Wright–Fisher neutral model of genetic variation. *Bioinformatics* 18:337–338.
- Ilegems E, Iwatsuki K, Kokrashvili Z, Benard O, Ninomiya Y, Margolskee RF. 2010. REEP2 Enhances Sweet Receptor Function by Recruitment to Lipid Rafts. *J. Neurosci.* 30:13774–13783.
- Ingman M, Kaessmann H, Pääbo S, Gyllensten U. 2000. Mitochondrial genome variation and the origin of modern humans. *Nature* 408:708–713.
- International Consortium for Blood Pressure Genome-Wide Association Studies, Ehret GB, Munroe PB, Rice KM, Bochud M, Johnson AD, Chasman DI, Smith AV, Tobin MD, Verwoert GC, et al. 2011. Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature* 478:103–109.
- Jacobson-Tepfer E. 2008. The emergence of cultures of mobility in the Altai mountains of Mongolia. *The Archeology of Mobility*.
- Jakobsson M, Scholz SW, Scheet P, Gibbs JR, VanLiere JM, Fung H-C, Szpiech ZA, Degnan JH, Wang K, Guerreiro R, et al. 2008. Genotype, haplotype and copy-number variation in worldwide human populations. *Nature* 451:998–1003.
- Jarvis JP, Scheinfeldt LB, Soi S, Lambert C, Omberg L, Ferwerda B, Froment A, Bodo J-M, Beggs W, Hoffman G, et al. 2012. Patterns of Ancestry, Signatures of Natural Selection, and Genetic Association with Stature in Western African Pygmies. *PLoS Genet* 8:e1002641.
- Jobling MA, Tyler-Smith C. 2003. The human Y chromosome: an evolutionary marker comes of age. *Nature Reviews Genetics* 4:598–612.
- Jobling M, Hurles M, Tyler-Smith C. 2013. *Human evolutionary genetics: origins, peoples & disease*. Garland Science
- Jorde LB, Watkins WS, Bamshad MJ, Dixon ME, Ricker CE, Seielstad MT, Batzer MA. 2000. The distribution of human genetic diversity: a comparison of mitochondrial, autosomal, and Y-chromosome data. *The American Journal of Human Genetics* 66:979–988.
- Kaessmann H, Wiebe V, Weiss G, Pääbo S. 2001. Great ape DNA sequences reveal a reduced diversity and an expansion in humans. *Nature genetics* 27:155–156.
- Kanehisa M, Goto S. 2000. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucl. Acids Res.* 28:27–30.
- Karafet T, Xu L, Du R, Wang W, Feng S, Wells RS, Redd AJ, Zegura SL, Hammer MF. 2001. Paternal population history of East Asia: sources, patterns, and microevolutionary processes. *The American Journal of Human Genetics* 69:615–628.
- Kay RF. 1985. Dental Evidence for the Diet of Australopithecus. *Annual Review of Anthropology* 14:315–341.
- Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D. 2002. The human genome browser at UCSC. *Genome research* 12:996–1006.
- Khlobystina. 1973. Origins and development of the first Bronze Age civilization in South Siberia. *Sovetskaya Arkeologiya* 1.

- Kim HL, Ratan A, Perry GH, Montenegro A, Miller W, Schuster SC. 2014. Khoisan hunter-gatherers have been the largest population throughout most of modern-human demographic history. *Nat Commun* [Internet] 5. Available from: http://www.nature.com/ncomms/2014/141204/ncomms6692/full/ncomms6692.html?WT.ec_id=NCOMMS-20141210
- Klein RG. 2009. *The Human Career: Human Biological and Cultural Origins*, Third Edition. University of Chicago Press
- Klein RJ, Zeiss C, Chew EY, Tsai J-Y, Sackler RS, Haynes C, Henning AK, SanGiovanni JP, Mane SM, Mayne ST, et al. 2005. Complement Factor H Polymorphism in Age-Related Macular Degeneration. *Science* 308:385–389.
- Kofler R, Schlötterer C. 2012. Gowinda: unbiased analysis of gene set enrichment for genome-wide association studies. *Bioinformatics* 28:2084–2085.
- Lachance J, Vernot B, Elbers CC, Ferwerda B, Froment A, Bodo J-M, Lema G, Fu W, Nyambo TB, Rebbeck TR, et al. 2012. Evolutionary history and adaptation from high-coverage whole-genome sequences of diverse African hunter-gatherers. *Cell* 150:457–469.
- Lahr MM, Foley RA. 1998. Towards a theory of modern human origins: geography, demography, and diversity in recent human evolution. *Yearbook of physical anthropology* 41:137–176.
- Lamason RL, Mohideen M-AP, Mest JR, Wong AC, Norton HL, Aros MC, Jurynec MJ, Mao X, Humphreville VR, Humbert JE. 2005. SLC24A5, a putative cation exchanger, affects pigmentation in zebrafish and humans. *Science* 310:1782–1786.
- Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, et al. 2001. Initial sequencing and analysis of the human genome. *Nature* 409:860–921.
- Laval G, Patin E, Barreiro LB, Quintana-Murci L. 2010. Formulating a Historical and Demographic Model of Recent Human Evolution Based on Resequencing Data from Noncoding Regions. *PLoS ONE* 5:e10284.
- Lees SH, Bates DG. 1974. The origins of specialized nomadic pastoralism: a systemic model. *American Antiquity*:187–193.
- Letouzey R. 1976. *Contribution de la botanique au problème d'une éventuelle langue pygmée*. Peeters Publishers
- Levine M. 1999. Late prehistoric exploitation of the Eurasian steppe. McDonald Inst of Archeological
- Li H, Durbin R. 2011. Inference of human population history from individual whole-genome sequences. *Nature* 475:493–496.
- Li H, Gu S, Cai X, Speed WC, Pakstis AJ, Golub EI, Kidd JR, Kidd KK. 2008. Ethnic related selection for an ADH Class I variant within East Asia. *PLoS ONE* 3:e1881.
- Li JZ, Absher DM, Tang H, Southwick AM, Casto AM, Ramachandran S, Cann HM, Barsh GS, Feldman M, Cavalli-Sforza LL. 2008. Worldwide human relationships inferred from genome-wide patterns of variation. *science* 319:1100–1104.
- Lindemann B. 1996. Taste reception. *Physiological reviews* 76:719–766.
- Li X, Glaser D, Li W, Johnson WE, O'Brien SJ, Beauchamp GK, Brand JG. 2009. Analyses of the Sweet Receptor Gene (*Tas1r2*) and Preference for Sweet Stimuli in Species of Carnivora. *J Hered* 100:S90–S100.
- Lombard M, Parsons I. 2015. Milk not Meat: The Role of Milk amongst the Khoe Peoples of Southern Africa. *Journal of African Archaeology* [Internet] 13. Available from: http://www.african-archaeology.de/index.php?page_id=154&journal_id=40&pdf_id=311

- Macaulay V, Hill C, Achilli A, Rengo C, Clarke D, Meehan W, Blackburn J, Semino O, Scozzari R, Cruciani F, et al. 2005. Single, Rapid Coastal Settlement of Asia Revealed by Analysis of Complete Mitochondrial Genomes. *Science* 308:1034–1036.
- Marean CW. 2010. Pinnacle Point Cave 13B (Western Cape Province, South Africa) in context: the Cape floral kingdom, shellfish, and modern human origins. *Journal of Human Evolution* 59:425–443.
- Marean CW, Bar-Matthews M, Bernatchez J, Fisher E, Goldberg P, Herries AI, Jacobs Z, Jerardino A, Karkanas P, Minichillo T. 2007. Early human use of marine resources and pigment in South Africa during the Middle Pleistocene. *Nature* 449:905–908.
- Marshall F. 1990. Origins of Specialized Pastoral Production in East Africa. *American Anthropologist* 92:873–894.
- McDougall I, Brown FH, Fleagle JG. 2005. Stratigraphic placement and age of modern humans from Kibish, Ethiopia. *Nature* 433:733–736.
- Mellars P. 2006. Why did modern human populations disperse from Africa ca. 60,000 years ago? A new model. *Proceedings of the National Academy of Sciences* 103:9381–9386.
- Meyerhof W. 2008. Human taste receptors. In: *Expression of Multidisciplinary Flavour Science—Proceedings of the 12th Weurman Symposium*. Interlaken, Switzerland. p. 3–12.
- Meyerhof W, Batram C, Kuhn C, Brockhoff A, Chudoba E, Bufe B, Appendino G, Behrens M. 2010. The Molecular Receptive Ranges of Human TAS2R Bitter Taste Receptors. *Chem. Senses* 35:157–170.
- Miller CT, Beleza S, Pollen AA, Schluter D, Kittles RA, Shriver MD, Kingsley DM. 2007. cis-Regulatory Changes in Kit Ligand Expression and Parallel Evolution of Pigmentation in Sticklebacks and Humans. *Cell* 131:1179–1189.
- Mitchell P. 2002. *The archaeology of southern Africa*. Cambridge University Press
- Mitchell P. 2010. Genetics and southern African prehistory: an archaeological view. *J Anthropol Sci* 88:73–92.
- Moore LG. 2001. Human genetic adaptation to high altitude. *High altitude medicine & biology* 2:257–279.
- Nachman MW, Crowell SL. 2000. Estimate of the Mutation Rate per Nucleotide in Humans. *Genetics* 156:297–304.
- Nei M. 1977. F-statistics and analysis of gene diversity in subdivided populations. *Annals of Human Genetics* 41:225–233.
- Nei M, Kumar S. 2000. *Molecular evolution and phylogenetics*. Oxford University Press
- Nei M, Roychoudhury AK. 1993. Evolutionary relationships of human populations on a global scale. *Molecular biology and evolution* 10:927–943.
- Nelson G, Hoon MA, Chandrashekar J, Zhang Y, Ryba NJP, Zuker CS. 2001. Mammalian Sweet Taste Receptors. *Cell* 106:381–390.
- Nikolsky Y, Bryant J eds. 2009. *PANTHER Pathway: An Ontology-Based Pathway Database Coupled with Data Analysis Tools* - Springer. In: *Methods in Molecular Biology*. Humana Press. Available from: http://link.springer.com/protocol/10.1007%2F978-1-60761-175-2_7#section=82252&page=1
- Norton HL, Kittles RA, Parra E, McKeigue P, Mao X, Cheng K, Canfield VA, Bradley DG, McEvoy B, Shriver MD. 2007. Genetic evidence for the convergent evolution of light skin in Europeans and East Asians. *Mol. Biol. Evol.* 24:710–722.

- Pagani L, Schiffels S, Gurdasani D, Danecek P, Scally A, Chen Y, Xue Y, Haber M, Ekong R, Oljira T, et al. 2015. Tracing the Route of Modern Humans out of Africa by Using 225 Human Genome Sequences from Ethiopians and Egyptians. *The American Journal of Human Genetics* 96:986–991.
- Palstra FP, Heyer E, Austerlitz F. 2015. Statistical inference on genetic data reveals the complex demographic history of human populations in Central Asia. *Mol Biol Evol*:msv030.
- Patin E, Laval G, Barreiro LB, Salas A, Semino O, Santachiara-Benerecetti S, Kidd KK, Kidd JR, Van der Veen L, Hombert J-M, et al. 2009. Inferring the Demographic History of African Farmers and Pygmy Hunter–Gatherers Using a Multilocus Resequencing Data Set. *PLoS Genet* 5:e1000448.
- Perry GH, Dominy NJ, Claw KG, Lee AS, Fiegler H, Redon R, Werner J, Villanea FA, Mountain JL, Misra R, et al. 2007. Diet and the evolution of human amylase gene copy number variation. *Nat. Genet.* 39:1256–1260.
- Pickrell JK, Coop G, Novembre J, Kudaravalli S, Li JZ, Absher D, Srinivasan BS, Barsh GS, Myers RM, Feldman MW, et al. 2009. Signals of recent positive selection in a worldwide sample of human populations. *Genome Res.* 19:826–837.
- Pinhasi R, Stock JT. 2011. *Human bioarchaeology of the transition to agriculture.* Wiley Online Library
- Pritchard JK, Pickrell JK, Coop G. 2010. The Genetics of Human Adaptation: Hard Sweeps, Soft Sweeps, and Polygenic Adaptation. *Current Biology* 20:R208–R215.
- Pritchard JK, Di Rienzo A. 2010. Adaptation – not by sweeps alone. *Nat Rev Genet* 11:665–667.
- Pritchard JK, Seielstad MT, Perez-Lezaun A, Feldman MW. 1999. Population growth of human Y chromosomes: a study of Y chromosome microsatellites. *Molecular biology and evolution* 16:1791–1798.
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, et al. 2007. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *The American Journal of Human Genetics* 81:559–575.
- Quintana-Murci L, Semino O, Bandelt H-J, Passarino G, McElreavey K, Santachiara-Benerecetti AS. 1999. Genetic evidence of an early exit of Homo sapiens sapiens from Africa through eastern Africa. *Nat Genet* 23:437–441.
- Ramachandran S, Deshpande O, Roseman CC, Rosenberg NA, Feldman MW, Cavalli-Sforza LL. 2005. Support from the relationship of genetic and geographic distance in human populations for a serial founder effect originating in Africa. *Proc. Natl. Acad. Sci. U.S.A.* 102:15942–15947.
- Rasmussen M, Guo X, Wang Y, Lohmueller KE, Rasmussen S, Albrechtsen A, Skotte L, Lindgreen S, Metspalu M, Jombart T, et al. 2011. An Aboriginal Australian Genome Reveals Separate Human Dispersals into Asia. *Science* 334:94–98.
- Reed DR, Zhu G, Breslin PAS, Duke FF, Henders AK, Campbell MJ, Montgomery GW, Medland SE, Martin NG, Wright MJ. 2010. The perception of quinine taste intensity is associated with common genetic variants in a bitter receptor cluster on chromosome 12. *Hum. Mol. Genet.*:ddq324.
- Roach JC, Glusman G, Smit AFA, Huff CD, Hubley R, Shannon PT, Rowen L, Pant KP, Goodman N, Bamshad M, et al. 2010. Analysis of Genetic Inheritance in a Family Quartet by Whole-Genome Sequencing. *Science* 328:636–639.
- Roberts RG, Jones R, Smith MA. 1990. Thermoluminescence dating of a 50,000-year-old human occupation site in northern Australia.

- Robino A, Mezzavilla M, Pirastu N, Dognini M, Tepper BJ, Gasparini P. 2014. A Population-Based Approach to Study the Impact of PROP Perception on Food Liking in Populations along the Silk Road. *PLoS ONE* 9:e91716.
- Roebroeks W, van Kolfschoten T. 1994. The earliest occupation of Europe: a short chronology. *Antiquity* 68:489–503.
- Rolls ET, Baylis LL. 1994. Gustatory, olfactory, and visual convergence within the primate orbitofrontal cortex. *The Journal of neuroscience* 14:5437–5452.
- Roper SD. 2006. Signaling in the chemosensory systems. *Cellular and Molecular Life Sciences CMLS* 63:1494–1500.
- Rose JI, Usik VI, Marks AE, Hilbert YH, Galletti CS, Parton A, Geiling JM, Černý V, Morley MW, Roberts RG. 2011. The Nubian Complex of Dhofar, Oman: An African Middle Stone Age Industry in Southern Arabia. *PLoS ONE* 6:e28239.
- Rozzi FVR, Koudou Y, Froment A, Le Bouc Y, Botton J. 2015. Growth pattern from birth to adulthood in African pygmies of known age. *Nat Commun* [Internet] 6. Available from: <http://www.nature.com/ncomms/2015/150728/ncomms8672/full/ncomms8672.html#acknowledgments>
- Ruff CB. 1994. Morphological adaptation to climate in modern and fossil hominids. *Am. J. Phys. Anthropol.* 37:65–107.
- Sabeti PC, Reich DE, Higgins JM, Levine HZP, Richter DJ, Schaffner SF, Gabriel SB, Platko JV, Patterson NJ, McDonald GJ, et al. 2002. Detecting recent positive selection in the human genome from haplotype structure. *Nature* 419:832–837.
- Sabeti PC, Schaffner SF, Fry B, Lohmueller J, Varilly P, Shamovsky O, Palma A, Mikkelsen TS, Altshuler D, Lander ES. 2006. Positive natural selection in the human lineage. *Science* 312:1614–1620.
- Sabeti PC, Varilly P, Fry B, Lohmueller J, Hostetter E, Cotsapas C, Xie X, Byrne EH, McCarroll SA, Gaudet R, et al. 2007. Genome-wide detection and characterization of positive selection in human populations. *Nature* 449:913–918.
- Sands B. 1998. The linguistic relationship between Hadza and Khoisan. *Language, Identity, and Conceptualization among the Khoisan* 15:265–283.
- Scally A, Durbin R. 2012. Revising the human mutation rate: implications for understanding human evolution. *Nat. Rev. Genet.* 13:745–753.
- Scheinfeldt LB, Soi S, Tishkoff SA. 2010. Working toward a synthesis of archaeological, linguistic, and genetic data for inferring African population history. *PNAS* 107:8931–8938.
- Schiffels S, Durbin R. 2014. Inferring human population size and separation history from multiple genome sequences. *Nat Genet* 46:919–925.
- Schiffman SS, Dackis C. 1975. Taste of nutrients: Amino acids, vitamins, and fatty acids. *Perception & Psychophysics* 17:140–146.
- Schlebusch CM, Gattepaille LM, Engström K, Vahter M, Jakobsson M, Broberg K. 2015. Human Adaptation to Arsenic-Rich Environments. *Molecular biology and evolution:msv046*.
- Schlebusch CM, Sjödin P, Skoglund P, Jakobsson M. 2012. Stronger signal of recent selection for lactase persistence in Maasai than in Europeans. *Eur. J. Hum. Genet.*
- Schlebusch CM, Skoglund P, Sjödin P, Gattepaille LM, Hernandez D, Jay F, Li S, Jongh MD, Singleton A, Blum MGB, et al. 2012. Genomic Variation in Seven Khoe-San Groups Reveals Adaptation and Complex African History. *Science* 338:374–379.

- Ségurel L, Austerlitz F, Toupance B, Gautier M, Kelley JL, Pasquet P, Lonjou C, Georges M, Voisin S, Cruaud C, et al. 2013. Positive selection of protective variants for type 2 diabetes from the Neolithic onward: a case study in Central Asia. *Eur J Hum Genet* 21:1146–1151.
- Ségurel L, Lafosse S, Heyer E, Vitalis R. 2010. Frequency of the AGT Pro11Leu polymorphism in humans: Does diet matter? *Ann. Hum. Genet.* 74:57–64.
- Shea JJ. 2003. The Middle Paleolithic of the East Mediterranean Levant. *Journal of World Prehistory* 17:313–394.
- Shigemura N, Iwata S, Yasumatsu K, Ohkuri T, Horio N, Sanematsu K, Yoshida R, Margolskee RF, Ninomiya Y. 2013. Angiotensin II Modulates Salty and Sweet Taste Sensitivities. *J. Neurosci.* 33:6267–6277.
- Shilov VP. 1975. Models of Pastoral Economies in the Steppe Regions of Eurasia in the Enolithic and Early Bronze Ages. *Sovietskaya Arkeologiya* 1:5–16.
- Shimada MK, Panchapakesan K, Tishkoff SA, Nato AQ, Hey J. 2007. Divergent haplotypes and human history as revealed in a worldwide survey of X-linked DNA sequence variation. *Molecular biology and evolution* 24:687–698.
- Simo-Servat O, Hernandez C, Simo R. 2013. Genetics in Diabetic Retinopathy: Current Concepts and New Insights. *Current Genomics* 14:289–299.
- Small DM, Prescott J. 2005. Odor/taste integration and the perception of flavor. *Experimental Brain Research* 166:345–357.
- Smith TM, Tafforeau P, Reid DJ, Grün R, Eggins S, Boutakiout M, Hublin J-J. 2007. Earliest evidence of modern human life history in North African early Homo sapiens. *PNAS* 104:6128–6133.
- Sponheimer M, Lee-Thorp JA. 1999. Isotopic Evidence for the Diet of an Early Hominid, *Australopithecus africanus*. *Science* 283:368–370.
- Stamatoyannopoulos JA. 2012. What does our genome encode? *Genome Res* 22:1602–1611.
- Steiner JE. 1974. Discussion Paper: Innate, Discriminative Human Facial Expressions to Taste and Smell Stimulation. *Annals of the New York Academy of Sciences* 237:229–233.
- Stringer C. 2002. Modern human origins: progress and prospects. *Philosophical Transactions of the Royal Society B: Biological Sciences* 357:563–579.
- Stringer CB, Grün R, Schwarcz HP, Goldberg P. 1989. ESR dates for the hominid burial site of Es Skhul in Israel. *Nature* 338:756–758.
- Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123:585–595.
- Tan J, Yang Y, Tang K, Sabeti PC, Jin L, Wang S. 2013. The adaptive variant EDARV370A is associated with straight hair in East Asians. *Hum Genet* 132:1187–1191.
- Tattersall I. 2009. Human origins: out of Africa. *Proceedings of the National Academy of Sciences* 106:16018–16021.
- Tavaré S, Balding DJ, Griffiths RC, Donnelly P. 1997. Inferring coalescence times from DNA sequence data. *Genetics* 145:505–518.
- Tishkoff SA, Reed FA, Friedlaender FR, Ehret C, Ranciaro A, Froment A, Hirbo JB, Awomoyi AA, Bodo J-M, Doumbo O, et al. 2009. The Genetic Structure and History of Africans and African Americans. *Science* 324:1035–1044.
- Tishkoff SA, Reed FA, Ranciaro A, Voight BF, Babbitt CC, Silverman JS, Powell K, Mortensen HM, Hirbo JB, Osman M, et al. 2007. Convergent adaptation of human lactase persistence in Africa and Europe. *Nat. Genet.* 39:31–40.

- Tishkoff SA, Varkonyi R, Cahinhinan N, Abbes S, Argyropoulos G, Destro-Bisol G, Drousiotou A, Dangerfield B, Lefranc G, Loiselet J, et al. 2001. Haplotype Diversity and Linkage Disequilibrium at Human G6PD: Recent Origin of Alleles That Confer Malarial Resistance. *Science* 293:455–462.
- Trinkaus E. 2005. Early modern humans. *Annu. Rev. Anthropol.* 34:207–230.
- Valoch K. 2013. The Early Palaeolithic site Stranska Skala I near Brno (Czechoslovakia). *Anthropologie* 51:67.
- Vandenbeuch A, Clapp TR, Kinnamon SC. 2008. Amiloride-sensitive channels in type I fungiform taste cells in mouse. *BMC neuroscience* 9:1.
- Verdu P, Austerlitz F, Estoup A, Vitalis R, Georges M, Théry S, Froment A, Le Bomin S, Gessain A, Hombert J-M, et al. 2009. Origins and Genetic Diversity of Pygmy Hunter-Gatherers from Western Central Africa. *Current Biology* 19:312–318.
- Vigilant L, Stoneking M, Harpending H, Hawkes K, Wilson AC. 1991. African populations and the evolution of human mitochondrial DNA. *Science* 253:1503–1507.
- Voight BF, Kudravalli S, Wen X, Pritchard JK. 2006. A map of recent positive selection in the human genome. *PLoS Biol.* 4:e72.
- Weiner S, Xu Q, Goldberg P, Liu J, Bar-Yosef O. 1998. Evidence for the use of fire at Zhoukoudian, China. *Science* 281:251–253.
- Wells RS, Yuldasheva N, Ruzibakiev R, Underhill PA, Evseeva I, Blue-Smith J, Jin L, Su B, Pitchappan R, Shanmugalakshmi S. 2001. The Eurasian heartland: a continental perspective on Y-chromosome diversity. *Proceedings of the National Academy of Sciences* 98:10244–10249.
- White TD, Asfaw B, DeGusta D, Gilbert H, Richards GD, Suwa G, Howell FC. 2003. Pleistocene homo sapiens from middle awash, ethiopia. *Nature* 423:742–747.
- Williamson SH, Hubisz MJ, Clark AG, Payseur BA, Bustamante CD, Nielsen R. 2007. Localizing Recent Adaptive Evolution in the Human Genome. *PLoS Genet* 3:e90.
- Wright S. 1943. Isolation by Distance. *Genetics* 28:114–138.
- Wright S. 1950. Genetical structure of populations. *Nature* 166:247–249.
- Yoshida R, Niki M, Jyotaki M, Sanematsu K, Shigemura N, Ninomiya Y. 2013. Modulation of sweet responses of taste receptor cells. *Semin. Cell Dev. Biol.* 24:226–231.
- Yotova V, Lefebvre J-F, Kohany O, Jurka J, Michalski R, Modiano D, Utermann G, Williams SM, Labuda D. 2007. Tracing genetic history of modern humans using X-chromosome lineages. *Human genetics* 122:431–443.
- Zhao GQ, Zhang Y, Hoon MA, Chandrashekar J, Erlenbach I, Ryba NJP, Zuker CS. 2003. The Receptors for Mammalian Sweet and Umami Taste. *Cell* 115:255–266.
- Zhou X, Stephens M. 2012. Genome-wide efficient mixed-model analysis for association studies. *Nat Genet* 44:821–824.