Lifestyle and genetic diversity: a study on African populations

Cristina Sofia de Sousa Cardoso e Valente dos Santos

Programa Doutoral em Biologia
Biologia
2015

Orientador
Professora Doutora Maria João Prata Martins Ribeiro, Professora Associada com Agregação, Faculdade de Ciências da Universidade do Porto

Coorientador
Doutora Maria Leonor Rodrigues de Sousa Botelho de Gusmão, Professor Adjunto, Universidade do Estado do Rio de Janeiro

Acknowledgments/Agradecimentos

Em primeiro lugar quero agradecer ao IPATIMUP por me ter recebido para realizar o meu trabalho de doutoramento.

Ao grupo de Genética Populacional e ao seu grupo leader António Amorim, por me ter integrado e permitido desenvolver o meu projeto de doutoramento.

Gostaría de agradecer à minha orientadora Maria João Prata.

Aos meus co-orientadores Leonor e Walther. Ao Walther apesar de não ter tido uma intervenção direta no trabalho (áreas de trabalho um pouco distintas, apenas isso!) foi sempre prestável. Leo mesmo estando longe tiveste um papel importantíssimo na realização deste trabalho. Mas isto é um até já!

E apesar de oficialmente não teres qualquer vínculo e obrigações na realização do meu trabalho de doutoramento é a ti Luís que tenho que agradecer do fundo do coração! Jamais esquecerei a tua ajuda!

A todas as pessoas que participaram menos ou mais ativamente no meu projeto um muito obrigada, nomeadamente à Marisa que realizou o seu trabalho de mestrado comigo!

Aos meus colegas de grupo um obrigado, Ana Goios, Cíntia, Verónica, Alexandra, Sandra, Ana Moleirinho; à Sofia, não só pela ajuda no laboratório, como na concepção do meu trabalho e ao Rui Manuel pelo carinho, apoio constante e pela nossas gargalhadas!

Iva estás longe mas estiveste sempre aqui pertinho de mim amiga! Às minhas "matemáticas" Nádia e Inês por serem quem são, e por todo o carinho e amizade.

Mafalda foi "nestes propósitos" que nos conhecemos mas tenho a certeza que nasceu aqui uma amizade muito especial.

À Inês porque é irrevogável o que me identifico contigo, maneira de estar na vida e afins...tenho a certeza que há coisas que tu entendes de maneira diferente de toda a gente…um obrigada sincero e ainda bem que as nossas vidas se cruzaram!

Aos meus amigos do sul, que apesar de estarem longe, estão sempre num lugar muito especial.

Um beijinho grande à minha mãe porque é A minha Mãe!

À minha mana, à minha "mais-que-tudo", porque se existe algum exemplo a seguir, és tu sem dúvida. Tenho muito orgulho em ser tua irmã!

E finalmente, como costumas dizer "the last but not the least" é a ti PAIZÃO que dedico este livrinho. Sem ti nada disto era possível!

Abstract

A major question in human population genetics is to understand the contribution of demographic history and selection to the patterns of genetic diversity that characterize human populations today. Identifying targets of selection in humans has been slow and difficult because is not easy to distinguish between the signatures left by selection from those due to demography. Gradually, however, studies of specific candidate *loci* together with genome-wide searches succeeded to ascertain a number of genes that show population differences due to adaptations especially to pathogens, climate and diet.

Concerning diet, an important change in human prehistory occurred with the rise of agriculture, beginning 10-12 Kya ago in the Near East. The Neolithic transition to modes of subsistence not relying in hunting and gathering is not only associated with new food resources but also with reduced dietary diversity, explaining the belief that the Neolithic represents the start of an era propitiating dietary adaptations.

In this study a candidate gene approach was applied to investigate whether differences in diets, roughly inferred from lifestyles of populations (hunter-gathering, agriculturalism and pastoralism), could have shaped diversity at genes that more or less directly play a role in the metabolism or in the taste perception of substances (including xenobiotics), that gain entry into the organism through food.

The first genes examined were *AGXT*, *PLRP2*, *MTRR*, *NAT2* and *CYP3A5*, since assumedly they might play a role in the metabolism of meat (*AGXT*), cereals (*PLRP2)*, folates (*MTRR* and *NAT2)* and salt (*CYP3A5*) in food*.* In each of those genes, variants known to have potential phenotypic effects were selected. In order to enrich the data for African populations, which still remain understudied, six groups were screened for the selected variants: agriculturalists from Angola, Equatorial Guinea and Mozambique, pastoralists from Uganda, and the hunter-gatherers Baka Pygmies from Gabon and Khoisan from Namibia. In addition a sample from Portuguese was also studied.

The analyses performed revealed a connection between patterns of genetic diversity at *PLRP2* and *NAT2* and lifestyles of populations, suggesting that both genes have evolved under a selective pressure related with diet. Contrarily, no signs were detected that *AGXT*, *MTRR* and *CYP3A5* could represent dietary adaptations in humans. Still, *CYP3A5* revealed the most unequivocal signature of selection, which in fact was found to be

strongly correlated with latitude but not with differences in eating habits between populations.

Concerning the taste genes, attention was focused on *TAS1R1*, *TAS1R3, TAS2R16* and *TAS238*, which are involved in the perception of sweet, bitter and *umami* tastes. In each gene individual variations reported to influence different taste modalities were selected and screened in five African populations with different lifestyles and in the Portuguese. No signs emerged that diversity at those *loci* had been shaped by a dietary variable differing according to lifestyle of populations, except in *TAS2R16,* and restrictively in Africa, where the frequency of a variant decreasing sensibility to cyanogenic glycosides showed to be especially well represented in farmer groups from Western-Central Africa, raising the question on whether the detection of those bitter compounds was less critical in African farmers than in non-food producers.

Some of the *TAS* genes analyzed presented unusual levels of differentiation across African, Asian and European populations, difficultly conciliated with the continental structure explainable by the dispersion history of human populations. This observation instigated the use of an alternative approach to address selection at the *TAS* genes, viewing which full sequences for different *TAS* regions were retrieved from the 1000 Genome database. Dissection of the patterns of molecular diversity led to identify several signatures of non-neutral evolution. Two genes from the *TAS1R* family (*TAS1R1* and *TAS1R3*), exhibited signals of positive selection, whereas *TAS2R38* presented the features expected for genes evolving under balancing selection. In *TAS2R16* a complex evolutionary history was detected.

In addition to the well-known role of *TAS* proteins as components of taste receptors in taste bud cells, it is being demonstrated that they are also involved in many other aspects related with human health. A better understanding of the functions of those proteins is still needed to understand the biological basis underlying what seems to be an adaptive evolution of these *TAS* genes in humans.

In the future, new theoretical models and statistical methods better suited to address selection targeting phenotypic traits for which account many alleles each with small effects, together with the development of the functional approaches necessary to validate the genomic signatures of selection, will make it possible to obtain refreshed insights on phenotypic adaptations in humans, as well as to understand at which point they were

prompted by the changes that accompanied the dispersion of humans out of Africa, or latter in the Neolithic, the transition to new subsistence economies.

Key words: Human adaptations, diet, metabolic pathways, taste receptors, gene diversity, selection

Resumo

Um dos grandes objetivos da Genética Populacional Humana é compreender a contribuição da história demográfica e da seleção, para os padrões de diversidade genética que atualmente caracterizam as populações humanas. A identificação de alvos de seleção tem sido lenta e difícil, por não ser fácil distinguir os sinais deixados pela seleção dos causados pela demografia das populações. Contudo, estudos debruçados sobre genes candidatos, juntamente com rastreios genómicos globais, paulatinamente têm trazido à luz diversos genes que apresentam diferenças populacionais causadas por processos de adaptação, nomeadamente a agentes patogénicos, clima e dieta. No que diz respeito à dieta, uma das grandes alterações ocorridas durante a pré-história humana foi subsequente ao aparecimento da agricultura, cujos vestígios mais antigos datam de há cerca de 10-12 mil anos, tendo sido encontrados no Médio Oriente. No Neolítico, a transição para modos de subsistência diferentes da simples caça e recoleção está não só associada com a introdução de novos tipos de alimentos, mas também com a redução da diversidade na dieta. Daí que seja visão comum presumir que o Neolítico representou o começo de uma nova era propícia ao desenvolvimento de adaptações dietéticas.

Neste estudo, foi aplicada uma abordagem dirigida a genes candidatos com o intuito de investigar se diferenças na dieta de populações, inferidas *grosso modo* pelas diferenças nos estilos de vida (caça-recoleção, agricultura e pastoralismo), podem ter moldado a diversidade em genes que têm algum papel quer ao nível do metabolismo quer na perceção do sabor dos alimentos ou substâncias (incluindo xenobióticos), que entram no organismo através da dieta.

O primeiro grupo de genes examinado incluiu *AGXT*, *PLRP2*, *MTRR*, *NAT2* e *CYP3A5*, presumindo-se que cada um deles podia influenciar o metabolismo da carne (*AGXT*), dos cereais (*PLRP2*), dos folatos (*MTRR* e *NAT2*) ou do sal (*CYP3A5*). Tendo em vista aumentar o nível de caracterização quanto a populações de África (continente que continua pouco estudado), seis populações africanas foram tipadas para as variantes selecionadas em cada gene: agricultores de Angola, Guiné Equatorial e Moçambique; pastores do Uganda, e caçadores-recoletores representados pelos Baka Pigmeus do Gabão e Khoisan da Namíbia. A população portuguesa também foi estudada, de forma a obter uma amostra de referência não africana.

As análises efetuadas revelaram existir uma certa ligação entre os padrões de diversidade genética nos genes *PLRP2* e *NAT2* e o estilo de vida das populações, sugerindo que ambos os genes podem ter estado sob forças seletivas relacionadas com a dieta. Contrariamente, em *AGXT*, *MTRR* e *CYP3A5* não foram encontrados quaisquer sinais que indicassem poder constituir adaptações dietéticas no homem. Em *CYP3A5* detetou-se, contudo, um acentuado sinal de seleção, muito fortemente correlacionado com a latitude mas sem relação aparente com diferenças nos hábitos alimentares das populações.

Quanto aos genes codificadores dos recetores do gosto, a análise incidiu sobre *TAS1R1*, *TAS1R3, TAS2R16* e *TAS2R38,* por estarem envolvidos na percepção dos sabores doce, amargo e *umami*. Em cada um deles foram selecionadas variações que reconhecidamente influenciam de algum modo a percepção de sabores, tendo também sido genotipadas em cinco populações africanas com diferentes estilos de vida, bem como na população portuguesa.

De uma forma geral, os resultados obtidos não sustentaram que a diversidade nesses *loci* tivesse sido influenciada por alguma variável dietética relacionada com o estilo de vida das populações. Porém, quanto ao gene *TAS2R16*, e restritivamente em populações africanas, a detecção de uma variante responsável pela diminuição da sensibilidade a glicósidos cianogénicos com frequência bastante comum em grupos de agricultores do Centro-Oeste africano, pode constituir um indício de que a deteção desses compostos, que provocam um sabor amargo, foi menos crítica nos agricultores africanos do que em grupos que não produzem alimentos.

Detetou-se, ainda, que alguns dos genes *TAS* analisados apresentavam níveis de diferenciação genética entre populações de África, Ásia e Europa, dificilmente explicáveis pela história da dispersão das populações humanas após a saída do continente africano. Esta observação levou a utilizar uma abordagem alternativa para pesquisar sinais de seleção, o que implicou recorrer às sequências completas relativas às regiões *TAS* disponíveis na base de dados 1000 Genomes. A análise dos padrões de diversidade molecular nestas sequências permitiu identificar diversos sinais que não se enquadram nas previsões do modelo de evolução neutra. Dois genes da família *TAS1R* (*TAS1R1* e *TAS1R3*) reuniam sinais de seleção positiva, enquanto o gene *TAS2R38* apresentava vestígios moleculares esperadas em genes sob efeito de seleção balanceadora. No gene *TAS2R16* a história evolutiva detetada foi bem mais complexa.

Encontra-se já bem estabelecida a função das proteínas *TAS* como componentes dos receptores do gosto, mas só recentemente se tem vindo a demonstrar o seu envolvimento em muitos outros mecanismos relacionados com a saúde humana. Assim, é ainda necessário aprofundar o conhecimento sobre a função destas proteínas para perceber melhor a base biológica do que parece ter sido um percurso evolutivo adaptativo dos genes *TAS* em humanos.

Espera-se que num futuro próximo, novos modelos teóricos e métodos estatísticos mais adequados para investigar seleção dirigida a características fenotípicas influenciadas por múltiplos genes, cada um com pequeno efeito, possam abrir novas perspetivas sobre a natureza das adaptações fenotípicas no homem, bem como perceber até que ponto foram desencadeadas pelas alterações que acompanharam a dispersão do homem para fora de África, ou mais tarde, no Neolítico, a transição para novas economias de subsistência.

Palavras-chave: Adaptações humanas, dieta, vias metabólicas, receptores do gosto, diversidade genética, seleção

Index

Figure Index

Abbreviations

AMH	Anatomically modern Human
AMOVA	Analysis of Molecular Variance
AMY	Amylase enzyme
EHH	Extended haplotype homozygosity
F_{ST}	Wright's fixation index
GPCR	G protein coupled receptor
HWE	Hardy-Weinberg Equilibrium
HIM	Hierarchical island model
IBD	Identity-by-descent
INDEL	Insertion or deletion of DNA bases
IM	Finite island model
LD	Linkage disequilibrium
LCT	Lactase enzyme
MAF	Minimum allele frequency
MDS	Multidimensional scaling
mtDNA	Mitochondrial DNA
Ne	Effective population size
PCR	Polymerase chain reaction
PROP	6-n-Propylthiouracil
PTC	Phenylthiocarbamide
SBE	Single base extension
SFS	Site frequency spectrum
SNP	Single nucleotide polymorphism
STR	Short tandem repeat
TMRCA	Time to most recent common ancestor
WGS	Whole-genome sequencing

CHAPTER 1. INTRODUCTION

1.1 Human Evolutionary Genetics

Human evolution is an intriguing process that has attracted attention from many different scientific areas that over time are giving important complementary insights to reconstruct the evolutionary history of our species. Among these areas, human genetics has always played a central role, but its importance grew dramatically in the last few decades as a result of the many progresses experienced in the field.

The ability to explore the genetic variation across the genome has increased with the unprecedented technological advances, making available new genetic information on extant and even on extinct human populations. In parallel, the number of analytical tools developed to deal with genetic data also augmented.

One of the aims in human evolutionary genetics is to understand the principal factors that underlie the distribution of genetic diversity among current human populations, attempting to put in a time scale when those factors started to act. The analysis of patterns of distribution of genetic variation can be used to answer many questions about human evolution, ranging from past demographic events until possible selective factors that have triggered genetic adaptations in human populations.

1.2 Modern Human Origins

During the last few decades the origin of modern humans has been strongly debated. For the purpose of simplicity, the different interpretations will be summarized in the following two most competing models i) the “Multiregional hypothesis” and ii) the ”Recent African hypothesis”, also known as “Out-of-Africa”.

The first model suggests that modern humans did not emerge in a single geographic region, assuming thus that non-African modern humans are descent from different populations of *Homo erectus* that underwent independent migrations occurring over a million years ago reaching different geographic areas (Oppenheimer 2012).

Multi-regionalists explain the tendency for modern regional populations to resemble each other morphologically, despite their putative distinct *H. erectus* ancestors, due to intensive gene flow between populations.

Contrarily, the “Out-of-Africa” model (Figure 1), sustains that all worldwide populations descent from anatomically modern humans firstly originated in Africa in the past 200 Kya, which then left the continent within the last 100 Kya replacing all pre-existing human groups outside Africa (Oppenheimer 2012).

Evidence supporting the “Multiregional Origin” theory is essentially provided by archeological data (Lahr 1994), while the “Out-of-Africa” hypothesis besides sustained by fossil and archeological records is additionally supported by extensive genetic data (Quintana-Murci, et al. 1999; Laval, et al. 2010). A scenario consistent with the origin of modern humans in Africa approximately 150–200 Kya ago has been revealed by numerous genetic studies based either in autosomal markers (Bowcock, et al. 1994; Tishkoff, et al. 1996; Chu, et al. 1998; Jin, et al. 1999), or maternally (Harihara, et al. 1988; Vigilant, et al. 1991; Ballinger, et al. 1992; Yao, et al. 2000; Yao, et al. 2002; Wen, et al. 2004; Wen, et al. 2005) and paternally (Su, et al. 1999; Jin and Su 2000; Su, et al. 2000; Ke, et al. 2001; Shi, et al. 2005; Shi, et al. 2008) inherited mitochondrial DNA and Y chromosome lineages, respectively.

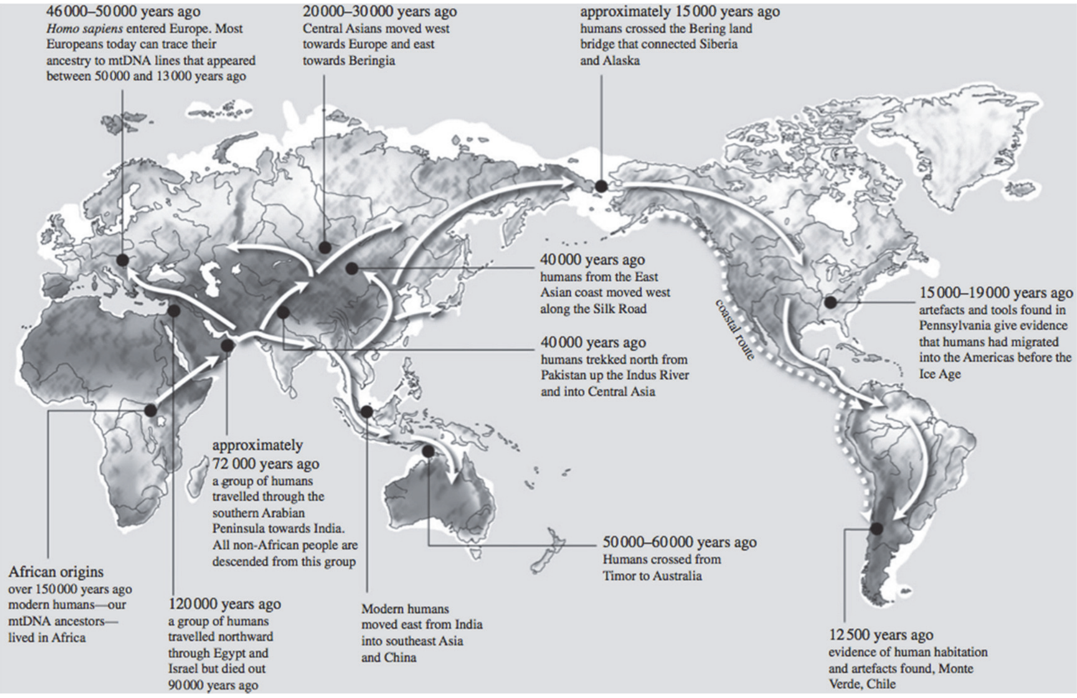

46 000–50 000 years ago
Homo sapiens entered Europe. Most Europeans today can trace their ancestry to mtDNA lines that appeared between 50 000 and 13 000 years ago
20 000–30 000 years ago
Central Asians moved west towards Europe and east towards Beringia
approximately 15 000 years ago
humans crossed the Bering land bridge that connected Siberia and Alaska
40 000 years ago
humans from the East Asian coast moved west along the Silk Road
coastal route
15 000–19 000 years ago
artefacts and tools found in Pennsylvania give evidence that humans had migrated into the Americas before the Ice Age
40 000 years ago
humans trekked north from Pakistan up the Indus River and into Central Asia
approximately
72 000 years ago
a group of humans travelled through the southern Arabian Peninsula towards India. All non-African people are descended from this group
50 000–60 000 years ago
Humans crossed from Timor to Australia
African origins
over 150 000 years ago modern humans—our mtDNA ancestors—lived in Africa
120 000 years ago
a group of humans travelled northward through Egypt and Israel but died out 90 000 years ago
Modern humans moved east from India into southeast Asia and China
12 500 years ago
evidence of human habitation and artefacts found, Monte Verde, Chile

Figure 1. **Narrative map of modern human dispersal** (Oppenheimer 2012).

The “Recent African hypothesis” model holds that after the emergence of modern humans in Africa (Figure 1 and 2) around 150-200 Kya, the migration towards Eurasia took place within the last 40–80 Kya, and finally colonization of the Americas has begun within the last 15-19 Kya.

The geographic expansion of modern humans out of Africa was predictably accompanied by a series of bottlenecks leading to relevant loss of genetic diversity in non-African populations. The prediction has received strong empirical support with the observation of clear founder effects in non-African populations, finding that has paved the way to hypothesize a serial founder model of migration in the history of non-Africans in which the dispersion of populations occurred in small steps, with each migration wave involving a sampling of variation from the previous population (Campbell and Tishkoff 2010).

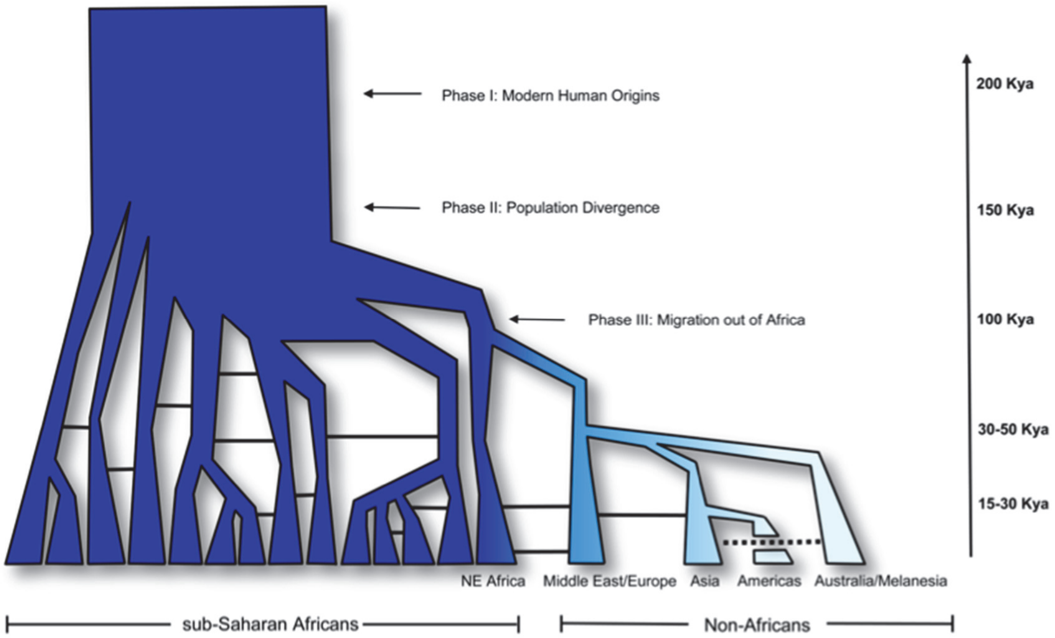


Figure 2. The “Recent African Origin” model of modern humans and population substructure in Africa (Campbell and Tishkoff 2010).

About the paths taken by Anatomically Modern Human (AMH) to expand to Middle East, two main migratory routes out of Africa have been hypothesized: i) the northern exit, involving a route of migration via North Africa and the Nile valley into the Levant with subsequent dispersal into both Europe and Asia (Mellars 2006), according to which the movement of AMH from the northeast to the Levant was via the Sinai Peninsula in Egypt after the crossing of the Sahara (Oppenheimer 2012); and ii) the southern exit via a coastal route crossing the Bab-el-Mandeb strait at the mouth of the Red Sea, and then AMH rapidly migrated along the South Asia coastline. The last hypothesis was apparently sustained by some features of the patterns of mtDNA diversity, namely by the demonstration that N and M sequences, the two primary sublineages of haplogroup L3 characteristic of sub-Saharan African populations, were already originated outside Africa, being both very well represented in East Eurasia, including South Asia, whereas in the Near East/Europe only N lineages are found that are younger and show lower genetic diversity compared to the sub-lineages of N from South Asia (Metspalu, et al. 2004; Macaulay, et al. 2005). The scenario of a single southern exit out of Africa was also strongly supported by the much later arrival of modern humans into Europe than into Australia (Oppenheimer 2012) where archeological evidence revealed the presence of human settlement dating to around 55-60 Kya (Quintana-Murci, et al. 1999; Mellars 2006; Stanyon, et al. 2009).

However, there are recent genetic studies that instead of a single exit route sustain a multiple-dispersal model for the out-of-Africa, holding that after a “southern route" dispersal into Asia as early as the late Middle Pleistocene, other separate dispersal of AMH into northern Eurasia must have occurred more recently, 25-38 Kya (Reyes-Centeno, et al. 2014; Fregel, et al. 2015).

The dispersal out of Africa of modern humans is indeed a subject far from being closed, and so in next few years certainly it will continue to be addressed as long as new genetic, archaeological and paleoanthropological evidence appear.

1.3 Natural Selection

Since AMH abandoned the homeland in Africa they had to explore new environments while crossing the globe (Scheinfeldt and Tishkoff 2013), having been exposed to new selective pressures. It is clear that during the colonization of novel habitats humans were confronted with different climates, food sources, pathogens and other factors that might have triggered specific genetic adaptations (Balaresque, et al. 2007).

The role of selection in recent human evolution has been intensely debated, which explains the intensive research undertaken in the last few years looking for evidence of selection in the human genome. The major challenge is being to identify clear signs of adaptation in humans, an issue that is framed in elucidating the relative contributions of natural selection and demographic history to extant patterns of genetic variation in human populations (Biswas and Akey 2006). In fact, if during the out of Africa humans experienced strong bottlenecks, the colonization of the rest of the world by modern humans was also accompanied by a series of founder effects followed by local population expansions. All these demographic events might have led to genetic signatures in current human populations that are difficult to discern from those expected under the action of selection.

In spite of these difficulties, recent genome wide studies have identified a large number of human *loci* showing signatures of recent selective sweeps (Voight, et al. 2006; Sabeti, et al. 2007; Williamson, et al. 2007), signatures that have also been captured even when highly sensitive approaches to search for selection were applied (Enard, et al. 2014).

Elucidating which traits are adaptive in humans is of fundamental interest to fully understand their evolution.

Most simply, natural selection can change frequencies in genetic variations in traits that influence fitness, in such a way that the favored traits increase in frequency in successive generations. There are, however, different modes of selection, each of them affecting DNA variation in a specific way.

1.3.1 Different Modes of Natural Selection

Natural selection can affect DNA diversity around or within a gene in diverse ways, depending on its mode of action. According to the induced effect in the frequency of variations at a locus, it can be classified in positive, negative or balancing selection (Figure 3).

Locus (or *loci*) under positive selective pressures, which are often thought to be associated with the development of new traits and local environmental adaptations, are expected to show unusual high levels of differentiation among populations (Vitalis, et al. 2014). The key characteristic of positive selection is that it causes an uncommon rapid rise in allele frequency occurring over a time short enough that recombination does not substantially break down the haplotype on which the selected mutation occurs (Sabeti, et al. 2002).

The signatures that positive selection can left in the genome vary according to three distinct modes: i) complete sweep, when a newly mutated allele rapidly increases in frequency until reaching fixation in the population, ii) partial sweep, when the allele only reaches high frequency but not fixation, iii) soft sweep, when selection operates on standing variations or simultaneously on multiple alleles of a locus (Ye and Gu 2011).

Independently of the case of sweep, if a mutation increases in frequency in a population as a result of positive selection, linked neutral variation will be dragged along with it, an effect that is referred to as genetic hitchhiking, which causes an overall reduction in genetic diversity that will be as more pronounced as lower the recombination in a region.

Contrarily, negative selection, also known as purifying selection, is the process leading to the decrease in frequency of a deleterious allele, which ultimately becomes extinct. The elimination of the deleterious allele by negative selection causes a reduction in variability at linked neutral or nearly neutral sites.

Finally balancing selection, is a mode of selection that may occur under two possible scenarios: i) multiple selective pressures act in distinct directions providing a conflict, instead of a single or many selective pressures in the same direction (Ye and Gu 2011); or ii) heterozygous individuals are favored, allowing to maintain both alleles in a population at a frequency equilibrium that is determined by the strength of selection. Conversely to *loci* under positive selection, those that are subjected to balancing selection expectedly will show unusually low levels of population differentiation (Vitalis, et al. 2014).

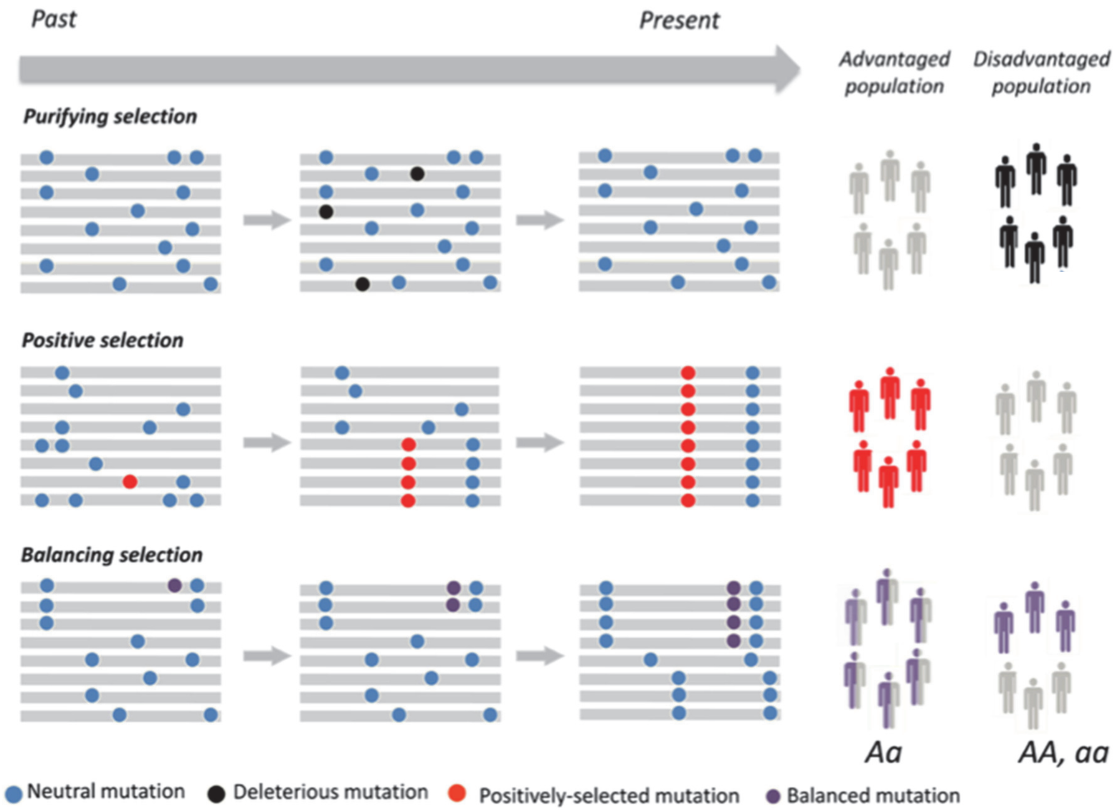


Figure 3. Types of selection and their legacy on the human genome (Quintana-Murci and Clark 2013).

1.3.2 Looking for evidence of Natural Selection

1.3.2.1 Candidate gene and Genome-Wide Studies

Studies that look for signatures of natural selection can be based on candidate gene or genome-wide approaches. The candidate gene approach is focused on genetic variations within specific genes recruited for study based on some *a priori* knowledge of their functional impact in a given trait, a given metabolic pathway or in other processes with biological relevance. Contrarily, genome-wide studies are in general biological hypothesis-free investing in massive genomic scans, and so millions of genetic variations are simultaneously interrogated for selection.

The single (or few) gene approach offers the advantage that genetic variants assumedly relevant in a putative adaptive trait can be prioritized and tested first. Contrarily to this hypothesis-driven (and so hypothesis-testing) approach, the examination of genome-wide scans, is more a hypothesis-generating approach, in the sense that only after the identification of signals of selection in specific genomic regions, it begins the fine search for the individual genetic variants (or genes or genetic pathways) responsible for the evidence of selection, which ultimately can elucidate the biological meaning of the detected signs (Vitti, et al. 2013).

In the last few years, the advances in genomic technology provided unprecedented opportunities for the examination of genome-wide patterns of diversity. We have witnesses a huge accumulation of data either for genome-wide Single Nucleotide Polymorphisms (SNPs) or even whole-genome sequencing (WGS) that has become progressively available for many different populations.

The more often interrogated genome-wide sets of common (those with 5% as minor allele frequency (MAF)) SNPs, have been those combining dense geographic sampling and dense SNP data, such as the Human Genome Diversity Panel CEPH (HGDP-CEPH) and phases I and II of the International HapMap Project (HapMap), considered to be fundamental platforms that enabled the identification of thousands genomic regions associated not only with disease susceptibility but also with common traits including adaptive ones (International HapMap, et al. 2007; Hindorff, et al. 2010).

As for whole-genome information, the primary sources freely available for public use are the *Complete Genomics 69 Genomes* and the *1000 Genomes*, both containing WGS data from global populations. WGS data have several advantages, including the removal of the bias introduced by variant ascertainment strategies, as well as the ability to directly capture rare and functional variants (Scheinfeldt and Tishkoff 2013).

Despite the prospects of genomic strategies as tools for detecting selection, they also have some limitations among which is the difficulty in distinguishing between locus-specific from genome-wide effects at many *loci*, which represents a challenging issue (Vitalis, et al. 2014). As Hernandez et al. 2011 argued, identifying the footprint of a selective sweep against a noisy genomic background is not easy, because patterns of genetic variation reflect the effects of multiple modes of natural selection as well as of demographic history, mutation, and recombination. Moreover, there is the need to cope with the confounding effect of background selection, which refers to a form of negative selection that acts over new deleterious mutations that continuously arise through the process of mutation (Vitti et al, 2013). The random new mutations with deleterious phenotypic effects tend to be eliminated by selection, which causes a reduction in variability at linked neutral or nearly neutral sites, generating the so called background selection, whose strength varies with the recombination rate, the magnitude of selection and the mutation rate.

It is also noteworthy to mention that most large-scale data sets were produced by high-throughput sequencing methods, which for the moment still raise important concerns. Unlike conventional Sanger sequencing, high-throughput sequencing rapidly produces massive amounts of sequence data, sacrificing quality for quantity (Han, et al. 2014). For the commonly used platforms, the error rates of readings are high, ranging from 0.001 to 0.01 per base read, and these rates further vary among different runs even with the same platform (Maruki and Lynch 2014). Due to high error rates (e.g., base-calling and alignment errors) the inferred genotypes from sequencing data might be often inaccurate, compromising then the quality of informative summary statistics (such as allele frequencies, heterozygosity, and nucleotide diversity) extracted from large-scale variation data. If these errors are not taken into account, any population genetic inferences based on the genotype calls can be misleading (Han, et al. 2014).

1.3.2.2 Different strategies to detect signatures of Natural Selection

Various approaches have been used to identify signatures of natural selection. In all of them, observed data are typically confronted with expectations assuming the null hypothesis of selective neutrality.

Within the wide number of methods that have been developed to detect selection, broadly they can be distinguished in those based on comparative data taken from different species, which are aimed at detecting old signatures of selection, and those using population data that allow the detection of more recent adaptive signatures. Only the last will be detailed here.

The most common approaches applied to address selection at a microevolutionary scale, that is, within a species, rely on the population genetics methods that will be next presented.

1.3.2.2.1 Population differentiation-based methods

Under a neutral model of evolution, differences in allele frequencies among populations are determined only by random genetic drift. The effect of drift is, however, influenced by the demographic history of populations, which is the case of human populations since their early emergence are known to have undergone many episodes of founder effects, population growth and admixture. Given that the current patterns of genetic diversity between human populations were highly determined by this complex demographic past (Przeworski, et al. 2000), all neutral *loci* are expected to show identical degree of differentiation between populations. Yet, if positive selection acts on an allele of a locus within one population but not within another, then the allele frequency will increase in the first population while reducing its total variability. As a consequence of the effects of positive selection, the level of differentiation between populations at that locus is predicted to increase compared to the expectation under neutrality.

This is the principle underlying the methods that rely on population differentiation to detect signs of selection. The most frequently used measure of population differentiation has been the Wright´s fixation index (F_{ST}), although several analogs were further proposed, including its Bayesian versions (Holsinger, et al. 2002; Gianola, et al. 2010; Bhatia, et al. 2013).

F_{ST}-based outlier tests have been widely applied to detect *loci* under selection even in genome-wide scans, since they permit to uncover different modes of selection. For instance, whereas an F_{ST} unusually high is indicative of positive selection, an F_{ST} unusually low suggests balancing selection (Hofer, et al. 2012). A major issue in F_{ST}-based methods is to identify how much variation in F_{ST} among *loci* would be expected in the absence of selection, that is, how to obtain the null distribution of F_{ST}.

In 1973, Lewontin and Krakauer (1973) originally presented a F_{ST}-outlier test, but since then a number of more refined tests were developed (for review see Meirmans and Hedrick (2011)), some of which make strict assumptions about demographic history. Among those, the method presented by Beaumont and Nichols (1996) finite Island Model (IM) assumes islands of equal sizes that exchange migrants at the same rate. Although this method has been one of the differentiation-based tests more widely used to detect selection in genomic scans, still it revealed some limitations. Indeed, when applied to species with demographic histories violating the simple assumptions of the method, it can generate a large excess of false positive *loci* pinpointed as being under selection (Excoffier, et al. 2009), since it does not take into account the complex genetic structure, intrinsic characteristic of human populations.

The recognition of this difficulty, led Excoffier, et al. (2009) to propose the Hierarchical Island Model (HIM) (Figure 4), presented by the authors as a more robust model to uncertainties about the exact number of groups and demes per group in a given system.

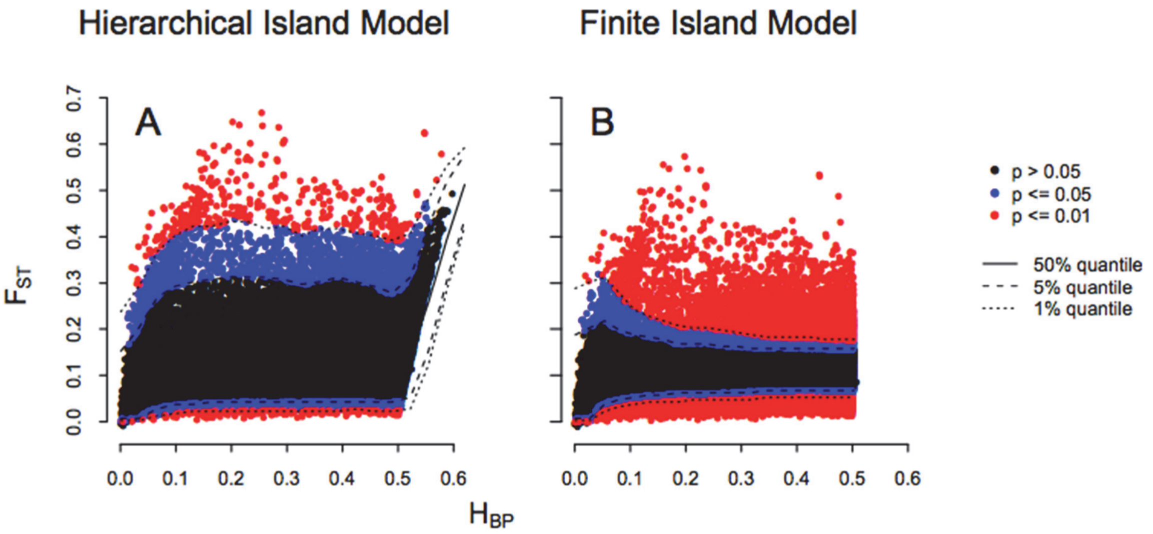


Figure 4. Detection of outlier SNPs based on *F_{ST}*.
Locus specific *P* values computed either based on the Hierarchical Island Model (A) and the Finite Island Model (B) (Hofer, et al. 2012), where F_{ST} is the genetic distance and H_{BP} is Heterozigozity values.

When Excoffier, et al. (2009) tested the method in human and non-human data sets, they obtained a much lower number of significant *loci* than under a non-hierarchical model, a result that represented an important achievement since the elimination of false positive *loci* from genome scans is critical to better determine on which specific class of genes selection is operating. This is especially important in the current era of large-scale projects continuously emerging, providing data on millions of genetic variants, in particular SNPs, while covering many human populations originated from different geographical locations (de Magalhaes and Matsuda 2012).

The approach mentioned above showed to be well suited to detect cases in which selection quickly drove an advantageous allele to high frequency, thereby generating extreme deviations from the neutral patterns of variation. However, when the action of selection over polygenic traits is being investigated, the approach might fail to capture hallmarks of selection, because polygenic traits, as the name indicates, are influenced by many *loci*, with each allele from different *loci* making a small contribution to the phenotype. Furthermore, it is known that most phenotypic variation is polygenic, and it is being demonstrated that selection acting on polygenic traits may lead to subtle shifts in allele frequency at many *loci*, resulting consequently in subtle signs in patterns of genetic variation, which might thus escape detection by conventional strategies to examine selection (Hancock, et al. 2010).

To overcome this difficulty, for the detection of beneficial alleles evolving under a polygenic selection model, approaches were developed that simultaneously consider the spatial distributions of allele frequencies and the underlying selective pressures that have shown to allow the detection of subtle but concordant signs of selection across different populations when controlled for specific environmental variables (Hancock, et al. 2010).

1.3.2.2.2 Patterns of *Linkage Disequilibrium*

Linkage Disequilibrium based methods have also been used to search for natural selection. The extent of *linkage disequilibrium* (LD)*,* defined as the nonrandom association of alleles at different *loci*, can be influenced by many factors, including mutation, population substructure, admixture or migration as well as natural selection. When *loci* are in significant LD*,* the *loci* should be treated together as haplotypes, and not independently.

The simplest scenario from which LD can be created is when a mutation arises on a specific chromosome, since the new variant will necessary appears in a given genetic background and so it will be in total LD with alleles at nearby variable sites. Over time, recombination and mutation in the genomic region result in progressive LD decay, reducing the size of the ancestral haplotype block such that, on average, older mutations will be found on smaller haplotype blocks than younger ones.

However, if an allele sweeps to high frequency in a population due to positive selection that will have a hitchhiking effect on its linked neighbor variants, and specific haplotypes will accordingly be dragged to increased frequency in the population.

This is the rational for using LD-based methods to identify signals of positive selection in the genome, where the goal is to look for extended regions of strong LD (that is for long haplotypes) assuming that an allele associated with unusually long-range LD and high population frequency might be a signature of positive selection, otherwise recombination and mutation would have broken down LD, shortening the extent of the haplotype block.

The simple and earliest measure to evaluate the strength of LD is commonly symbolized by *D*, as originally proposed by Lewontin and Kojima (1960). *D* quantifies the deviation between the observed frequency of a two-locus haplotype and its expected value assuming random association between alleles. Meanwhile, several alternative measures based on *D* have been devised such as *D'* (Lewontin 1964) and r^2 (Hill and Robertson 1968), among others (Ardlie, et al. 2002; Slatkin 2008).

One of the most used LD-based methods to test for selection relies on the calculation of the Extended Haplotype Homozygosity statistic (EHH), where Extended Haplotype Homozygosity from a core region (hypothetically under selection) is defined as the probability that two randomly chosen chromosomes carrying the same core haplotype of interest are identical by descent, as assayed by homozygosity at all SNPs (Sabeti, et al. 2002). A core haplotype with unusually high frequency and high homozygosity that extends over large regions is indicative of a mutation that rose to elevated frequency faster than expected under neutral evolution, which likely represents a sign of a selective sweep. The method identifies tracts of homozygosity within a 'core' haplotype (by other words, extended haplotypes transmitted without recombination), using the EHH statistic, which ranges from 0, when all haplotypes are different, to 1, when complete homozygosity results in extended haplotypes all identical (Sabeti, et al. 2002).

Concerning the number of SNPs to include in the EHH analysis, the optimum amount is very variable because it depends on SNP density and on levels of *LD*, two characteristics that vary much across the genome (Gattepaille and Jakobsson 2012).

Besides EHH, other haplotype-based statistics have been implemented that were demonstrated to exhibit high power to detect positive selection over a large range of allele frequencies (Voight, et al. 2006; Barreiro, et al. 2008) as well as to be insensitive to background selection, which is accepted to be a confounder factor that differentially affects the haplotypes sharing the ancestral or derived allele (Fagny, et al. 2014).

In more recent years, a number of methods based on identity-by-descent (IBD) analyses have also been developed to detect several types of selection (Albrechtsen, et al. 2010). Although EHH and IBD-based approaches fundamentally rely on the same principles and both take advantage on the same patterns in genomic regions to look for selection, IBD-based approaches have greater power to detect selection on standing variation then EHH-based methods (Albrechtsen, et al. 2010).

1.3.2.2.3 Tests of neutrality

The distribution of allele frequencies across polymorphic sites in the genome, known as the Site Frequency Spectrum (SFS) is a staple of population genetics, having been commonly used to reveal broad patterns of selection (Williamson, et al. 2005; Boyko, et al. 2008) and to make inferences on demographic history (e.g. population expansions, bottlenecks, or migrations) (Marth, et al. 2004; Williamson, et al. 2005). The SFS represents the basis of many neutrality tests that permit to assess the goodness-of-fit of the standard neutral model on specific data sets. Tests of selective neutrality have been developed for samples either drawn from single populations or from multiple populations. Most of them are based on the comparison of some of the summary statistics of the SFS in a data set to their expected distribution according to standard models of population genetics, i.e., the Wright–Fisher model and related ones (Achaz 2008; Vitti, et al. 2013).

When a selected allele and its nearby hitchhiker region sweep toward fixation, this will shift the distribution of alleles in the population. So, as a consequence of the sweep it will occur a population-wide reduction in the genetic diversity around the selected locus. New mutations appearing in this homogeneous background will be necessarily rare because they have emerged recently in the population. This excess of rare variants segregating at the nearby sites of the selected variant will be the hallmark of the selective force, which will create a shift in the SFS.

Although many statistical approaches have been developed to detect this sign in the SFS, the first and one of the most widely used tests to distinguish between a DNA sequence evolving according to the neutral mutation theory and one evolving under a non-neutral process is Tajima's *D* (Tajima 1989), which is considered a fairly simple and robust test, though only applicable when dealing with a single population.

Tajima's *D* is the standardized difference between two estimates of theta ($\theta=4N e\mu$), one of which is the Watterson (1975) estimator defined as the number of segregating sites (*S*) in a sample of n sequences, and the second is the mean pairwise differences between sequences (π) in the sample (Tajima 1983). In broad terms Tajima's compares the average number of pairwise differences between individuals with the total number of segregating sites in a population, and might assume negative or positive values, both corresponding to departures from the neutral expectations. Under the standard neutral model, Tajima's *D* is expected to be near zero. Given that low frequency alleles contribute less to the number of pairwise differences in a data set than do alleles of moderate frequency, a significant positive value of Tajima's *D* reveals a relative excess of intermediate frequency alleles, as expected under a model of population subdivision or an old balanced polymorphism, whereas a significant negative value of *D* reflects a relative excess of rare variants, as might be expected after a population exponential growth or after a "selective sweep" in which a new or rare variant was favored and quickly rose towards fixation in the population (Przeworski, et al. 2000).

Based on a similar approach to Tajima's *D*, other neutrality tests have been proposed among which the most popular are the Fu and Li's *D* statistic (Fu and Li 1993) and the Fay and Wu's H test (Fay and Wu 2000). Fu and Li's *D* shares many characteristics with Tajima's *D*: a negative value indicates an excess of singletons, which would also give a negative Tajima D, and a positive value indicates a relative deficiency of singletons, which typically (though not necessarily) would also give a positive Tajima *D* value.

In contrast to Tajima's *D*, which measures departures from neutrality that are reflected in the difference between low-frequency and intermediate frequency alleles, Fay and Wu's *H* measures departures from neutrality that impact in the difference between high-frequency and intermediate-frequency alleles. Thus, while *D* is very sensitive to population expansion because the number of segregating sites responds more rapidly to changes in population size than the nucleotide heterozygosity, *H* is less sensitive, but also powerful to detect recent positive selection (Zeng, et al. 2006).

It is worth noting, that since this family of tests relies on assumptions about the demographic history of the populations in which the data sets were obtained, the interpretation of the results they provide can be much challenging. Each test applied individually rarely affords unambiguous evidence of selection, and most typically only the combination of different tests may allow to distinguish population expansions, bottlenecks or other demographic events from different modes of selection (Zeng, et al. 2006). As before mentioned, departures from the neutral expectations might also be consistent with a variety of events that do not involve selection at all.

1.4 Human Adaptations

After the first spread out of Africa around 100 Kya, modern humans started occupying almost all the remainder world's landmasses, having encountered new environments and ecosystems. In addition, over the last thousand years, conditions within Africa also have changed. All these alterations fueled the idea that there has been ample scope for the action of natural selection in recent human evolution (Coop, et al. 2009).

In the last years, many studies emerged investigating the role of human adaptations in molding differences between individuals and/or between population groups that is in shaping the patterns of population variation across the genome.

However, demonstrating clearly that selection has operated on a given gene is a difficult task because it is necessary to prove simultaneously that patterns of variation at the gene are not consistent with neutrality, that there is a functional difference between specific alleles, and in addition that such functional difference would result in a phenotypic effect that could provide any selective advantage or disadvantage.

Despite the efforts undertaken, up to now the successful demonstration of the three aspects, at least to some degree, has been only achieved in few cases, provided by studies essentially coming from candidate gene approaches where strong biological hypotheses were implicit.

Amongst the most consensually recognized examples of recent human adaptations are included genes involved in resistance to *vivax* malaria (Ko, et al. 2011; Ralph and Coop 2013), skin pigmentation (Khan and Khan 2010; de Magalhaes and Matsuda 2012) and lactose intolerance (Tishkoff, et al. 2007; Breton, et al. 2014; Macholdt, et al. 2014; Ranciaro, et al. 2014; Macholdt, et al. 2015). In all these cases, studies began with the hypothesis that a known phenotypic trait is adaptive, and subsequently genetic evidence for selection was discovered.

More recently, taking advantage of the progress in genomic technology that made possible the examination of genome-wide and whole-genome sequencing data, studies of human adaptation tended to move from specific genes, known to be involved in certain traits, towards looking for signatures of selection across the genome (de Magalhaes and Matsuda 2012; Enard, et al. 2014).

Many genome-wide scans for selection have been undertaken, resulting in the report of hundreds of *loci* or genomic regions showing signals of positive selection, with patterns of variations characteristic of advantageous mutations that have spread quickly through populations (Akey, et al. 2002; Bustamante, et al. 2005; Voight, et al. 2006; International HapMap, et al. 2007; Sabeti, et al. 2007; Williamson, et al. 2007; Akey 2009; Pickrell, et al. 2009). Still, moving from candidate genomic regions to the underlying adaptive mutation has been difficult, and in fact few previously unsuspected adaptive traits and variants have been elucidated by GWS (Grossman, et al. 2013).

Even so, large-scale scans of positive selection have showed various functional groups of genes enriched in positive selection signatures, including those involved in sensory perception, immunity, spermatogenesis, regulation of transcription or neural functions (Clark, et al. 2003; Nielsen, et al. 2005; Arbiza, et al. 2006; Gaya-Vidal and Alba 2014).

Gene ontology analyses also revealed that biological processes such as ‘Lipid and fatty acid binding’ or ‘Glycogen metabolism’ were enriched for genes located in candidate genomic regions shaped by selection, which as a whole was interpreted as an added evidence for local adaptation to food sources in humans, supporting thus the role of diet as a major selective pressure responsible for shaping genetic diversity in human populations (Hofer, et al. 2012).

1.4.1 The Neolithic dietary changes

It is generally assumed that adaptation to factors that change with time and/or vary from region to region, have played an important role in the evolution of human populations. However, due to the complex bases of genetic adaptations, characterized by successive and contradictory selective forces acting on the same genes due to changing conditions, it has not been easy to prove adaptive explanations in humans (Balaresque, et al. 2007).

Amongst the factors more commonly considered to have triggered adaptations are included climate, pathogens and diet. The two first exerted obvious selective pressures, and represent those for which the earliest and perhaps the strongest evidence were obtained (Hancock, et al. 2008; Coop, et al. 2009; Coop, et al. 2010; Hancock, et al. 2011). Yet, dietary adaptations were also unequivocally identified in the human genome.

Food represents one of the most important environmental factors for humans, who throughout history have undergone several dietary transitions. Local adaptations to regionally specific dietary components might have been critical to sculpt the human genome, and consequently to lay some of the bases for differentiation of human populations (Ye and Gu 2011).

A remarkable dietary transformation in human history emerged with the domestication of plants and animals in the beginning of the Neolithic, roughly 10-12 Kya years ago. Then, besides the unprecedented increase in population density, humans were confronted with new infectious diseases, new environments, and new food sources (Figure 5) (Sabeti, et al. 2002). It was during the Neolithic Period that a dramatic change in human subsistence pattern began to occur with the transition from nomadic lifestyles to those more sedentary associated with pastoral and agricultural societies (Scheinfeldt and Tishkoff 2013).

The past 10 Kya have thus been some of the most interesting times in human history, being considered the period when many important genetic adaptations and disease resistances arose (Sabeti, et al. 2002).

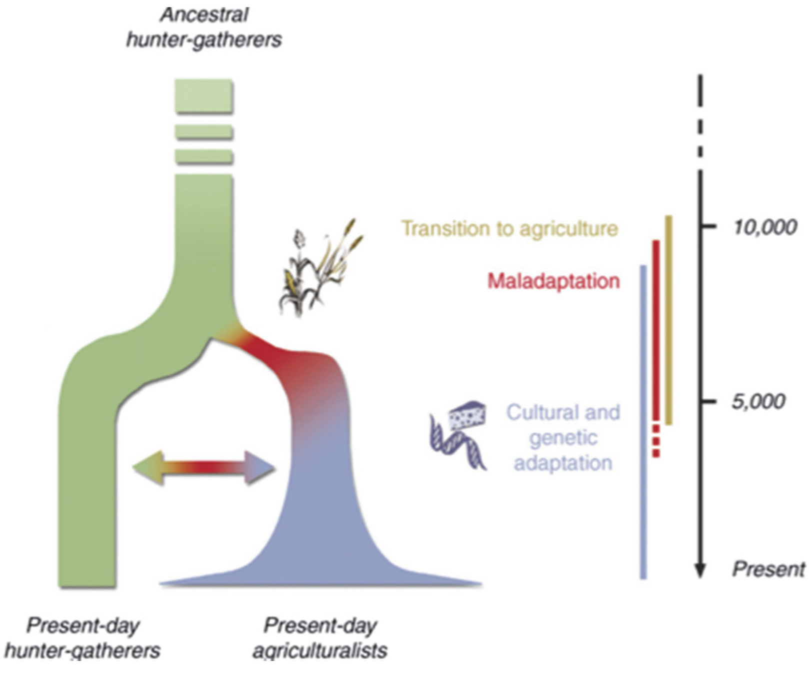

Ancestral hunter-gatherers
Transition to agriculture
Maladaptation
Cultural and genetic adaptation
10,000
5,000
Present
Present-day hunter-gatherers
Present-day agriculturalists

Figure 5. Human evolutionary history in terms of lifestyle, population size and diet.
Note that this diagram does not imply that hunter-gatherers have not continued to evolve until the present day. We simply know much less about the environmental changes to which they have had to adapt (Patin and Quintana-Murci 2008).

Although agriculture developed independently at multiple locations, archeological and genetic studies support that the earliest Neolithic culture appeared in the Near East followed in China, Americas and later in West Africa (Figure 6). Multiple animal domestication events occurred around the same time, with evidence of the first pastoral societies dating back to 10-12 Kya ago in North Africa and West Asia (Hanotte, et al. 2002).

After that, the dietary regimen of hunter-gatherer, which was the traditional economy up to the Neolithic transition, were slowly replaced by other afforded by the new food resources brought by the agricultural and pastoralism practices.

According to many authors, these changes in subsistence patterns were accompanied by a burden of diseases that nowadays constitute a serious public health problem. Among them is type 2 diabetes, concerning which as early as 1960's it was proposed the "thrifty genotype hypothesis" as an evolutionary explanation for its current very elevated incidence (Neel 1962).

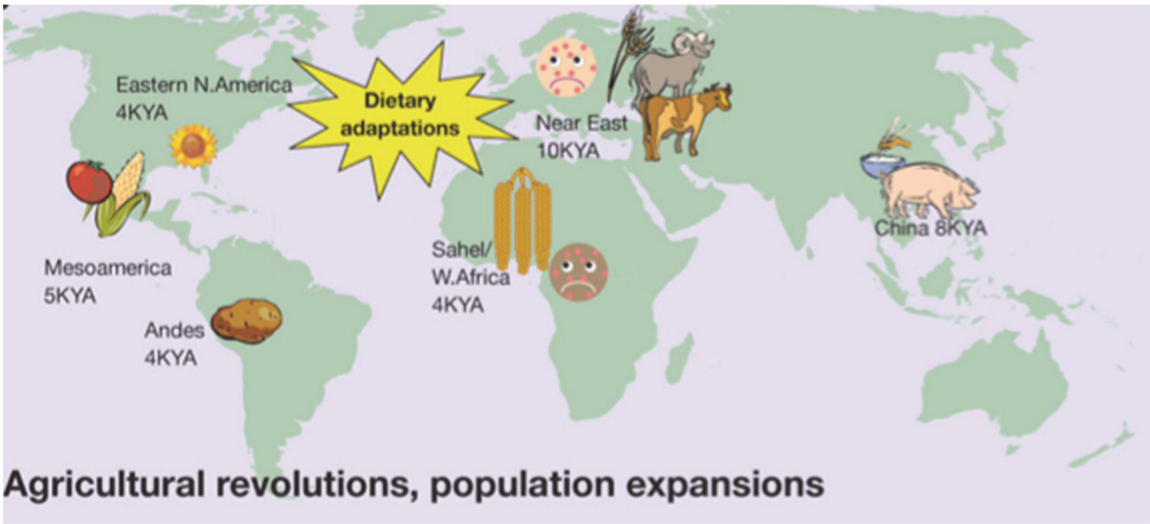


Figure 6. Agricultural revolution leading to population expansions and new disease challenges. Symbols indicate a selection of locally domesticated plants and animals (Balaresque, et al. 2007).

In pre-Neolithic times, subsistence of human populations was based on fishing, hunting wild animals and gathering wild plants. Comparatively to the diet of hunter-gatherers, that of farmers had much reduced nutritional intake diversity probably because controlled breeding was only possible for few plants and animals (Patin and Quintana-Murci 2008). Besides less diverse food intake, farming populations adopted an energy-rich and carbohydrate dense diet composed principally of cereals, fatty meats dairy products, and starchy root vegetables (e.g. potatoes and yams). The use of alcoholic beverages and salt is also associated to farmer dietary customs (Patin and Quintana-Murci 2008). Since humans were not accustomed to the new nutrients (e.g. maltose from starch, lactose from milk, gluten from wheat, alcohol, etc.), or at least to the substantially increased amounts of some nutrients introduced in the changed diet, it seems likely that some adjustments were necessary in the enzymatic machinery to metabolize those substances (Patin and Quintana-Murci 2008).

1.4.1.1 Adaptations in enzymatic pathways involved in food metabolism

Some recent studies have already shown that the Neolithic dietary modification exerted indeed a selective pressure strong enough to modify the expression profiles of digestive enzymes. Among them is salivary amylase, the enzyme responsible for digestion of starch, which is a key component of the diet of farmers. The amylase gene (*AMY1*), whose number of copies varies highly between individuals, encodes salivary amylase. In 2007 Perry, et al. (2007) demonstrated that individuals from populations with high-starch diets had on average more *AMY1* copies than those with traditionally low-starch diets, sustaining the idea that higher numbers of copies of *AMY1* have evolved in response to a shift towards diets containing more starch since the Neolithic times (Perry et al. 2007).

Lactase-phlorizin hydrolase, commonly referred to as lactase, is another digestive enzyme showing a genomic signature consistent with a post-Neolithic dietary adaptation. The ability to digest lactose, the sugar in milk, rapidly declines after weaning in most humans, due to the decreased expression of the gene encoding for lactase, the enzyme that breakdowns lactose. However, individuals from northern Europe and pastoralist populations from Africa, the Arabian Peninsula and central Asia with a tradition of fresh milk production and consumption retain in adulthood what is known as “lactase-persistence” trait, indicating that it represents a recent adaptive trait in humans related with the spread of pastoralist communities (Ranciaro, et al. 2014).

The two genes above mentioned, *AMY1* and *LCT*, are now widely seen as the best examples of diet-related genes that adapted rapidly to human dietary changes, retaining strong signs of the evolutionary forces that shaped their current pattern of genetic diversity across populations.

However, other studies interrogating additional genes involved in the metabolism of farming-derived nutrients, have also indicated that differences in subsistence strategies and diets between human populations worked as selective pressures during evolution, though exerting relatively subtle effects in most of the genes identified as adaptive. Of note the work of Hancock, et al. (2010), one of the most comprehensive studies up to now performed, which was based on a genome-wide scan aimed at identifying targets of adaptations to diet, mode of subsistence, and ecoregion. Some interesting signals detected in the work were adaptations induced by dietary specializations, with the overall data suggesting that genetic adaptations to dietary differences in human populations may be rather widespread. It was identified, for instance, a significant correlation between diets poor in folates and genetic variations lying in two genes that codify enzymes from the acid folic pathway: methionine synthase *MTRR* and *NAT2*. Since a remarkable reduction in folates availability in food was presumably subsequent to the subsistence and nutritional Neolithic shift, those findings were interpreted as evidence for genetic adaptations experienced by foraging (i.e., hunting and gathering) and nonforager (i.e., pastoral and agricultural) populations (Hancock, et al. 2010).

In the same study, the most convincing signal of an adaptation to a dietary specialization has come from a variation within the gene *PLRP2,* which codes for pancreatic lipase-related protein 2, an enzyme responsible for the breakdown of galactolipids present in the cellular membranes of plants. The allele resulting in a more active enzyme was found to be more common in populations relying in diets with high content in cereals (farmers), indicating therefore that diversity at *PLRP*2 might have evolved as an adaptive response to a specialized diet (Hancock, et al. 2010).

A major difficulty in studies aiming to evaluate whether a relationship exists between a genetic variant and a diet-related variable is to obtain adequate biological samples from human populations well defined ethnologically for which detailed nutritional information (obtained from field observations) is available (Patin and Quintana-Murci 2008). Actually, it might be quite misleading to use a shared lifestyle as synonymous of a shared traditional diet, because differences in lifestyles do not necessarily involve radical changes in dietary habitudes (Schoeninger 2001). On the other hand, an assumed contemporary lifestyle might not mean an ancient lifestyle typical of a population, since there are known cases of human groups that only recently adopted a given subsistence mode (Oota, et al. 2005).

1.4.2 Taste Perception

The selective pressures related to the changes in food availability and diet composition during human evolution probably have acted over multiple biological processes. In the search for those that might be candidate adaptations to human dietary changes, taste comes out as a logic target, because it is a major determinant of food selection (Ye and Gu 2011). Other perceptions of food, like vision and olfaction, also influence the dietary preferences, but unlike vision and olfaction, which function in diverse behavioral contexts, the sense of taste is the dominant regulator and driver of feeding behavior (Yarmolinsky, et al. 2009).

From an evolutionary perspective, in mammals, particularly in omnivorous species due to the variety and complexity of the feeding strategy, the sense of taste serves two principal functions: a) enable the evaluation of the toxicity of food, contributing to avoid the hazards of accidental toxin ingestion, while helping individuals to choose food based on the content in nutrients; and b) initiate physiological reflexes that prepares the body to metabolize and absorb foods once they have been ingested (Chaudhari and Roper 2010; Breslin 2013).

In present day humans, it is know that variations in taste preference partially arise from genetic differences and may have important consequences for food selection, nutrition, and health (Shigemura, et al. 2009).

Given that, the sense of taste is an important guide to ingest food, it is clearly of interest to understand whether the inclusion of new food in dietary habitudes during human evolution, also worked as a selective pressure to promote changes in taste perception, increasing and/or decreasing sensibility to specific tastes.

In humans, taste *stimuli* are commonly categorized into five “so-called” basic taste modalities- salty, sour, sweet, *umami*, and bitter, although additional qualities such as fatty, metallic and others were also proposed to must be considered basic tastes (Chaudhari and Roper 2010).

Taste *stimuli* interact with taste receptors (Figure 7), which function as chemoreceptors that initiate an afferent signal transmitted to the brain resulting then in a taste perception.

There are multiple types of taste receptors that account for the molecular recognition and cellular processing that occurs in specific cells from the oral taste buds. Importantly, however, many of the proteins that underlie transduction for sweet, bitter, and *umami* tastes are also expressed in sensory cells lining the stomach and intestine (reviewed in Chaudhari and Roper (2010)).

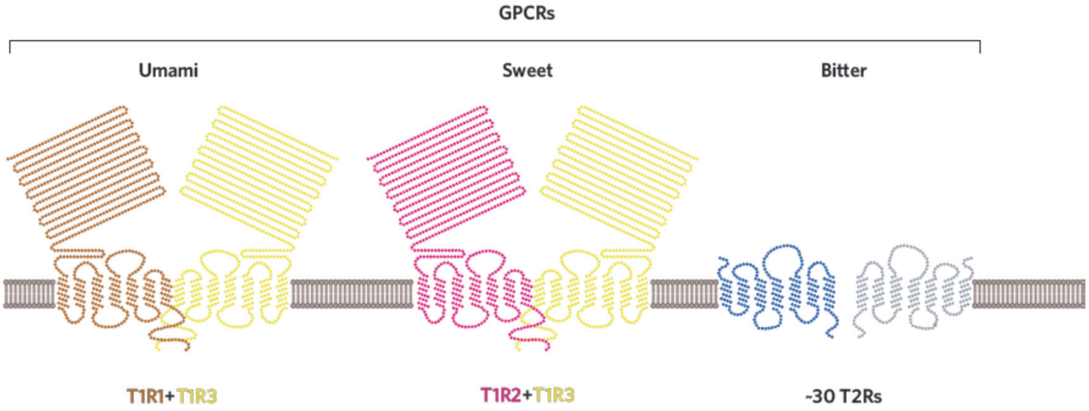


Figure 7. Taste receptors for *umami*, sweet and bitter tastes. Adapted from Chandrashekar, et al. (2006).

1.4.2.1 Bitter

Of the different taste sensations, bitterness was the first taste response known to be influenced by genetic factors. The discovery dates back to the early 1930's when it was realized that people could be either "taster" or "non-taster" of bitter compounds like phenylthiocarbamide (PTC) and propylthiouracil (PROP) (Blakeslee 1932; Fox 1932). The perception of bitter is mediated by a wide family of taste receptors known as TAS2Rs, which belongs to the class of G protein-coupled receptors that comprise the largest protein superfamily in mammalian genomes with the characteristic seven domains spanning the plasma membrane. The encoding genes, *TAS2R*s, experienced family expansion by tandem gene duplication during mammalian evolution (Bachmanov and Beauchamp 2007). Up to now 38 *TAS2R*s functional genes and 5 pseudogenes were annotated in human genomic databases, located in 3 clusters on chromosomes 5, 7 and 12 (Bachmanov and Beauchamp 2007). Because most human *TAS2R* genes exhibit elevated diversity of coding sequence, it is commonly presumed that *TAS2R* genetic polymorphisms must account for the individual differences in bitter taste sensitivity, but in truth such relationship has only demonstrated for a few cases (Bachmanov and Beauchamp 2007). One of them is *TAS2R38,* which codes for the receptor that has as ligands PTC and PROP (Campbell, et al. 2012), the two before mentioned synthetic compounds that elicit bitter taste. Variations within *TAS2R38* were clearly proven to influence the ability to taste or not those compounds (Behrens, et al. 2013; Robino, et al. 2014; Sharma and Kaur 2014), playing also a role in the perception of glucosinolates, a kind of bitter-tasting compounds widely found in many plants (Gorovic, et al. 2011).

Another well-established relationship is between the receptor encoded by *TAS2R16* and the ability to detect β-d-glucopyranosides, which are naturally occurring bitter compounds characterized by toxic cyanogenic activity that are also present in a wide range of plants. Polymorphic variations at *TAS2R16,* known to confer strong differential response to β-d-glucopyranosides such as the salicin derived from the willow bark (Campbell, et al. 2013), likewise have been suggested to influence preference for foods or beverages containing naturally occurring bitter compounds (Imai, et al. 2012).

1.4.2.2 Sweet

The ability to perceive sweet taste is a common trait in a range of animals that probably reflects the importance of simple carbohydrates as a dietary energy source (Kare 1984). For humans, the most common natural taste stimuli described as sweet are sugars, which are important nutrients for animals from many different species ranging from insects to mammals (Bachmanov, et al. 2011).

In addition to sugars, there are other chemicals (commonly referred to as sweeteners) that also evoke the sensation of sweetness in humans and are palatable to many other animals (Bachmanov, et al. 2011).

In mammals, sweetness perception is originated when sweeteners interact with taste receptor proteins from the T1R family, known as TAS1R in humans, expressed in taste receptor cells in taste buds of the oral cavity (Bachmanov, et al. 2002). This family also belongs to the superfamily of G protein–coupled receptors.

The *TAS1R* gene family contains 3 genes, *TAS1R1, TAS1R2 and TAS1R3*, which are located in a cluster on chromosome 1.

In humans, sweet taste is largely mediated by a heterodimer receptor composed of *TAS1R2* plus *TAS1R3* proteins. Both components of the *TAS1R2-TAS1R3* dimer entail an extracellular domain, called the Venus Flytrap Module (VFTM), containing a ligand-binding pocket that responds to a large variety of sweet substances, such as caloric sugars (glucose, fructose and sucrose), sweet amino acids and synthetic dipeptides (e.g., aspartame). Natural and artificial sugars (e.g., sucrose, glucose, and sucralose) bind to the VFT domains of both *TAS1R2* and *TAS1R3*, whereas dipeptide sweeteners bind only to the TAS1R2 VFT domain (Fernstrom, et al. 2012).

Besides *TAS1R2-TAS1R3*, other sweet taste receptors are supposed to exist but its number is still unresolved (Garcia-Bailo, et al. 2009).

1.4.2.3 *Umami*

Umami is a Japanese name that means “good flavor" or "good taste”, which is used to describe the savory sensation produced by meaty, savory flavor. It was just only in the beginning of the 20th century that *umami* joined the group of basic tastes, which from then on were the five qualities: sour, salty, bitter, sweet and *umami* (Ikeda 2002).

The main substance eliciting *umami* taste in humans is the amino acid *L*-glutamate, typically as its sodium salt monosodium glutamate (MSG) that is commonly present in many food including cheese, meat or mushrooms. Some purine nucleotides such as 5'-ribonucleotides guanosine 5'-monophosphate (GMP) and inosine 5'-monophosphate (IMP) also elicit the *umami* taste (Toda, et al. 2013).

At the molecular level, the *umami* processing seems to be quite related to that of sweet taste, with the perception of *umami* being also mediated by G-protein-coupled receptors of the *TAS1R* family. For now, the best known *umami* receptor is a heteromer composed by TAS1R1 and TAS1R3 proteins (Garcia-Bailo, et al. 2009).

Although variability in human *umami* taste perception still remains poorly understood, it has been demonstrated that genetic variations in both genes encoding the *TAS1R1-TAS1R3* heterodimer directly affects *umami* taste sensitivity (Shigemura, et al. 2009; Raliou, et al. 2011).

1.5 Africa as a Case Study

Africa is the world's second largest continent in size after Asia, covering an area of about 30,328,000 km², 6% of the Earth's total surface and 20.4% of the total land area. However, it is the most compact continent in terms of shape, measuring approximately 8,050 km from north to south as well as from east to west. It is surrounded by the Mediterranean Sea to the north, both the Suez Canal and the Red Sea along the Sinai Peninsula to the northeast, the Indian Ocean to the southeast, and the Atlantic Ocean to the west. Another distinguishing feature of the African continent is the fact that the Equator runs almost exactly through its middle, giving it a markedly symmetrical arrangement of climatic and flora conditions in both its northern and southern portions ranging from equatorial and tropical forest through tropical savannahs and deserts to Mediterranean conditions (Boateng 1978).

1.5.1 People and Ethnicity in Africa

The demographic history of the African continent is very complex, having covered fluctuations in population size, migration, admixture and extensive population structure. The occurrence and duration of some of these demographic events were often correlated with known major environmental changes and/or cultural developments (Campbell and Tishkoff 2010).

Currently, the continent contains a population of over 1,100,000,000 inhabitants accounting for about 15.5% of the world's human population (2013 World population data sheet - http://www.prb.org/). Africa remains the least urbanized region in the world, although currently facing an historic period of demographic change. In fact, while in the beginning of the 1990-decade two thirds of the population lived in rural areas, in less than two decades the proportion of the continent-wide population residing in settlements classified as cities grew to the 40%. The corresponding prevision for 2050 is to be around 60% in a 2,000,000,000 inhabitants landscape.

Nowadays, the population is geographically very unevenly distributed and there are significant intra-regional urbanization differences: North Africa and South Africa regions have the highest urbanization rates of the continent, contrasting with the East Africa region that remains as the least urbanized region of the world (Programme and Africa 2008; Programme and Programme 2010).

In terms of ethnicity, Africa is characterised by an extraordinary high level of ethnic diversity. However, in its origins, development and meaning to different people at different times, ethnicity is a complex issue (Keese 2010).

The complexity of the concept of ethnicity is reflected in the difficulty to describe its meaning. According to the Encyclopaedia Britannica *ethnicity* refers to the identification of a group based on a perceived cultural distinctiveness that makes the group into a "people", involving aspects such as language, music, values, art, styles, literature, family life, religion, ritual, food, naming, public life, and material culture.

It is the sharing of such kind of characteristics that permit to define in the broad sense ethnic groups, whose number in Africa, although still debatable, has been estimated at around 2,000 (http://www.ethnologue.com).

1.5.2 Lifestyles

The classification of cultures of human societies into basic subsistence systems is still a common practice in anthropology today due to its usefulness for understanding human cultural diversity. The focus on economic differences proved to be useful since the main characteristics of a culture are directly related to its economy.

Modes of subsistence are based on the use of different sources to obtain food and other needs, involving also the technology and work arrangements required to effectively capture and utilize resources (Haviland, et al. 2010).

Cultures have been divided into two basic subsistence types: hunter-gatherer and food-producing societies. The hunter-gatherers rely on food-foraging, which is the oldest and most universal human adaptation (in the sense of an adjustment of an organism to its environment) and typically involves geographic mobility (Haviland, et al. 2010). Foragers are those peoples who base their subsistence completely or predominantly by some combination of gathering, collecting, hunting, fishing, trapping, or scavenging the resources available in the plant and animal communities around them, without domestication (Panter-Brick, et al. 2001). Food-producing societies began to emerge with the domestication of plants and animals around 10,000 years ago. Agriculture or farming is the activity consisting in the cultivation of animals, plants, *fungi*, and other products used to sustain human groups. It is known as horticulture when practiced in a small scale. A branch of agriculture is the pastoralism subsistence in which people make their living by rising livestock. But while horticulture led to settlements that are more permanent, pastoralism required mobility to seek out pasture and water. Currently in Africa, there are human groups that depend on different types of economies.

Among the African hunter-gatherer groups are the Dorobo in Tanzania and Kenya; !Kung in Namibia, Botswana and Angola; Pygmies in Rwanda, Burundi, Uganda, the Democratic Republic of the Congo, the Republic of Congo, the Central African Republic, Cameroon, the Equatorial Guinea, Gabon, Angola, Botswana, Namibia, Madagascar, and Zambia; San in South-Africa.

Some of the pastoralism societies in sub-Saharan Africa are the Karimojong in Uganda or the Maasai in Tanzania and Kenya.

Agricultural societies are, however, predominant throughout the entire continent. Their establishment in sub-Saharan Africa was associated to massive migrations and settlements of Bantu-speaking people, as will be presented below.

1.5.3 African languages and Bantu expansion

Another side of the ethnic diversity present in Africa is reflected in the high number of languages spoken in the continent: over 2,000 indigenous languages besides many creoles, pidgins and *lingua francas* (Epstein and Kole 1998; Heine and Nurse 2000). Excluding these last languages, which were introduced over the past two millennia, the indigenous ones can be classified into four large phyla (Greenberg, et al. 1990) (Figure 8):

- Niger-Congo, 1,436 languages (including the large Bantu family);
- Afroasiatic, 371 languages (some of them spoken exclusively in the Middle East);
- Nilo-Saharan, 196 languages;
- Khoisan, 35 languages.

Khoisan or click languages were once widespread over much of the southern and eastern Africa. However, currently less than 100,000 people who live in marginal areas of Botswana, South Africa, Namibia, and Tanzania speak them (Shoup 2011).

Niger-Congo phyla is the largest in Africa both in number of speakers, over 400 million, and geographical distribution. The Niger Congo group is divided into six major family groups: Kwa, Mande, Voltaic, Atlantic, Bantu and Adamawa (Shoup 2011). The Bantu group, which contains more than 500 languages, is the most widespread of the Niger-Congo sub-branches and one of the largest of the world. According to linguistics, genetics and archaeology, the population dispersal of Bantu speakers from its presumed home area in West Africa into central, eastern and Southern Africa began around 3,000-5,000 years ago (Gann and Duignan 1972).

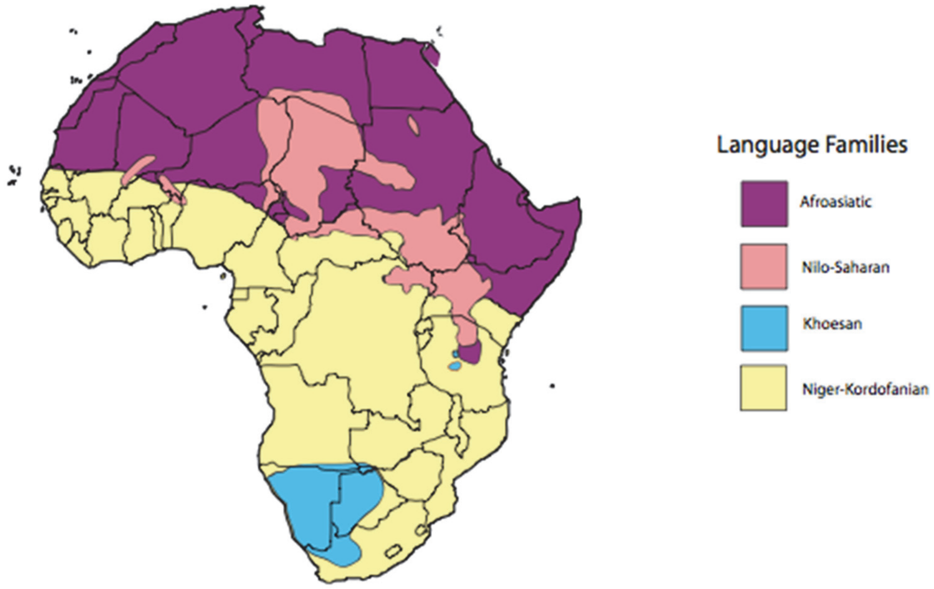


Figure 8. Map of Africa colored by the language family spoken in each region (Scheinfeldt, et al. 2010).

Due to the absence of written records, there is no certainty regarding the homeland of Bantu speakers, often assumed to be located it in the border between Nigeria and Cameroon (Hombert and Hyman 1999), or the routes taken by Bantu groups in their spread over the continent. However, there is no doubt that the Bantu migrations had important demographic and cultural significance, since they brought technologies in metalworking and agriculture that supported large, settled populations, whereas hunter-gathering and herding peoples were forced out of productive agricultural areas into more marginal zones such as arid steppes, deserts and mountains.

Concerning the Bantu expansion, two main dispersal scenarios have been proposed: i) The “early split“ model (Figure 9a) of Eastern and Western branches, posits that the East branch travelled north of the rainforest as far as the Great Lakes, approximately 4000 years ago, before heading south to cover the eastern half of Africa, while the West branch expanded through the rainforest, emerging on the south side to cover the western half of the continent; and ii) The ”late split” model (Figure 9b) argues that there was an initial movement through the rainforest and a later divergence of the Eastern branch from the branch of the other side approximately 2 Kya (Pakendorf 2011; de Filippo, et al. 2012).

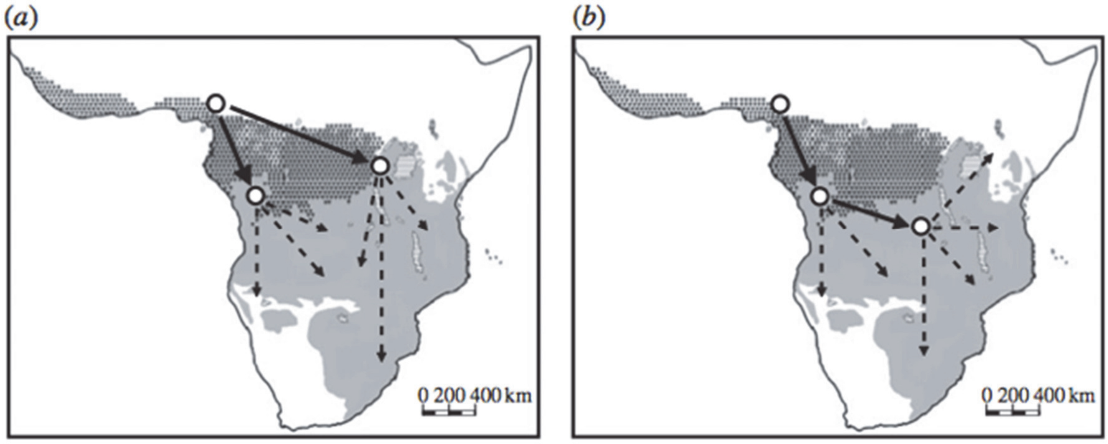


Figure 9. The two main models of Bantu migrations.
The current area occupied by Bantu languages is shaded in grey, and the extent of the rainforest is indicated by the darker shading. (a) Early-split and (b) late-split (de Filippo, et al. 2012).

1.5.4 The importance of genetic studies in Africa

Africa represents the ancestral homeland of all humankind, the oldest populated continent, and the source of millions of people of the recent African diaspora. The continent harbours a very heterogeneous ethnic composition (Attah-Poku 1998) and high levels of genetic diversity within and between populations.

Because of the pivotal role of African populations in human history, characterizing their patterns of genetic diversity and *linkage disequilibrium* is recognised to be crucial for reconstructing human evolution and for understanding the genetic basis of complex diseases (Tishkoff and Verrelli 2003). Despite that, Africa remained for long largely uncharacterized from the genetic point of view, and only recently African populations begun to be more systematically scrutinized [for a review see Campbell and Tishkoff (2010) and Tishkoff and Williams (2002)], regarding different genetic issues such as:

i) Patterns of Genetic Variation: Higher levels of genetic diversity have been detected in African populations compared to non-African ones, with analyses of haploid (mtDNA and Y-chromosome), X-chromosomal or autosomal genetic markers. The first results date back to the 80 decade of the last century (Johnson, et al. 1983; Papiha, et al. 1991; Lucotte, et al. 1994; Paabo 1995), and were recently confirmed by studies based on genome-wide data (Elhassan, et al. 2014);

ii) Population Structure: Strong substructure was identified in extant African populations, particularly between hunter-gatherers and other populations. A considerable amount of genetic diversity may had been already present at an early stage of modern human evolution, as indicated by studies suggesting that ancestral populations were geographically structured before the out of Africa migration of modern humans (Gunz, et al. 2009; Schlebusch, et al. 2013; Elhassan, et al. 2014);

iii) Migration and Admixture: It was demonstrated that one of the most significant demographic events in the recent African history was the geographic expansion of Bantu-speaking populations (Gomes, et al. 2010; Barbieri, et al. 2014; Russell, et al. 2014);

iv) Disease patterns: While non-communicable diseases represent the major health concern in the developed world, infectious or communicable diseases are still the most prevalent diseases in the developing world, including sub-Saharan Africa. Accordingly, the search of genetic variations that affect susceptibility to infectious diseases like cholera, AIDS, and many others, has been an important topic of research in African populations (Mengel, et al. 2014; Skelton, et al. 2014). In addition, the efforts to characterize the genetic structure in Africa have also in mind the implications of biomedical relevance.

CHAPTER 2. OBJECTIVES

On a worldwide scale, there is substantial phenotypic variation among human populations, implying a plasticity that in a certain extent might have arose through processes of adaptive evolution. After the emergence of modern humans in Africa, human populations have expanded to occupy a vast range of habitats, facing so new climates, pathogens and food resources, having also diversified in terms of subsistence modes. All these changes make it possible that there has been ample scope for the action of selection during human evolution. However, much uncertainty still persists on the role of selection in shaping global or local patterns of human genetic diversity, especially as regards selection driven by diet-related pressures.

Currently, African populations still remain among the most understudied major human groups, thereby hindering the analysis of factors that have accounted for the extant distribution of human genetic diversity.

Given this, the principal aims of this work were:

- To characterize African populations relying in distinct modes of subsistence for polymorphisms in genes involved in food metabolism and taste perception;
- To look for evolutionary factors that have modeled patterns of diversity at those genetic markers in African populations;
- To investigate in a wider geographical context whether metabolic adaptations emerged as response to the diversification of dietary regimens that occurred after the Neolithic;
- To search for signatures of selection through the analysis of levels of population differentiation at specific “diet-related” *loci* comparatively to genome wide levels of differentiation;
- To examine sequence diversity at genes involved in taste perception, in order to obtain insights on the factors that have underlain the evolution in human populations of these important but still understudied genes.

CHAPTER 3. RESULTS

3.1 “Exploring the relationship between lifestyles, diets and genetic adaptations in humans”

BMC Genetics, 2015

(doi: 10.1186/s12863-015-0212-1)

3.1.1 Introduction

The most remarkable dietary change over the recent history of human populations was that associated with the change from food collection to food production (Luca, et al. 2010), which occurred independently and in different times in separate parts of the world marking the beginning of the Neolithic, a transition that in some regions dates back to 12,000 years ago. The domestication of plants and animals prompted the conditions that would brought about new modes of subsistence as well as new food habits as a consequence of the shift in the availability and exploitation of dietary resources (Patin and Quintana-Murci 2008; Luca, et al. 2010). Genetic adaptations to dietary specializations are thought to have represented advantageous evolutionary solutions in humans, however it is still unclear the extent to which dietary factors have created selective pressures acting on genes that play roles in food-related metabolic pathways. Recent studies have revealed genomic signatures of adaptations likely driven by diet-related pressures (Balaresque, et al. 2007; Luca, et al. 2010; Lachance, et al. 2012). In addition, candidate genes approaches had already provided tight evidence for genetic adaptations to differences in nutrient consumption such as at the lactase and amylase genes (Bersaglieri, et al. 2004; Perry, et al. 2007; Tishkoff, et al. 2007; Jones, et al. 2013; Liu, et al. 2013; Vitalis, et al. 2014).

Other metabolic-related genes have been hypothesized to constitute dietary adaptations, among which are included: *AGXT,* coding for alanine:glyoxylate aminotransferase, the enzyme responsible for the transamination of glyoxylate into glycine (Danpure 1997; Caldwell, et al. 2004; Segurel, et al. 2010); *PLRP2,* coding for pancreatic lipase-related protein 2, involved in galactolipids hydrolysis, (Lowe 2002; Sias 2004; De Caro, et al. 2008; Hancock, et al. 2010); *MTRR,* encoding for methionine synthase reductase, an enzyme acting in the complex folate pathway (Shaw, et al. 2009; Hancock, et al. 2010); *NAT2* coding for N-acetyltransferase 2, a phase-II enzyme involved in the detoxification of a wide number of xenobiotics (Patin, et al. 2006; Luca, et al. 2008; Magalon, et al. 2008; Sabbagh, et al. 2008; Hancock, et al. 2010; Sabbagh, et al. 2011); and *CYP3A5,* coding for cytochrome P-450 3A5, a member of the CYP3A enzymes that are involved in the oxidative metabolism of many endogenous substrates and xenobiotics, which is implied in sodium homeostasis (Thompson, et al. 2004; Young, et al. 2005; Thompson, et al. 2006; Zhang, et al. 2010).

Genetic variation in *AGXT* was tentatively linked with meat content in diets, *PLRP2* with richness in cereals (Hancock, et al. 2010), both *MTRR* and *NAT2* with availability of folate in foods and *CYP3A5* with health conditions that are influenced by dietary salt intake (Thompson, et al. 2004; Zhang, et al. 2010). However, for these five genes results so far obtained were either contradictory (e.g. *AGXT*), or not yet replicated (e.g. *MTRR* and *PLRP2*), or not clear enough to ascertain whether they can indeed represent genetic adaptations to any dietary variable. This prompted us to address the issue applying the same research strategy simultaneously to *AGXT*, *CYP3A5*, *MTRR*, *NAT2* and *PLRP2*.

Thus, assuming that current modes of subsistence are still good surrogates of main diets in which populations have traditionally relied, the aim of this study was to gain further insights into the relationship between diet-related variables in populations and patterns of diversity at variations in above mentioned five genes.

Functional variants within *AGXT*, *PLRP2, MTRR, NAT2* and *CYP3A5* were examined in six sub-Saharan populations with distinct modes of subsistence and also in one European population that was also screened to generate a non-African reference group. Results were then combined with previously published information for other African and Eurasian populations to evaluate the contribution of geography and mode of subsistence or other diet-related variables to explain the patterns of genetic diversity observed for the five genes.

3.1.2 Material and Methods

Samples and DNA extraction

A total amount of 361 individuals from six sub-Saharan African populations with different modes of subsistence were analyzed, i) 144 were farmers/agriculturalists from three populations: 32 from Angola (ANG), 82 from Equatorial Guinea (EQG) and 30 from Mozambique (MOZ); ii) 116 were herders/pastoralists from the Karamoja region in Northern Uganda (UGN); and iii) 101 belonged to two hunter-gatherer/forager populations: 39 Baka Pygmies from Gabon (BPY) and 62 Ju/hoansi from Tsumkwe, a small settlement in North Eastern Namibia (KNA). Forty-eight individuals from the Portuguese population (PTG) were additionally studied. Each population was assigned to a specific mode of subsistence and main diet component according to Murdock Ethnographic Atlas (http://lucy.ukc.ac.uk/cgi-bin/uncgi/Ethnoatlas/atlas.vopts), Encyclopedia of World Cultures, Africa: An Encyclopedia for Students (John Haley 2002).

From blood samples stored in FTA™ cards (Whatman), total DNA was extracted using a standard phenol-chlorophorm protocol (Sambrook 1989) in the sample from Equatorial Guinea, while in the remaining samples it was done as described in previous works: Angola (Beleza, et al. 2004), Mozambique (Alves, et al. 2001), Uganda (Gomes, et al. 2009), Baka Pygmies (Batini, et al. 2011), Khoisan (Marks, et al. submitted) and Portugal.

Genotyping

The screened SNPs were c.32C>T (rs34116584; NM_000030.2 *AGXT* gene), c.1074G>A (rs4751995; NM_005396.3 *PLRP2* gene), c.1130A>G (rs162036; NM_024010.2 *MTRR* gene), c.219-237G>A (rs776746; NM_000777.3 *CYP3A5* gene) and the following 4 variants in *NAT2* gene (NM_000015.2): c.191G>A (rs1801279), c.341T>C (rs1801280), c.590G>A (rs1799930) and c.857G>A (rs1799931). The last 4 SNPS, defining the alleles *NAT2*14*, *NAT2*5*, *NAT2*6* and *NAT2*7*, respectively, were selected in this study because reportedly they allow inferring the acetylator phenotypes with high accuracy (Hein and Doll 2012).

A pair of primers was designed for each SNP using Primer3 software ver. 4.0. Possible secondary structures or interactions between primers were checked with AutoDimer software ver. 1.0 (Vallone and Butler 2004) (Table S1). A multiplex Polymerase Chain Reaction (PCR) system was developed to co-amplify the regions containing the eight SNPs. The genotyping methodology was based in a multiplex minisequencing reaction, using the Single Base Extension (SBE) (Eap, et al. 2007) reaction kit (Applied Biosystems). SBE primers were designed and tested identically as described above. Poly tails of varying lengths were attached to the 5' end of each primer in order to avoid identical fragment sizes, allowing simultaneous typing of multiple variants in the same reaction (Table S2). SBE products were run on ABI 3130 Genetic Analyser (AB Applied Biosystems) and the electropherograms were analyzed using GeneMapper software ver. 4.0. (Applied Biosystems, Foster City, USA), based on fragment size inferred with GeneScan-120 size standard.

Chromosomal locations and genomic segments from the 5 genes were obtained using the latest version of the human genome assembly GRCh37 (http://www.ensembl.org/).

Statistical analysis

The Arlequin software ver. 3.5.1 (Excoffier and Lischer 2010) was used to estimate allele or haplotype frequencies, to test for Hardy-Weinberg Equilibrium (HWE) and to calculate genetic distances (F_{ST}). Regarding the *NAT2* gene, to account for *linkage disequilibrium* (LD) between SNPs, from the unphased multilocus genotypic data, haplotypic frequencies defined by the 4 SNPs were computationally estimated also with Arlequin software ver. 3.5.1. (Excoffier and Lischer 2010) and the acetylation phenotypes deduced from the pair of haplotypes carried by each subject (Hein and Doll 2012).

In order to determine whether means of allele frequencies were statistically different between groups of populations defined according to various criteria, One-Way ANOVA tests were performed. The Spearman Rank Correlation Score Corrected (SRCSC) was used to evaluate the correlation between allele frequencies and latitude. Both statistical analyses were performed on the website for statistical computation VassarStats (http://vassarstats.net).

The graphical representation of a F_{ST} distance matrix was constructed by means of the Multidimensional Scaling (MDS) procedure implemented in StatSoft, Inc. (2007), Statistica version 8.0 (www.statsoft.com).

To investigate possible signals of selection we obtained the neutral distribution of F_{ST}s conditional on heterozygozity based on genotypic data from a validated panel of 52 assumedly neutral SNPs for human identification (Sanchez, et al. 2006), using the 61 African and Eurasian populations contained in the SNP*for*ID browser, to draw a neutral dispersion cloud, 50,000 coalescent simulations of 100 demes were carried on Arlequin software. For variants at *AGXT*, *PLRP2*, *MTRR*, *NAT2* and *CYP3A5,* $F_{ST,}$ average heterozygosity within populations (*Het*) and the scaled heterozygosity *Het*/(1-F_{ST}) were also computed. The null sampling distribution of the empirical F_{ST} values was calculated using two distinct models, both implemented in Arlequin software ver. 3.5.1 (Excoffier and Lischer 2010): a) the classical Island Model at migration-drift equilibrium (IM), conventionally referred to as *F*dist approach, proposed by Beaumont and Nichols (1996), considering together all populations; and b) the Hierarchical Island Model (HIM) more recently recommended by Excoffier and collaborators (Excoffier, et al. 2009), for which populations were clustered in 5 groups (Europe, Africa, Middle East, South and East Asia) when a worldwide scale was assumed, or in 4 groups (Northern, Southern, Western and Eastern Africa) when only Africa was considered. Since F_{ST} strongly correlates with heterozygosity (Beaumont and Nichols 1996), the empirical *P*-value for each locus was calculated within bins of 20,000 SNPs grouped according to Minimum Allele Frequency (MAF), as the proportion of the bivariate probability distribution which was less probable than the estimated values, in the same way as calculated in the DFdist software package (Beaumont and Nichols 1996) .

Hierarchical Analysis of Molecular Variance (AMOVA) was performed in the Arlequin software ver. 3.5.1 defining population groups according to: i) mode of subsistence, ii) main diet component, and iii) geography in the context of Eurasia and Africa. Additionally for *CYP3A5,* populations groups were also defined according to latitude and location North or South the Tropic of Cancer.

Finally for *linkage disequilibrium* (LD) patterns analyses, the .ped files containing the data sets were first manipulated with gPLINK software (http://pngu.mgh.harvard.edu/purcell/plink/) (Purcell, et al. 2007) and then exported to Haploview software ver. 4.1 (Barrett, et al. 2005) to calculate *D'* (normalized *D,* where *D'*=*D*/*D*max measures the LD strength) and r^2 (squared correlation coefficient measure of LD between the two *loci*) parameters and visualize the LD plots, considering a window of 200, 300 and 500 Kb for *PLRP2*, *CYP3A5* and *NAT2* genes, respectively.

Comparative data

Viewing comparative analyses, data for other populations were retrieved from the database dbCLINE (http://genapps.uchicago.edu/software.html) from the Di Rienzo laboratory, which includes information from the International HaPMap Project Phase III (http://hapmap.ncbi.nlm.nih.gov/), the Human Genome Diversity Project Panel-Centre d'Etude du Polymorphisme Humain (HGDP-CEPH); and also we included data already published in other works (Bryc, et al. 2010; Henn, et al. 2011; Pagani, et al. 2012; Schlebusch, et al. 2012). The populations' abbreviations used in Figure 10 and 11 are MAN (Mandenka), NBT (North Bantu), MOB (Mozabite), SBT (South Bantu), LUH (Luhya), TIG (Tigray), AMH1 (Amhara1), AFA (Afar), YOR (Yoruba), AMH2 (Amhara2), ARC (Ari

Cultivator), WOL (Wolayta), ANU (Anuak), ARB (Ari Blacksmith), JPT (Japanese), NAX (Naxi), LAH (Lahu), MIA (Miaozu), TUJ (Tujia), SHE (She), YIZ (Yizu), CAM (Cambodian), HAN (Han), DAU (Daur), DAI, (DAI), BER (Italian from Bergamo), RUS (Russian), TSC1 (Tuscan1), SAR (Sardinian), FRC (French), BAS (Basque), TUSC2 (Tuscan2), ORC (Orcadian), ADY (Adygei), DRU (Druze), PAL (Palestinian) KAL (Kalash), PAT (Pathan), GUJ (Gujarati), BUR (Burusho), SIN (Sindhi), XIB (Xibo), GUM (Gumuz), SOM (Somali), MAS (Maasai), MON (Mongolian), YAK (Yakut), BAL (Balochi), MAK (Makrani), TU (Tu), UYG (Uygur), BED (Bedouin), BRA (Brahui), HAZ (Hazara), SAN (San), BIP (Biaka Pygmies), KUV (Kung Vasekela), MBP (Mbuti Pygmies) KHO1 (Khomani2), SAD (Sadawe), HAD (Hadza), HEZ (Hezhen), NYU (Naukan Yup'ik), ORO (Orogen) and MCH (Maritime Chukchee). More detailed information is summarized in Table S3.

3.1.3 Results

Locus by locus analysis

The observed genotypic distributions (Table S4) did not revealed significant departures from Hardy-Weinberg expectations after applied the Bonferroni’s correction for multiple tests. Estimates of allele frequencies for the five *loci* in the seven studied populations are shown in Table 1, and for each locus results here and previously obtained will be dissected in the following sections.

Table 1. Derived allele frequencies.

POP	**c.32C>T**	**c.1074G>A**	**c.1130A>G**	**c.191G>A**	**c.341T>C**	**c.590G>A**	**c.857G>A**	**c.219-237G>A**
	(*AGXT*)	**(*PLRP2*)**	**(*MTRR*)**	**(*NAT2*14*)**	**(*NAT2*5*)**	**(*NAT2*6*)**	**(*NAT2*7*)**	**(*CYP3A5*)**
ANG	0.0000±0.0000	0.3261±0.0691	0.5294±0.0856	0.1522±0.0530	0.2046±0.0748	0.3636±0.0725	0.0000±0.0000	0.2400±0.0604
EQG	0.0482±0.0166	0.3214±0.0360	0.3563±0.0363	0.0977±0.0225	0.3588±0.0536	0.1786±0.0296	0.0233±0.0115	0.1429±0.0270
MOZ	0.0370±0.0257	0.2333±0.0546	0.5500±0.0642	0.1429±0.0540	0.2500±0.0884	0.2857±0.0697	0.0000±0.0000	0.1167±0.0414
UGN	0.0727±0.0175	0.3945±0.0331	0.3835±0.0339	0.0699±0.0187	0.3902±0.0575	0.3085±0.0337	0.0055±0.0055	0.2336±0.0289
BPY	0.0147±0.0146	0.1891±0.0455	0.3846±0.0551	0.0263±0.0184	0.1842±0.0536	0.2568±0.0508	0.0000±0.0000	0.1447±0.0404
KNA	0.0000±0.0000	0.0242±0.0138	0.1371±0.0309	0.0000±0.0000	0.0656±0.0239	0.0484±0.0193	0.0968±0.0266	0.2097±0.0366
PTG	0.1915±0.0406	0.5106±0.0516	0.1383±0.0356	0.0000±0.0000	0.5000±0.0903	0.2021±0.0414	0.0532±0.0232	0.9022±0.0301

Populations’ abbreviations as referred in Material and Methods.

AGXT

In the *AGXT* gene, we studied the variant c.32C>T, concerning which the derived allele T had been previously suggested to play an adaptive role in populations traditionally relying in meat-rich diets (Caldwell, et al. 2004; Danpure 2006). The hypothesis was specifically investigated by Caldwell, et al. (2004) who reported on frequency data sustaining the model, a conclusion for which much accounted the observation of the highest frequency of the derived allele in the Sweden Saami, who have a long history of consuming high amounts of animal products (Caldwell, et al. 2004; Danpure 2006). Though, later, revisiting the question with a better coverage of Central Asian populations, Segurel, et al. (2010) failed to find increased allele frequencies across populations with diets richer in meat comparatively to those less meat rich, challenging this way the adaptive model proposed for the variation.

In this study, in terms of meat content in diets of African populations, it was assumed that in general farmers rely less in meat than pastoralists or hunter-gatherers, in accordance with a recent review from ethnographic compilations of hunter-gatherer diets indicating that animal food comprises their dominant energy source (Cordain, et al. 2002). Among the seven sub-Saharan populations examined, the frequency of the derived allele at c.32C>T ranged from 0 to 7.27% without showing any pattern of variation that could be connected with mode of subsistence or meat content in diets of populations. For instance, it was absent both from the farmers from Angola and from the Khoisan hunter-gatherers, although the first are representative of less meat consumers groups while the second are from more meat consumers ones. In the sample from Portugal, considered to be a farming population with a mixed diet reasonably balanced regarding animal and plant food resources, the derived allele reached 19.15%, a frequency higher than registered in any of the African populations regardless of its mode of subsistence or reliance upon meat.

To integrate the results in a more comprehensive distribution, data for c.32C>T was retrieved from the literature on populations for which information on the relative predominance of meat in their diets was available (Table S3). There were results only for populations from Africa and Eurasia, among which the average frequency of the derived allele was 0.081 across the set of populations assigned to have high meat consumption, while it was, 0.133, across the populations with low-meat consumption. Actually neither the overall differences in allele frequencies within the ”low-meat” and “high-meat” groups were statistically significant (*P*=0.0710, One-Way ANOVA), nor the trend in the frequency distribution sustained the hypothesis that the allele could be positively selected in meat-rich diet populations.

Furthermore, if the broad geographical distribution of c.32C>T in Africa and Eurasia conformed well the major population clusters commonly identified by random neutral genetic markers, intriguingly in Asia, where there is a high dispersion of gene frequencies, the extreme values were reported for two populations in rather close geographical proximity but with distinct traditional lifestyles: in the Tajiks, a group of sedentary agriculturalists from Western Tajikistan the derived allele was very well represented (26.9%), whereas in the Kazaks from Western Uzbekistan, who are traditionally nomadic herders whose diet mainly consists of meat, milk and dairy products, the allele only occurred marginally (1.7%).

From these analyses, no connection emerged between the frequency distribution of c.32C>T in *AGXT* and lifestyle of populations.

PLRP2

In this gene it was focused on c.1074G>A, a variant that causes a premature truncation of the pancreatic lipase-related protein 2 resulting in a more active version of the enzyme. In a very recent genome-wide scan for selection in human populations Hancock, et al. (2010) identified in this variant a convincing signal of adaptation to a dietary specialization, since the derived allele was found to be significantly more common in populations relying in diets with high content in cereals (farmers) than in other populations.

As long as it is known, the association was not further investigated except in the present study, where among the screened African groups, the derived allele was detected to be quite common in the three farmers' groups (23.3%-32.6%) as well as in the herders from Uganda (39.5%). Comparatively, the two hunter-gatherers groups showed lower frequencies, especially the Ju/hoansi (2.4%). The sample from Portugal showed the highest frequency in this study, 51.1% (Table 1).

As a whole, these results do not conflict with the hypothesis that the distribution of c.1074G>A might be related to the weight of cereals in diets, in the sense that at least within Africa, farmer populations tended to have higher frequencies of the derived allele compared to hunter-gatherers who rely less in cereals. These results were then put in a wide-ranging context, recruiting information on c.1074G>A for African and Eurasian populations from several sources, and maintaining the classification in populations that specialize and that do not specialize on cereals when originally presented (Table S3). As shown in Figure 10A and 10B, the frequency of the truncated allele was found to be more common across populations with cereal-rich diets (average frequency 35.8% in Africa; 49.1% in Eurasia + Africa) than across those less dependent on cereals (average frequency 22.7% in Africa; 22.9% in Eurasia + Africa), differences that were statistically significant either in Africa (*P*=0.0050, One-Way ANOVA) or in Eurasia + Africa (*P*=<0.0001, One-Way ANOVA). Comparing herders and hunter-gatherers, both integrated in the group of cereal less rich populations (Figure 10), mean frequency was respectively 41.3% and 18.6% in Africa, and 46% and 19.5% in Eurasia + Africa, with both differences being again statistically significant (*P*=0.0055 for Africa; *P*=<0.0001, for Eurasia + Africa, One-Way ANOVA). Considering Africa and Eurasia together, the trend that can be extracted from the whole data points to a decreasing frequency gradient of the derived allele at c.1074G>A from populations more specialized on cereals towards those less relying on them, as was also captured by the MDS plot shown in Figure 10C, where it is visible some structure between farmer, herder and hunter-gatherer populations.

Globally these results suggest that diversity at *PLRP2* was shaped by selective pressures that differed according to populations' lifestyle.

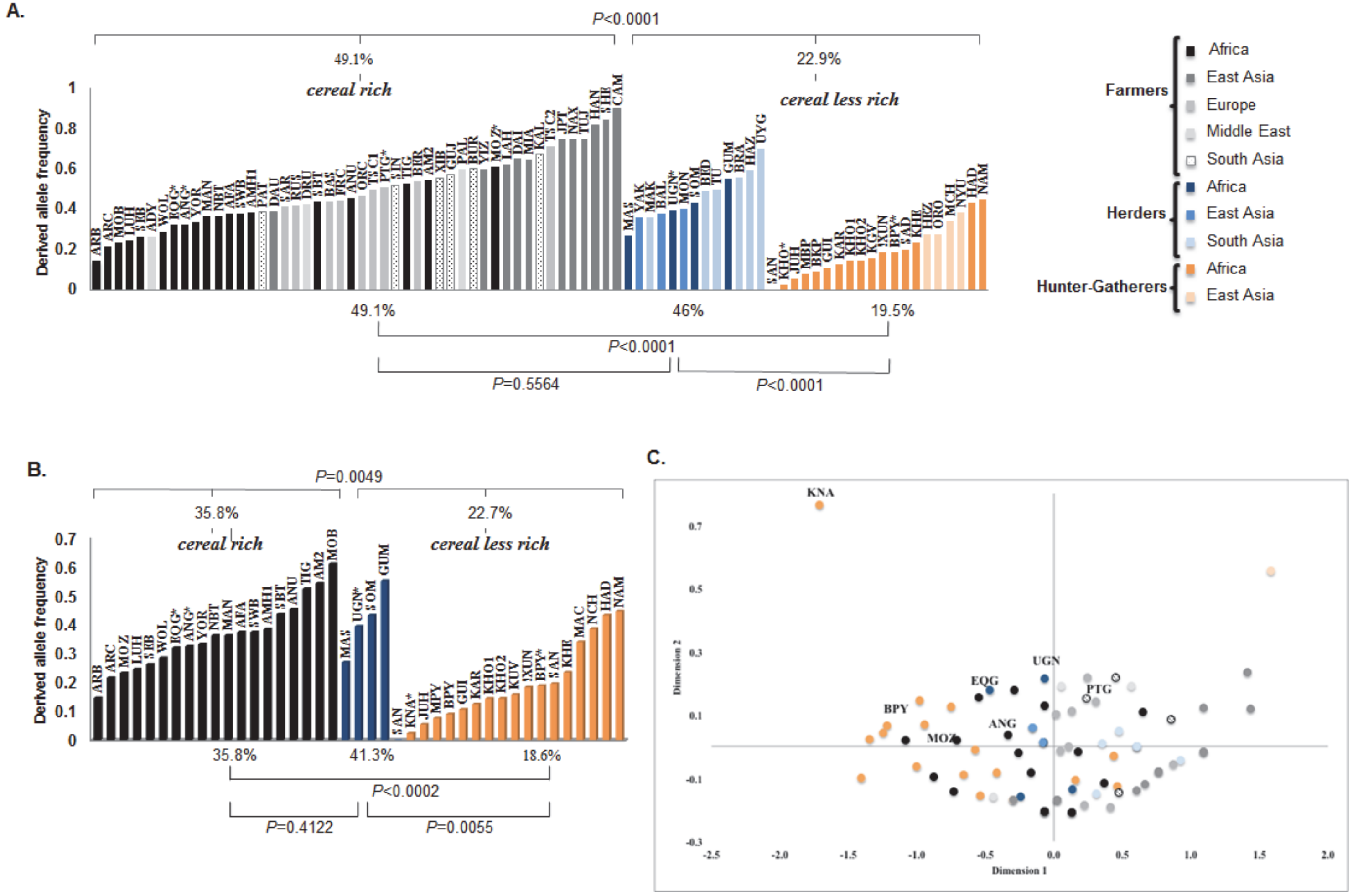


Figure 10. Allele frequencies and MDS plot for *PLRP2*.
P values of ANOVA One-Way test in (A) African + Eurasian and (B) African populations' group; MDS plot of pairwise genetic distances between populations (C). In the MDS plot different colors represent distinct lifestyles: hunter-gatherers (orange), herders (blue) and farmers (black). * Populations addressed in this study. Populations' abbreviations are referred in Material and Methods .

MTRR

Within *MTRR* we examined the common variation affecting levels of enzymatic activity c.1130A>G, since it was another candidate adaptive genetic variation identified in the before mentioned genome-wide study (Hancock, et al. 2010). Before, *MTRR* had received high attention in association studies, having been implicated, for instance, with risk for *spina bifida* (Shaw, et al. 2009). However, its adaptive role to dietary specializations was addressed in only one work where c.1130A>G was found to be strongly correlated with diets containing mainly the folate-poor foods roots and tubers (Hancock, et al. 2010). The results obtained in this work revealed that the derived allele was quite common in most African groups, peaking in the agriculturists from Angola and Mozambique with values of 0.529 and 0.550, respectively (Table 1). Both estimates are similar to that described in the Yoruba (0.548), the only African group with a diet principally relying on roots and tubers addressed in a previous study (Hancock, et al. 2010). So, at least in Africa high frequencies of this allele can be found in populations without having such a dietary specialization. Furthermore, no indication arose that the distribution of c.1130A>G could be related to the dietary availability in folates, which is generally thought to be lower in non-forager populations (pastoral and agricultural) than in hunter-gatherers (Luca, et al. 2008). In fact, in the hunter-gatherers Baka, in the herders from Uganda and in the farmers from Equatorial Guinea, the derived allele occurred at similar frequencies (0.385, 0.384, 0.356, respectively) despite the differences in mode of food production. In the hunter-gatherers Ju/honasi from Namibia, the allele occurred at the lowest frequency in Africa (0.137) but with a magnitude similar to that found in the European sample (0.138), considered as a representative of an agriculturalist society (Table 1). To interpret these results under a wide framework of African and Eurasian populations, frequency data were recruited once more from the literature (Table S3 from supplementary material), and the combined information allowed to realize that the distribution of c.1130A>G fitted well the pattern generally provided by neutral markers, not appearing to be influenced by the mode of subsistence or the relative folate content in diets of populations from Eurasia and Africa. In East Asia, for instance, the two highest values of the derive allele were present in the Tu (0.4), nomadic herders, and in the Hezhen (0.333), mainly hunters and fishers, but nonetheless in the foragers Orogen and Yakut, who also live in East Asia, the allele was absent or very rare (see Table 1).

So, for the variation c.1130A>G in *MTRR,* the current patterns of diversity do not indicate that it could represent an adaptation to mode of subsistence of human populations.

NAT2

The dietary availability in folates had also been previously hypothesized to be a modulator of genetic diversity at the gene that encodes for *NAT2* (N-acetyltransferase 2) (Luca, et al. 2008). Individuals can be classified in fast, intermediate or slow acetylator phenotypes, which are determined by the haplotypic composition defined by genetic variations at the *NAT2* locus. Evidence for the diet-related hypothesis provided by Luca, et al. (2008) was reinforced with the recent findings by the same people more recently (Luca, et al. 2010), based on a more comprehensive analysis of *NAT2* worldwide genetic diversity, that were also compatible with a model holding that the slow acetylator phenotypes were selectively favored in populations relying in dietary regimens with reduced folate supply, whereas the fast acetylators were neutral or even advantageous in the presence of folate-rich diets, as those thought to be fulfilled by hunter-gatherers. To extent the population coverage of previous works, frequencies of *NAT2* haplotypes and acetylator phenotypes were also estimated in this study (Table 2). The distribution of haplotypes was very heterogeneous across African populations, but in line with previous observations the prevalence of the slow acetylator phenotype in the two hunters-gatherers groups (Khoisan, 1.6%; Baka Pygmies, 13.5%) was significantly much lower than in the three agriculturalists groups or in the Ugandan pastoralists, all displaying values up to 37.4% (*P*=0.0139, One-Way ANOVA). In the Portuguese the slow acetylator phenotype accounted for the high proportion of 52.2%, which falls within the range typical from other European populations (Sabbagh, et al. 2011).

Table 2. Haplotype and phenotype frequencies for *NAT2* gene.

HAPLOTYPE	ANG	EQG	MOZ	UGN	BPY	KNA	PTG
ACAG	-	-	-	0.0056	-	-	-
ACGG	-	0.0091	0.0250	0.0064	-	-	-
ATGG	0.1053	0.0596	0.1250	0.0769	0.0135	-	-
GCGG	0.2894	0.3659	0.2250	0.3658	0.1622	0.0738	0.5000
GTAA	-	0.0150	-	0.0111	-	0.0010	-
GTAG	0.3947	0.1975	0.3000	0.300	0.2568	0.0482	0.1980
GTGG	0.2105	0.3117	0.300	0.2342	0.5676	0.7633	0.2604
GCGA	-	-	0.0250	-	-	-	-
GTGA	-	0.0350	-	-	-	0.1137	0.0417
ATAA	-	0.0063	-	-	-	-	-
Slow	0.5500	0.3735	0.4500	0.6750	0.1351	0.0164	0.5218
Intermediate	0.3000	0.5301	0.5000	0.2375	0.5135	0.3770	0.3478
Fast	0.1500	0.0964	0.0500	0.0875	0.3514	0.6066	0.1304

Populations' abbreviations as referred in Material and Methods.

Next, the present data was contrasted with other results before published for Eurasian and African populations (see populations on Table S3), confining the analysis to c.590G>A, which defines allele *NAT2*6*, because it was the variation with more information accumulated for populations representatives of the three modes of subsistence.

From Figure 11A, which shows the allelic distribution of c.590G>A across Africa and Eurasia, it becomes clear that its prevalence is scarcely influenced by the continent where populations are located. However, some connection arises with systems of food production and acquisition given that in the whole set of African and Eurasian populations foraging groups tended to exhibit statistically significant lower frequencies of the derived allele compared to populations dependent on agricultural and pastoral resources (see in Figure 10 the *P*-values of One-Way ANOVA). Between pastoralists and agriculturalists, no significant differences were detected, which means that the clustering of c.590G>A frequencies only showed correspondence with populations that are food producers or food collectors, an observation that otherwise fully meets that reported by Sabbagh, et al. (2011), and the results even more recent published by the same team (Patillon, et al. 2014).

In brief, the analyses reinforce previous indications that *NAT2* has evolved under a selective factor influenced by human diet.

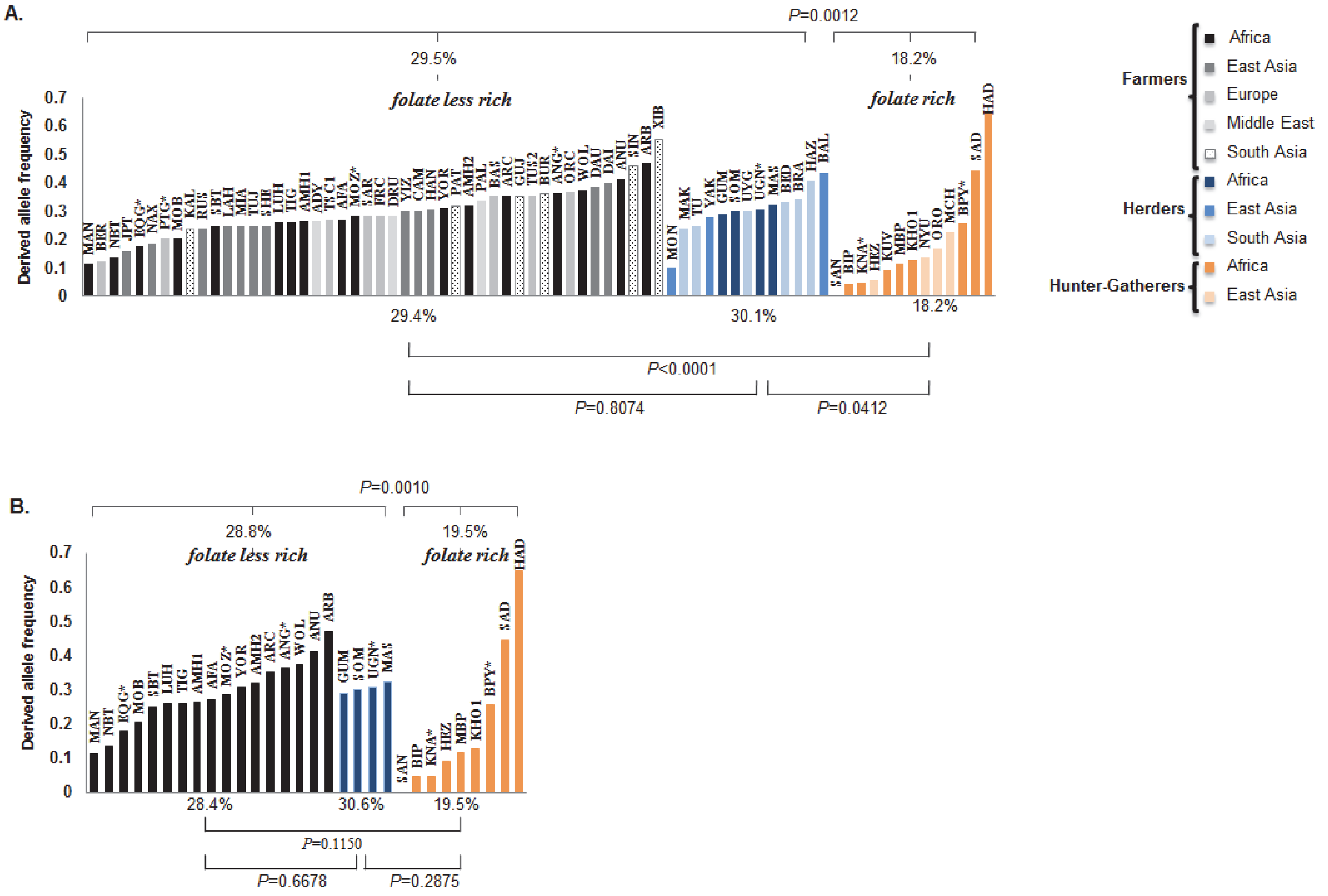


Figure 11. Allele frequencies for *NAT2*.
P values of ANOVA One-Way test in (A) worldwide and (B) African populations' group hunter-gatherers (orange), herders (blue) and farmers (black), * populations addressed in this study. Populations' abbreviations are referred in Material and Methods.

CYP3A5

With regard to *CYP3A5,* the intronic variation c.219-237G>A was screened, commonly referred to *CYP3A5*1/*3* polymorphism, in which the derived allele A results in a premature stop codon that reduces protein expression. It has been firmly demonstrated that the variation possesses a very unusual worldwide distribution whereby the frequency of *CYP3A5*3* is significantly correlated with latitude (Thompson, et al. 2004).

*CYP3A5*1/*3* likely influences salt and water retention and risk for salt-sensitive hypertension (Thompson, et al. 2004), exerting an effect on blood pressure that is determined by interactions with dietary salt intake (Eap, et al. 2007; Zhang, et al. 2010). Since anthropological evidence indicates that diet of hunting and gathering people is usually characterized by low level of salt intake, being often considered as a surrogate of the preagricultural humans' diet, lately praised as a model of well balanced food consumption (Eaton and Eaton 2000), we asked whether diversity at *CYP3A5*1/*3* could be related with diet of populations.

Thus, this variation was screened in six African populations, among whom the derived allele was only moderately represented, but suggestively it was in two farmer groups that the lowest and the highest frequencies were found (11.7% and 24.0% in the groups from Mozambique and Angola, respectively), disfavoring thus any link between lifestyle and differences in allele frequency across populations. In the Portuguese, the allele reached the very elevated value of 90.2%, which it is usual in populations from Europe where *CYP3A5*3* varies quiet narrowly being near-fixation in most populations (Thompson, et al. 2004). Again, the present data was combined with those retrieved from the literature (see Table S3), and with an enlarged coverage of African and Eurasian populations, confirming in fact that the frequency of the low expressor allele significantly increased with distance from the equator (Figure 12A) (SRCSC=0.7540; *P*<0.0001). When the relationship was assessed separately in each of the three continents, no significant rank correlation was observed in Africa (SRCSC=0.1058; *P*=0.2438) or in Europe (SRCSC=0.4183; *P*=0.1310), but in Asia the correlation coefficient was again statistically significant (SRCSC=0.5724; *P*=0.0002). Interestingly, in Asia, where the average allele frequency was 0.793, the significant correlation can be explained since the lowest values are consistently present in populations from the South of the continent, located very near or already inside the intertropical zone. In Africa, the frequency of the allele drastically declines to an average value of 0.286 when inferred from a panel of populations' majority located inside the tropical zone. In Europe, which is fully situated in a temperate climatic region, the average frequency reaches 0.903. Therefore, being or not located in the tropical zone seems to be a factor that strongly influences the distribution of *CYP3A5*1/*3* alleles (see Figure 12B). These analyses led to conclude that *CYP3A5* was the target of a selective factor determined by the geographic location of human populations.

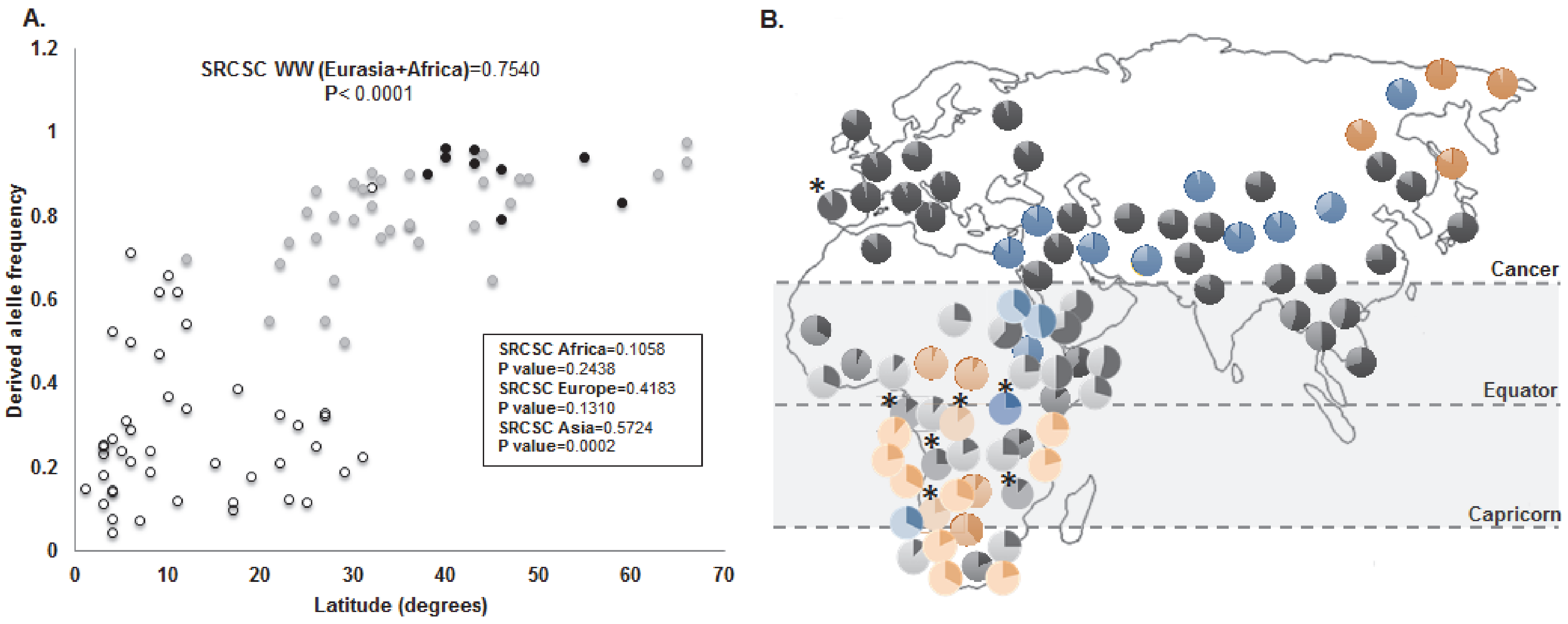


Figure 12. Distribution of *CYP3A5*3* in Africa and Eurasia and correlation with latitude.
Correlation plot between latitude and allele frequencies in African (open dots), European (black dots) and Asian populations (grey dots) (A). Map representing the distribution of *CYP3A5*3* across Africa and Eurasia (B): ancestral allele frequency (light pie) and the derived allele (dark pie); hunter-gatherers (orange pie), herders (blue pie) and farmers (black pie), * populations addressed in this study.

Hierarchical AMOVA

Hierarchical AMOVA was performed to determine the relative contribution of geography, mode of subsistence and different diet-related variables to the genetic structure observed in the SNPs at *AGXT*, *PLRP2, MTRR*, *NAT2* (only for that defining *NAT2*6*) and *CYP3A5*, hereinafter referred for simplicity as uniquely by their gene symbols (Table 3).

Geography was found to significantly account to explain the total genetic variance across Africa and Eurasia at *AGXT, PLPR2, MTRR,* and *CYP3A5*, but not at *NAT2*. The contribution of geography was especially high in *CYP3A5* in which it amounted to a very high proportion, 40.4% of total diversity. For this variation it was further assessed the effect of i) latitude and ii) the location North or South the Tropic of Cancer, leading to realize that for *CYP3A5* the highest value of F_{CT} (which measures the proportion of variance among groups) was achieved when populations north of the Tropic of Cancer were grouped against the southern ones, attaining then 44.9% of total diversity.

Concerning mode of subsistence, it was found to be a considerable modulator of diversity at *PLRP2*, explaining 8.8% of the total diversity at the locus, while also accounting to residual proportions of diversity at *NAT2* (1.6%) and *AGXT* (1.5%). When the criterion to group populations was the content in diets of cereals (for *PLPR2*), meat (for *AGXT*), folates (for *NAT2* and *MTRR*) or salt (for *CYP3A5*), significant F_{CT} values were only observed at *PLRP2,* in which the more or less reliance in cereals contributed to 6.5% of the total variance, and at *NAT2,* where differences in the dietary richness in folates explained 3% of the locus diversity.

Table 3. AMOVA analysis under different criteria.

	c.32C>T (*AGXT*)	*P*-value	c.1074G>A (*PLRP2*)	*P*-value	c.1130A>G (*MTRR*)	*P*-value	c.590G>A (*NAT2*)	*P*-value	c.219-237G>A (*CYP3A5*)	*P*-value
Mode of subsistence	**1.5**	**0.0355**	**8.8**	**0.0000**	-0.6	0.8793	**1.6**	**0.0046**	4.9	0.0542
Main diet component	0.4	0.1478	**6.5**	**0.0001**	-0.2	0.4712	**3.0**	**0.0019**	2.5	0.1004
Geography	**3.9**	**0.0001**	**8.5**	**0.0000**	**3.4**	**0.0003**	0.2	0.2149	**40.4**	**0.0000**
Latitude	-	-	-	-	-	-	-	-	**37.8**	**0.0000**
Above/below Tropic of Cancer	-	-	-	-	-	-	-	-	**44.9**	**0.0000**

Significant differences are highlighted in **bold**.

Signals of selection

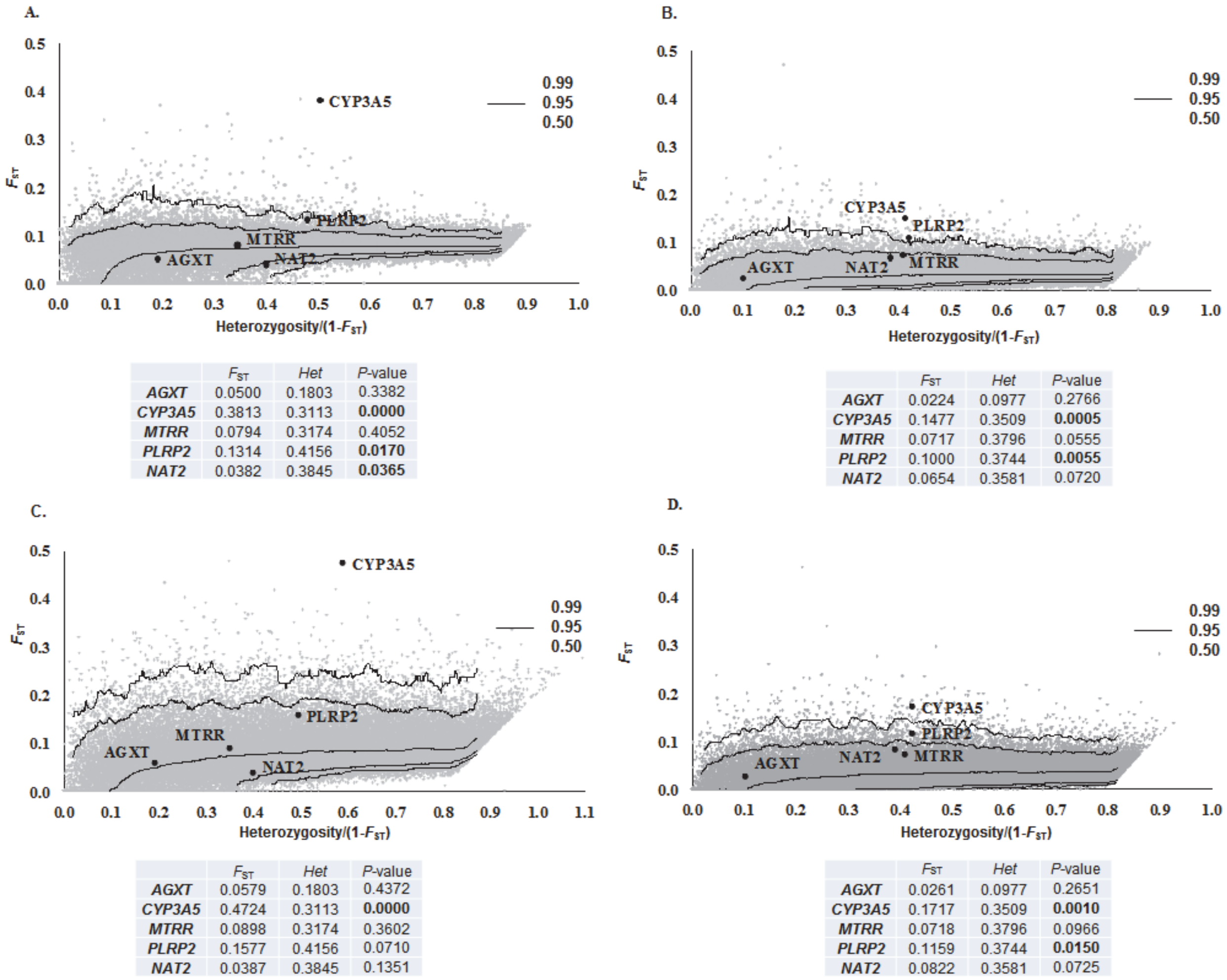
To dissect better whether from the levels of genetic differentiation across Africa and Eurasia signs of selection could be captured, it was used a conventional F_{ST}-based approach that assumes that genetic differentiation among populations is expectedly higher or lower for *loci* under directional or balanced selection, respectively, expected under neutrality.

Viewing that, it had firstly generated null sampling distribution of the empirical F_{ST} employing two different models, the finite Island Model (IM), which assumes the classical island model at migration-drift equilibrium (Beaumont and Nichols 1996); and the Hierarchical Island Model (HIM), in which populations samples are assigned to different groups, allowing for increased migration rates between populations within groups than between groups (Excoffier 2009). Besides portraying more realistically the demographic history of human populations, HIM was shown to produce a low rate of false positive signs comparatively to IM, when used to test *loci* for selection (Excoffier 2009).

The simulated null-distributions are presented in Figure 13 where are also shown the F_{ST} values plotted against scaled heterozygosity estimated for the SNPs at *AGXT*, *PLRP2, MTRR, CYP3A5* and *NAT2* (Excoffier, et al. 2009).

Considering simultaneously Africa and Eurasia and using as reference the IM distribution, the F_{ST}s for *MTRR* and *AGXT* did not differed significantly from the null expectations (Figure 13A). By contrast, the global differentiations at *PLRP2, NAT2* and *CYP3A5*, all lied outside the 95% confidence region of the neutral distribution, though showing departures with opposite directions: whereas the F_{ST} coefficient for *NAT2* was significantly smaller than expected, the coefficients for *CYP3A5* and *PLRP2* were both significantly larger (*P*-values in Figure 13A). The outlier position is especially remarkable in the case of *CYP3A5* that presented the exceedingly high F_{ST} coefficient of 0.381, almost five times greater compared to the average empirical neutral level of 0.079 between African and Eurasian populations. These results suggest that *NAT2* could have been under balanced or negative selection whist both *PLRP2* and *CYP3A5* might well have been modeled by positive selection. Taken into account the F_{ST} null distribution simulated under the HIM (Figure 13C), the F_{ST}s for *NAT2* and *PLRP2* lost the condition of significant outliers and the unique differentiation that remained significantly higher than the neutral expectations was at *CYP3A5.* Simulations were also carried out considering separately Africa and Eurasia.

While in Eurasia none of the five assessed SNPs revealed to be outsiders in the distributions simulated under the simple or the hierarchical island models (results not shown), noteworthy in Africa the differentiations at *PLRP* and *CYP3A5* were significantly higher than expected under the neutral expectations derived from the two demographical models (Figure 13B and 13D).



A.

	F_{ST}	*Het*	*P*-value
AGXT	0.0500	0.1803	0.3382
CYP3A5	0.3813	0.3113	**0.0000**
MTRR	0.0794	0.3174	0.4052
PLRP2	0.1314	0.4156	**0.0170**
NAT2	0.0382	0.3845	**0.0365**

B.

	F_{ST}	*Het*	*P*-value
AGXT	0.0224	0.0977	0.2766
CYP3A5	0.1477	0.3509	**0.0005**
MTRR	0.0717	0.3796	0.0555
PLRP2	0.1000	0.3744	**0.0055**
NAT2	0.0654	0.3581	0.0720

C.

	F_{ST}	*Het*	*P*-value
AGXT	0.0579	0.1803	0.4372
CYP3A5	0.4724	0.3113	**0.0000**
MTRR	0.0898	0.3174	0.3602
PLRP2	0.1577	0.4156	0.0710
NAT2	0.0387	0.3845	0.1351

D.

	F_{ST}	*Het*	*P*-value
AGXT	0.0261	0.0977	0.2651
CYP3A5	0.1717	0.3509	**0.0010**
MTRR	0.0718	0.3796	0.0966
PLRP2	0.1159	0.3744	**0.0150**
NAT2	0.0822	0.3581	0.0725

Figure 13. Distribution of F_{ST} vs. Scaled Heterozygosity expected under neutral models.
Joint distributions in African + Eurasian (A) and African populations (B) under Island Model (IM); and joint distributions in African + Eurasian (C) and African populations (D) under Hierarquical Island Model (HIM). It is represented the 99% confidence regions of the null distribution. Black dots represent the observed measures in the studied genes, referred for simplicity as uniquely by their gene symbols; significant differences after Bonferroni's correction for multiple tests are highlighted in **bold.**

LD Patterns

In order to assess whether the examined genetic variants were in fact those responsible for the selective signals detected *PLRP2, NAT2* and *CYP3A5,* the patterns of *linkage disequilibrium* (LD) surrounding were explored each of the three genes, viewing which a genomic window was considered that encompassed the adjacent genes. In Table 4 are presented the non-synonymous variants showing significant *D'* and r^2 values with our target SNPs, identified in African populations, which were the unique with genome data available.

Table 4. *Linkage disequilibrium* including *D'* and r^2 parameters.

Gene (target SNP)	SNP*	*D'*	r^2	Gene location	DNA changes	Functional consequences	Reference
PLRP2 (rs4451995)	rs2305204	1.0	0.036	*PLRP1*	c.1242G>C	-	Pagani, et al. (2012)
		NA	NA				Henn, et al. (2012)
		NA	NA				Schlebusch, et al. (2012)
	rs1049125	0.884	0.052	*PLRP1*	c.1382T>C	-	Pagani, et al. (2012)
		NA	NA				Henn, et al. (2012)
		NA	NA				Schlebusch, et al. (2012)
	rs4751996	0.981	0.962	*PLRP2*	c.1084G>A	-	Pagani, et al. (2012)
		NA	NA				Henn, et al. (2012)
		NA	NA				Schlebusch, et al. (2012)
NAT2 (rs1799930)	rs1801280	NA	NA	*NAT2*	c.341T>C	Associated with slow acetylator due to N-acetyltransferase enzyme variant (acetylation slow phenotype)	Pagani, et al. (2012)
		1.0	0.142				Henn, et al. (2012)
		NA	NA				Schlebusch, et al. (2012)
	rs1208	NA	NA	*NAT2*	c.803G>A		Pagani, et al. (2012)
		1.0	0.360				Henn, et al. (2012)
		NA	NA				Schlebusch, et al. (2012)

NA - data not available; * SNP non-synonymous

The correspondent LD plots for each gene across different African populations are present in Figures S1-S7, located in supplementary material. For *CYP3A5* and *NAT2* no significant LD was detected with neighbor genes. Within each of the two genes, high LD was only found between our target SNP at *NAT2* and the linked variants rs1801280 and rs1208, both associated with decreased enzyme activity like rs1799930. Although this makes it difficult to discriminate the effects of the three variants, which can be concluded that the selective signal detected at *NAT2* is related with variations that affect enzyme activity in a similar direction. As for the gene *PLRP2*, it was found to be located in a region of considerable LD with *PLRP1*, a downstream gene that codes for pancreatic lipase-related protein 1. Within *PLRP1* two non-synonymous (rs2305204 and rs1049125), whose functional consequences are unknown, are in strong LD with our target SNP at *PLRP2*, which in addition was at high LD with rs4751996, a non-synonymous substitution of unknown functional effect, also located in *PLRP2*.

3.1.4 Discussion

The analysis of patterns of human genetic diversity at wide geographical scales can disclose remarkable features difficultly explained by demographic events or pure neutral processes that rather might represent the first symptoms of environmental adaptations.

In this study a special attention to variations in *AGXT*, *PLRP2*, *MTRR*, *NAT2* and *CYP3A5* was taken, five genes assumedly involved in the metabolism of substances (including xenobiotics) that gain entry into the organism through dietary food stuffs, for which it has been previously posited that they could represent instances of gene-culture coevolution in humans (Sabbagh, et al. 2008; Hancock, et al. 2010; Segurel, et al. 2010).

Out of those genes*, PLRP2*, *NAT2* and *CYP3A5* were found to present signs in their distribution patterns evoking the action of environmental selective pressures, though of diverse nature and strength.

The most unequivocal signature of selection was associated with *CYP3A5* that displayed a level of inter-population differentiation dramatically surmounting even the most conservative neutral expectations. Contrarily to our starting hypothesis, however, the amount of salt presumed to be ingested across main dietary habits did not accounted for the distribution of *CYP3A5*, which instead was highly determined by the geographical location of populations in the North or in the South of the Tropic of Cancer. So, the analyses here undertaken fully support previous findings indicating that *CYP3A5*3* evolved under a selective pressure determined by an environmental factor correlated with latitude (Thompson, et al. 2004), but also add accuracy to the interpretation pointing toward a factor shared by regions located above or below the Northern Tropic. *CYP3A5* has been intensively explored in the context of the genetic factors contributing to hypertension susceptibility, known to vary widely across different human populations. Nearly 40 years ago Gleibermann (1973) proposed the "sodium retention" hypothesis, according to which the high rate of hypertension in certain populations could partially be due to a genetic background that was environmental adaptive, presuming that efficient salt retaining mechanisms might had been advantageous in the hot savanna climate where humans first emerged. More recently, it was argued that hypertension susceptibility was ancestral in humans, and that differential susceptibility arose due to distinct selective pressures after the Out-of-Africa expansion of modern humans (Zhang, et al. 2010). *CYP3A5* is being often quoted to address the evolutionary perspective of hypertension susceptibility, due to the demonstrated role of CYP3A5 enzymes in sodium homeostasis, even though the many studies that analyzed the relationship between *CYP3A5* genotypes and blood pressure/hypertension have provided quite inconsistent results (reviewed in Lamba, et al. (2012)). So, together with the clarification of the link between *CYP3A5* and blood pressure, future lines of research should pay more attention to the role of *CYP3A5* enzymes in the physiological processes related with thermoregulation and/or with neutralization of effects of sunlight exposure. In the highly heat stressful intertropical region, there is a regular need to deal with the threat of dehydration, which may raise complicated physiological responses in wet or dry climates under which the efficient control of heat loss likely differs. Interestingly, the involvement of *CYP3A5* in such responses seems to obtain support from the recent discovery of an osmosensitive transcriptional control of human *CYP3A4*, *CYP3A7*, and *CYP3A5* that revealed increased mRNA expressions under ambient hypertonicity (Kosuge, et al. 2007).

Concerning *PLRP2*, the explorations here undertaken led in essence to corroborate the findings of Hancock, et al. (2010), indicating that diversity at the locus is somehow connected with mode of subsistence in populations. In fact, the assessed truncated allele showed to be significantly more frequent in farmers comparatively to groups not relying in farming, with the general trend, inferred from the whole set of African and Eurasian populations, pointing to a clinal decrease in frequency from farmers, next pastoralists towards hunter-gatherers. In addition, the global differentiation at this variant fell outside the neutral expectations, except when the HIM model was used in the tests for selection in Africa plus Eurasia. Hancock, et al. (2010) have associated the worldwide distribution of *PLRP2* to the content in cereals in diets of populations, on the grounds of the important role of the protein encoded by *PLRP2* in plant-based diets once, unlike other pancreatic lipases, this enzyme hydrolyzes galactolipids, which are the main triglyceride component in plants (Hancock, et al. 2010). However, the recent demonstration that the truncated allele addressed in their (and in the present) study exhibits near absence of secretion makes it unlikely that the encoded product may contribute to plant lipid digestion in humans (Xiao, et al. 2011), which seemingly undermines the biological basis originally proposed. In the meanwhile, new insights arose on the physiological role of *PLRP2*, suggesting for instance a likely major influence in fat digestion in newborns (Berton, et al. 2007). This refreshed information opens new perspectives that deserve future investigation to clarify whether cereal content or other *PLRP2* substrate, or amount of substrate, differing in hunter-gatherers, herders and farmers is the factor that exerted the selective pressure contributing to shape the current pattern of *PLRP2* diversity in human populations.

Before, however, it is necessary to overcome the uncertainty raised by the presence of significant LD between *PLRP2* and *PLRP1.* The two genes encode lipases that show high sequence homology and that assumedly participate in dietary fat metabolism, although not being yet clarified their differences in substrate specificity. Consequently, for the moment is not possible to discriminate between which variations at *PLRP2* or *PLRP1* are the best candidates to be causative of the selective signals detected.

With respect to *NAT2*, in agreement with earlier studies (Luca, et al. 2008; Sabbagh, et al. 2011) it was also detected that the average frequency of a slow acetylator allele was statistically lower in hunter-gatherers than in food-producer populations, when Eurasian and African populations were taken as a set. Furthermore, the same slow acetylator variant, which is the widespread *NAT2*6* allele, revealed an unusual low level of geographical structure across Africa and Eurasia, indicating that it was subjected to drifting constrains that likely could arise under the action of a mode of selection resulting in such a homogeneous allele distribution. In the tests for selection involving *NAT2*, significant departure from the neutral expectation was captured when the simple IM demographic model was assumed, which can raise uncertainty on whether signs of more subtle selective pressures might mistakenly escape the stringency of the HIM. The adaptive evolution of *NAT2* has been supported by a number of different studies (Patin, et al. 2006; Luca, et al. 2008; Magalon, et al. 2008; Sabbagh, et al. 2008; Talbot, et al. 2010; Sabbagh, et al. 2011; Hein and Doll 2012), including the examination of *NAT2* sequence data whose patterns of diversity made it plausible that slow-acetylating variants have been subject to weak selective pressures (Patin, et al. 2006). However, it is yet to be clarified the nature of such pressures. Luca, et al. (2008) tentatively claimed that it could be related with the diminished availability of folates in diet brought with the shift from economies relying in hunting and gathering to those based on farming and herding of domesticates. In line with the hypothesis, very recently a significant correlation was reported between *NAT2* acetylator phenotypes and historical dietary habitudes in India with the slow acetylator prevalence being higher in regions where is higher the proportion of vegetarians populations (Khan, et al. 2013). A major problem with this folate-related model is the still non-demonstrated role of *NAT2* human enzymes in folate metabolism (Sim, et al. 2008). Endogenous substrates for human *NAT2* are not known, although being well established that *NAT2* catalyzes the acetylation of many xenobiotics (Sim, et al. 2008). Since the exposure to xenobiotic substances or to concentration of xenobiotics likewise must had altered along the change in diets experienced by producers of food resources, the role of this kind of substances in shaping diversity at *NAT2* deserves further attention.

In summary, added evidence was provided that diversity at *PLRP2* and *NAT2* harbor signatures of genetic adaptations that might have been triggered by the diversification of modes of subsistence and diets that human populations began to experience after the rise of the Neolithic. Before that, modern humans had started to leave their original homeland in Africa, traveling out of the continent to colonize all regions in the globe. They progressively reached a wide range of new environments, facing new selective pressures that may have contributed to shape human genetic diversity. In *CYP3A5*, the compelling correlation with regions North and South the Tropic of Cancer, makes it likely that it belongs to the yet not fully understood catalog of genetic adaptations triggered by environmental stresses. Furthermore, the genetic signatures that *CYP3A5* harbors seem strength enough to have been driven by very long-lasting selection.

Finally, it was not possible to confirm the hypotheses at stake that *AGXT* and *MTRR* could also be diet-selected genes, since their diversity patterns could be well reconciled with demographic history at least of African and Eurasian populations.

3.1.5 Conclusions

In this study, we found signs that *PLRP2, NAT2* and *CYP3A5*, three genes assumedly involved in the metabolism of substances (including xenobiotics) that gain entry into the organism through dietary food stuffs, can represent instances of gene-culture coevolution in humans. Concerning *PLRP2*, it is still needed to clarify whether the signal detected is not a hitchhiking effect of its neighbor *PLRP1*. In addition, it is also necessary to demonstrate which the biological mechanisms were, and the environmental factors involved as well as their interactions, to understand the nature of the selective pressures that contributed to shape current patterns of genetic diversity at those *loci*. Furthermore, functional studies are needed to demonstrate the putative biological impact of the variations assessed, which ultimately would also exclude that the detected signs of selection could be due to other variations in *linkage disequilibrium* with those that were here examined.

3.1.6 Supplementary Material

Table S1. Amplification primers sequences.

Gene	SNPs	Primer forward	Primer reverse	Size (bp)
AGXT	c.32C>T	TCCACCAATCCTCACCTCTC	CCCCACTCCTACCTGGTACA	288
PLRP2	c.1074G>A	CTGGAGGATTCAGAGCTTGG	CAAAGCAATCCTGATGTACCC	163
MTRR	c.1130A>G	CGTGATCTGCCCTAACAGTG	ACAGCATCAGGGCTGTTACC	135
NAT2	c.191G>A; c.341T>C; c.590G>A; c.857G>A	TAACATGCATTGTGGGCAAG	GAGTTGGGTGATACATACACAAGG	797
CYP3A5	c.219-237G>A	ACCACCCAGCTTAACGAATG	GGTCCAAACAGGGAAGAGATA	94

Table S2. Mini-sequencing primers sequences.

Gene	SNP	Minisequencing primer seq 5'>3'	Size (bp)	Detection	Mutation	Final size (bp)
AGXT	c.32C>T	GCTTGAGCAGGGCCTTG	17	A/G	C/T	18
PLRP2	c.1074G>A	ctgacaaGTATTTCTTTGGACAGGTTG	20	A/G	A/G	28
MTRR	c.1130A>G	tgactaaactaggtgccacgtcgtgaaagtctgacaaGAAAATAAAGGCAGACACAA	20	A/G	A/G	58
NAT2	c.191G>A	gtgaaagtctgacaaaactgactaaactaggtgccacgtcgtgaaagtctgacaaGACCTGGAGA**Y**ACCACCCACCC	22	T/C	A/G	78
	c.341T>C	ccacgtcgtgaaagtctgacaaCCTTCTCCTGCAGGTGACCA	20	T/C	T/C	43
	c.590G>A	tgacaaaactgactaaactaggtgccacgtcgtgaaagtctgacaaACTTATTTA**C**GCTTGAACCTC	21	A/G	A/G	68
	c.857G>A	ctaggtgccacgtcgtgaaagtctgacaaTGCCCA**M**ACCTGGTGATG	18	A/G	A/G	48
CYP3A5	c.219-237G>A	aagtctgacaaGGTCCAAACAGGGAAGAGATA	21	T/C	A/G	33

aactgactaaactaggtgccacgtcgtgaaagtctgacaa is the tale used by convention; in **bold** are highlighted the polymorphisms inside the sequencing primers.

Table S3. Detailed data on mode of subsistence, diet and geography of the populations used for comparative analyses.

			AMOVA groups classification					
	Population	**n**	***Mode of subsistence***	***Main diet component***	***Geography***	***Latitude***	***Intertropical zone***	**Reference**
c.32C>T (AGXT)	Angolan	29	agriculturalist	meat-poor diet	Africa	NA	NA	*this study*
	Mozambican	30	agriculturalist	meat-poor diet	Africa	NA	NA	*this study*
	Equatorial Guinea	87	agriculturalist	meat-poor diet	Africa	NA	NA	*this study*
	Nigerian	62	agriculturalist	meat-poor diet	Africa	NA	NA	Caldwell, et al. (2004)
	Ethiopian	69	agriculturalist	meat-poor diet	Africa	NA	NA	Thomas, et al. (2002)
	Portuguese	47	agriculturalist	meat-poor diet	Europe	NA	NA	*this study*
	British (North Wales)	82	agriculturalist	meat-poor diet	Europe	NA	NA	Caldwell, et al. (2004)
	Turkey (Anatolia)	88	agriculturalist	meat-poor diet	Europe	NA	NA	Thomas, et al. (2002)
	Armenian	73	agriculturalist	meat-poor diet	Europe	NA	NA	Caldwell, et al. (2004)
	Norwegian	76	agriculturalist	meat-poor diet	Europe	NA	NA	Weale, et al. 2001)
	Ashkenazi Jews (Germany)	73	agriculturalist	meat-poor diet	Europe	NA	NA	Thomas, et al. (2002)
	Indian (Mombai)	84	agriculturalist	meat-poor diet	South Asia	NA	NA	Thomas, et al. (2002)
	Tajiks (Fergana)	17	agriculturalist	meat-poor diet	East Asia	NA	NA	Segurel, et al. (2010)
	Tajiks (Gharm)	23	agriculturalist	meat-poor diet	East Asia	NA	NA	Segurel, et al. (2010)
	Tajiks (Pejinkent)	24	agriculturalist	meat-poor diet	East Asia	NA	NA	Segurel, et al. (2010)
	Tajiks (Douchambe)	26	agriculturalist	meat-poor diet	East Asia	NA	NA	Segurel, et al. (2010)
	Sichuan Chinese	86	agriculturalist	meat-poor diet	East Asia	NA	NA	Caldwell, et al. (2004)
	Ugandan	110	pastoralist	meat-poor diet	Africa	NA	NA	*this study*
	Kasaks (Karakalpakia)	30	pastoralist	meat-rich diet	East Asia	NA	NA	Segurel, et al. (2010)
	Mongolian	80	pastoralist	meat-rich diet	East Asia	NA	NA	Caldwell, et al. (2004)
	Kyrgyz1 (Narin)	20	pastoralist	meat-rich diet	East Asia	NA	NA	Segurel, et al. (2010)
	Kasaks (Buraka)	49	pastoralist	meat-rich diet	East Asia	NA	NA	Segurel, et al. (2010)
	Kyrgyz2 (Narin)	26	pastoralist	meat-rich diet	East Asia	NA	NA	Segurel, et al. (2010)
	Turkmen (Karakalpakia)	34	pastoralist	meat-rich diet	East Asia	NA	NA	Segurel, et al. (2010)
	Kyrgyz (Andijan)	32	pastoralist	meat-rich diet	East Asia	NA	NA	Segurel, et al. (2010)

Table S3 (continued)

	Karakalpaks	30	pastoralist	meat-rich diet	East Asia	NA	NA	Segurel, et al. (2010)
	Saami (Sweden)	34	pastoralist	meat-rich diet	Europe	NA	NA	Thomas, et al. (2002)
	Khoisan Namibia	62	hunter-gatherer	meat-rich diet	Africa	NA	NA	*this study*
	Baka Pygmies	39	hunter-gatherer	meat-rich diet	Africa	NA	NA	*this study*
	Khomani Bushman San (B)	35	hunter-gatherers	meat-rich diet	Africa	NA	NA	Henn, et al. (2011)
	Hadza (Tanzania)	20	hunter-gatherers	meat-rich diet	Africa	NA	NA	Henn, et al. (2011)
	Sandawe (Tanzania)	35	hunter-gatherers	meat-rich diet	Africa	NA	NA	Henn, et al. (2011)
c.1074G>A (PLRP2)	Angolan	29	agriculturalist	specialized in cereals	Africa	NA	NA	*this study*
	Mozambican	30	agriculturalist	specialized in cereals	Africa	NA	NA	*this study*
	Equatorial Guinean	87	agriculturalist	specialized in cereals	Africa	NA	NA	*this study*
	N. Bantu	12	agriculturalist	specialized in cereals	Africa	NA	NA	HGDP-CEPH
	Mandenka	24	agriculturalist	specialized in cereals	Africa	NA	NA	HGDP-CEPH
	S. Bantu	8	agriculturalist	specialized in cereals	Africa	NA	NA	HGDP-CEPH
	Luhya	88	agriculturalist	specialized in cereals	Africa	NA	NA	HapMap Ph3
	Mozabite	30	agriculturalist	specialized in cereals	Africa	NA	NA	HGDP-CEPH
	Yoruba	25	agriculturalist	specialized in cereals	Africa	NA	NA	HGDP-CEPH
	Portuguese	47	agriculturalist	specialized in cereals	Europe	NA	NA	*this study*
	Basque	24	agriculturalist	specialized in cereals	Europe	NA	NA	HGDP-CEPH
	Italian (Bergamo)	14	agriculturalist	specialized in cereals	Europe	NA	NA	HGDP-CEPH
	French	29	agriculturalist	specialized in cereals	Europe	NA	NA	HGDP-CEPH
	Orcadian	16	agriculturalist	specialized in cereals	Europe	NA	NA	HGDP-CEPH
	Russian	25	agriculturalist	specialized in cereals	Europe	NA	NA	HGDP-CEPH
	Sardinian	28	agriculturalist	specialized in cereals	Europe	NA	NA	HGDP-CEPH
	Tuscan1	8	agriculturalist	specialized in cereals	Europe	NA	NA	HGDP-CEPH
	Tuscan2	89	agriculturalist	specialized in cereals	Europe	NA	NA	HapMap Ph3
	Adygei	17	agriculturalist	specialized in cereals	Middle East	NA	NA	HGDP-CEPH
	Druze	48	agriculturalist	specialized in cereals	Middle East	NA	NA	HGDP-CEPH
	Palestinian	51	agriculturalist	specialized in cereals	Middle East	NA	NA	HGDP-CEPH

Table S3 (continued)

	Burusho	25	agriculturalist	specialized in cereals	South Asia	NA	NA	HGDP-CEPH
	Kalash	25	agriculturalist	specialized in cereals	South Asia	NA	NA	HGDP-CEPH
	Pathan	25	agriculturalist	specialized in cereals	South Asia	NA	NA	HGDP-CEPH
	Sindhi	25	agriculturalist	specialized in cereals	South Asia	NA	NA	HGDP-CEPH
	Xibo	9	agriculturalist	specialized in cereals	South Asia	NA	NA	HGDP-CEPH
	Gujarati	88	agriculturalist	specialized in cereals	South Asia	NA	NA	HapMap Ph3
	Cambodian	11	agriculturalist	specialized in cereals	East Asia	NA	NA	HGDP-CEPH
	Dai	10	agriculturalist	specialized in cereals	East Asia	NA	NA	HGDP-CEPH
	Han	45	agriculturalist	specialized in cereals	East Asia	NA	NA	HGDP-CEPH
	Japanese	31	agriculturalist	specialized in cereals	East Asia	NA	NA	HGDP-CEPH
	Lahu	10	agriculturalist	specialized in cereals	East Asia	NA	NA	HGDP-CEPH
	Miaozu	10	agriculturalist	specialized in cereals	East Asia	NA	NA	HGDP-CEPH
	Naxi	10	agriculturalist	specialized in cereals	East Asia	NA	NA	HGDP-CEPH
	She	10	agriculturalist	specialized in cereals	East Asia	NA	NA	HGDP-CEPH
	Tujia	10	agriculturalist	specialized in cereals	East Asia	NA	NA	HGDP-CEPH
	Yizu	10	agriculturalist	specialized in cereals	East Asia	NA	NA	HGDP-CEPH
	Daur	10	agriculturalist	specialized in cereals	East Asia	NA	NA	HGDP-CEPH
	Afar (Ethiopia)	12	agriculturalist	specialized in cereals	Africa	NA	NA	Pagani, et al. (2012)
	Amhara1	NA	agriculturalist	not specialized in cereals	Africa	NA	NA	HapMap Ph3
	Amhara2	26	agriculturalist	specialized in cereals	Africa	NA	NA	Pagani, et al. (2012)
	Anuak (Ethiopia)	23	agriculturalist	specialized in cereals	Africa	NA	NA	Pagani, et al. (2012)
	Ari Blacksmith (Ethiopia)	17	agriculturalist	specialized in cereals	Africa	NA	NA	Pagani, et al. (2012)
	Ari cultivator (Ethiopia)	24	agriculturalist	specialized in cereals	Africa	NA	NA	Pagani, et al. (2012)
	Tigray (Ethiopia)	21	agriculturalist	specialized in cereals	Africa	NA	NA	Pagani, et al. (2012)
	Wolayta (Ethiopia)	8	agriculturalist	specialized in cereals	Africa	NA	NA	Pagani, et al. (2012)
	SEBantu (Herero Namibia)	12	agriculturalist	specialized in cereals	Africa	NA	NA	Schlebusch, et al. (2012)
	SWBantu (South Africa)	20	agriculturalist	specialized in cereals	Africa	NA	NA	Schlebusch, et al. (2012)
	Ugandan	110	pastoralist	not specialized in cereals	Africa	NA	NA	*this study*

Table S3 (continued)

	Masaai	143	pastoralist	not specialized in cereals	Africa	NA	NA	HapMap Ph3
	Bedouin	49	pastoralist	not specialized in cereals	South Asia	NA	NA	HGDP-CEPH
	Brahui	25	pastoralist	not specialized in cereals	South Asia	NA	NA	HGDP-CEPH
	Hazara	25	pastoralist	not specialized in cereals	South Asia	NA	NA	HGDP-CEPH
	Makrani	25	pastoralist	not specialized in cereals	South Asia	NA	NA	HGDP-CEPH
	Uygur	10	pastoralist	not specialized in cereals	South Asia	NA	NA	HGDP-CEPH
	Tu	10	pastoralist	not specialized in cereals	South Asia	NA	NA	HGDP-CEPH
	Mongolan	10	pastoralist	not specialized in cereals	East Asia	NA	NA	HGDP-CEPH
	Balochi	25	pastoralist	not specialized in cereals	East Asia	NA	NA	HGDP-CEPH
	Yakut	25	pastoralist	not specialized in cereals	East Asia	NA	NA	HGDP-CEPH
	Somali (Ethiopia)	17	pastoralist	not specialized in cereals	Africa	NA	NA	Pagani *et al* (2012)
	Gumuz (Ethiopia)	19	pastoralist	not specialized in cereals	Africa	NA	NA	Pagani *et al* (2012)
	Nama (Namibia)	20	pastoralist	not specialized in cereals	Africa	NA	NA	Schlebusch, et al. (2012)
	Uganda	110	pastoralist	not specialized in cereals	Africa	NA	NA	*this study*
	Masaai	143	pastoralist	not specialized in cereals	Africa	NA	NA	HapMap Ph3
	Oroqen	10	hunter-gatherer	not specialized in cereals	East Asia	NA	NA	HGDP-CEPH
	Hezhen	10	hunter-gatherer	not specialized in cereals	East Asia	NA	NA	HGDP-CEPH
	Maritime Chukchee	NA	hunter-gatherer	not specialized in cereals	East Asia	NA	NA	HGDP-CEPH
	Naukan Yup'ik	NA	hunter-gatherer	not specialized in cereals	East Asia	NA	NA	HGDP-CEPH
	Khoisan Namibia	62	hunter-gatherer	not specialized in cereals	Africa	NA	NA	*this study*
	San	7	hunter-gatherer	not specialized in cereals	Africa	NA	NA	HGDP-CEPH
	Baka Pygmies	39	hunter-gatherer	not specialized in cereals	Africa	NA	NA	*this study*
	Biaka Pygmy	36	hunter-gatherer	not specialized in cereals	Africa	NA	NA	HGDP-CEPH
	Mbuti Pygmy	15	hunter-gatherer	not specialized in cereals	Africa	NA	NA	HGDP-CEPH
	Khomani2 (South Africa)	35	hunter-gatherers	not specialized in cereals	Africa	NA	NA	Henn, et al. (2011)
	Hadza (Tanzania)	20	hunter-gatherers	not specialized in cereals	Africa	NA	NA	Henn, et al. (2011)
	Sandawe (Tanzania)	35	hunter-gatherers	not specialized in cereals	Africa	NA	NA	Henn, et al. (2011)
	Khwe (Angola)	17	hunter-gatherers	not specialized in cereals	Africa	NA	NA	Schlebusch, et al. (2012)

Table S3 (continued)

	!Xun (Angola)	19	hunter-gatherers	not specialized in cereals	Africa	NA	NA	Schlebusch, et al. (2012)
	Gui and Ghana (Botswana)	15	hunter-gatherers	not specialized in cereals	Africa	NA	NA	Schlebusch, et al. (2012)
	Ju/'hoansi (Namibia)	18	hunter-gatherers	not specialized in cereals	Africa	NA	NA	Schlebusch, et al. (2012)
	Karretjie (South Africa)	20	hunter-gatherers	not specialized in cereals	Africa	NA	NA	Schlebusch, et al. (2012)
	Khomani2 (South Africa)	39	hunter-gatherers	not specialized in cereals	Africa	NA	NA	Schlebusch, et al. (2012)
	Kung Vasekela	NA	hunter-gatherers	not specialized in cereals	Africa	NA	NA	HGDP-CEPH
c.1130A>G (MTRR)	Angolan	29	agriculturalist	folate poor-diet	Africa	NA	NA	*this study*
	Mozambican	30	agriculturalist	folate poor-diet	Africa	NA	NA	*this study*
	Equatorial Guinean	87	agriculturalist	folate poor-diet	Africa	NA	NA	*this study*
	N. Bantu	12	agriculturalist	folate poor-diet	Africa	NA	NA	HGDP-CEPH
	Mandenka	24	agriculturalist	folate poor-diet	Africa	NA	NA	HGDP-CEPH
	S. Bantu	8	agriculturalist	folate poor-diet	Africa	NA	NA	HGDP-CEPH
	Luhya	88	agriculturalist	folate poor-diet	Africa	NA	NA	HapMap Ph3
	Mozabite	30	agriculturalist	folate poor-diet	Africa	NA	NA	HGDP-CEPH
	Yoruba	25	agriculturalist	folate poor-diet	Africa	NA	NA	HGDP-CEPH
	Portuguese	47	agriculturalist	folate poor-diet	Europe	NA	NA	*this study*
	Basque	24	agriculturalist	folate poor-diet	Europe	NA	NA	HGDP-CEPH
	Italian (Bergamo)	14	agriculturalist	folate poor-diet	Europe	NA	NA	HGDP-CEPH
	French	29	agriculturalist	folate poor-diet	Europe	NA	NA	HGDP-CEPH
	Orcadian	16	agriculturalist	folate poor-diet	Europe	NA	NA	HGDP-CEPH
	Russian	25	agriculturalist	folate poor-diet	Europe	NA	NA	HGDP-CEPH
	Sardinian	28	agriculturalist	folate poor-diet	Europe	NA	NA	HGDP-CEPH
	Tuscan1	8	agriculturalist	folate poor-diet	Europe	NA	NA	HGDP-CEPH
	Tuscan2	89	agriculturalist	folate poor-diet	Europe	NA	NA	HapMap Ph3
	Adygei	17	agriculturalist	folate poor-diet	Middle East	NA	NA	HGDP-CEPH
	Druze	48	agriculturalist	folate poor-diet	Middle East	NA	NA	HGDP-CEPH
	Palestinian	51	agriculturalist	folate poor-diet	Middle East	NA	NA	HGDP-CEPH
	Burusho	25	agriculturalist	folate poor-diet	South Asia	NA	NA	HGDP-CEPH

Table S3 (continued)

	Kalash	25	agriculturalist	folate poor-diet	South Asia	NA	NA	HGDP-CEPH
	Pathan	25	agriculturalist	folate poor-diet	South Asia	NA	NA	HGDP-CEPH
	Sindhi	25	agriculturalist	folate poor-diet	South Asia	NA	NA	HGDP-CEPH
	Xibo	9	agriculturalist	folate poor-diet	South Asia	NA	NA	HGDP-CEPH
	Gujarati	88	agriculturalist	folate poor-diet	South Asia	NA	NA	HapMap Ph3
	Cambodian	11	agriculturalist	folate poor-diet	East Asia	NA	NA	HGDP-CEPH
	Dai	10	agriculturalist	folate poor-diet	East Asia	NA	NA	HGDP-CEPH
	Han	45	agriculturalist	folate poor-diet	East Asia	NA	NA	HGDP-CEPH
	Japanese	31	agriculturalist	folate poor-diet	East Asia	NA	NA	HGDP-CEPH
	Lahu	10	agriculturalist	folate poor-diet	East Asia	NA	NA	HGDP-CEPH
	Miaozu	10	agriculturalist	folate poor-diet	East Asia	NA	NA	HGDP-CEPH
	Naxi	10	agriculturalist	folate poor-diet	East Asia	NA	NA	HGDP-CEPH
	She	10	agriculturalist	folate poor-diet	East Asia	NA	NA	HGDP-CEPH
	Tujia	10	agriculturalist	folate poor-diet	East Asia	NA	NA	HGDP-CEPH
	Yizu	10	agriculturalist	folate poor-diet	East Asia	NA	NA	HGDP-CEPH
	Daur	10	agriculturalist	folate poor-diet	East Asia	NA	NA	HGDP-CEPH
	Afar (Ethiopia)	12	agriculturalist	folate poor-diet	Africa	NA	NA	Pagani, et al. (2012)
	Amhara1	NA	agriculturalist	folate poor-diet	Africa	NA	NA	HapMap Ph3
	Amhara2 (Ethiopia)	26	agriculturalist	folate poor-diet	Africa	NA	NA	Pagani, et al. (2012)
	Anuak (Ethiopia)	23	agriculturalist	folate poor-diet	Africa	NA	NA	Pagani, et al. (2012)
	Ari Blacksmith (Ethiopia)	17	agriculturalist	folate poor-diet	Africa	NA	NA	Pagani, et al. (2012)
	Ari cultivator (Ethiopia)	24	agriculturalist	folate poor-diet	Africa	NA	NA	Pagani, et al. (2012)
	Tigray (Ethiopia)	21	agriculturalist	folate poor-diet	Africa	NA	NA	Pagani, et al. (2012)
	Wolayta (Ethiopia)	8	agriculturalist	folate poor-diet	Africa	NA	NA	Pagani, et al. (2012)
	SEBantu (Herero Namibia)	12	agriculturalist	folate poor-diet	Africa	NA	NA	Schlebusch, et al. (2012)
	SWBantu (South Africa)	20	agriculturalist	folate poor-diet	Africa	NA	NA	Schlebusch, et al. (2012)
	Ugandan	110	pastoralist	folate poor-diet	Africa	NA	NA	*this study*
	Masaai	143	pastoralist	folate poor-diet	Africa	NA	NA	HapMap Ph3

Table S3 (continued)

	Bedouin	49	pastoralist	folate poor-diet	South Asia	NA	NA	HGDP-CEPH
	Brahui	25	pastoralist	folate poor-diet	South Asia	NA	NA	HGDP-CEPH
	Hazara	25	pastoralist	folate poor-diet	South Asia	NA	NA	HGDP-CEPH
	Makrani	25	pastoralist	folate poor-diet	South Asia	NA	NA	HGDP-CEPH
	Uygur	10	pastoralist	folate poor-diet	South Asia	NA	NA	HGDP-CEPH
	Tu	10	pastoralist	folate poor-diet	South Asia	NA	NA	HGDP-CEPH
	Mongolian	10	pastoralist	folate poor-diet	East Asia	NA	NA	HGDP-CEPH
	Balochi	25	pastoralist	folate poor-diet	East Asia	NA	NA	HGDP-CEPH
	Yakut	25	pastoralist	folate poor-diet	East Asia	NA	NA	HGDP-CEPH
	Somali (Ethiopia)	17	pastoralist	folate poor-diet	Africa	NA	NA	Pagani, et al. (2012)
	Gumuz (Ethiopia)	19	pastoralist	folate poor-diet	Africa	NA	NA	Pagani, et al. (2012)
	Nama (Namibia)	20	pastoralist	folate poor-diet	Africa	NA	NA	Schlebusch, et al. (2012)
	Ugandan	110	pastoralist	folate poor-diet	Africa	NA	NA	*this study*
	Masaai	143	pastoralist	folate poor-diet	Africa	NA	NA	HapMap Ph3
	Oroqen	10	hunter-gatherer	folate rich-diet	East Asia	NA	NA	HGDP-CEPH
	Hezhen	10	hunter-gatherer	folate rich-diet	East Asia	NA	NA	HGDP-CEPH
	Maritime Chukchee	NA	hunter-gatherer	folate rich-diet	East Asia	NA	NA	HGDP-CEPH
	Naukan Yup'ik	NA	hunter-gatherer	folate rich-diet	East Asia	NA	NA	HGDP-CEPH
	Khoisan Namibia	62	hunter-gatherer	folate rich-diet	Africa	NA	NA	*this study*
	San	7	hunter-gatherer	folate rich-diet	Africa	NA	NA	HGDP-CEPH
	Baka Pygmies	39	hunter-gatherer	folate rich-diet	Africa	NA	NA	*this study*
	Biaka Pygmy	36	hunter-gatherer	folate rich-diet	Africa	NA	NA	HGDP-CEPH
	Mbuti Pygmy	15	hunter-gatherer	folate rich-diet	Africa	NA	NA	HGDP-CEPH
	Khomani1 (South Africa)	35	hunter-gatherers	folate rich-diet	Africa	NA	NA	Henn, et al. (2011)
	Hadza (Tanzania)	20	hunter-gatherers	folate rich-diet	Africa	NA	NA	Henn, et al. (2011)
	Sandawe (Tanzania)	35	hunter-gatherers	folate rich-diet	Africa	NA	NA	Henn, et al. (2011)
	Khwe (Angola)	17	hunter-gatherers	folate rich-diet	Africa	NA	NA	Schlebusch, et al. (2012)
	!Xun (Angola)	19	hunter-gatherers	folate rich-diet	Africa	NA	NA	Schlebusch, et al. (2012)

Table S3 *(continued)*

	Gui and Ghana (Botswana)	15	hunter-gatherers	folate rich-diet	Africa	NA	NA	Schlebusch, et al. (2012)
	Ju/'hoansi (Namibia)	18	hunter-gatherers	folate rich-diet	Africa	NA	NA	Schlebusch, et al. (2012)
	Karretjie (South Africa)	20	hunter-gatherers	folate rich-diet	Africa	NA	NA	Schlebusch, et al. (2012)
	Khomani2 (South Africa)	39	hunter-gatherers	folate rich-diet	Africa	NA	NA	Schlebusch, et al. (2012)
	Kung Vasekela	NA	hunter-gatherers	folate rich-diet	Africa	NA	NA	HGDP-CEPH
c.590G>A (NAT2*6)	Angolan	29	agriculturalist	folate poor-diet	Africa	NA	NA	*this study*
	Mozambican	30	agriculturalist	folate poor-diet	Africa	NA	NA	*this study*
	Equatorial Guinean	87	agriculturalist	folate poor-diet	Africa	NA	NA	*this study*
	N. Bantu	12	agriculturalist	folate poor-diet	Africa	NA	NA	HGDP-CEPH
	Mandenka	24	agriculturalist	folate poor-diet	Africa	NA	NA	HGDP-CEPH
	S. Bantu	8	agriculturalist	folate poor-diet	Africa	NA	NA	HGDP-CEPH
	Luhya	88	agriculturalist	folate poor-diet	Africa	NA	NA	HapMap Ph3
	Mozabite	30	agriculturalist	folate poor-diet	Africa	NA	NA	HGDP-CEPH
	Yoruba	25	agriculturalist	folate poor-diet	Africa	NA	NA	HGDP-CEPH
	Portuguese	47	agriculturalist	folate poor-diet	Europe	NA	NA	*this study*
	Basque	24	agriculturalist	folate poor-diet	Europe	NA	NA	HGDP-CEPH
	Italian (Bergamo)	14	agriculturalist	folate poor-diet	Europe	NA	NA	HGDP-CEPH
	French	29	agriculturalist	folate poor-diet	Europe	NA	NA	HGDP-CEPH
	Orcadian	16	agriculturalist	folate poor-diet	Europe	NA	NA	HGDP-CEPH
	Russian	25	agriculturalist	folate poor-diet	Europe	NA	NA	HGDP-CEPH
	Sardinian	28	agriculturalist	folate poor-diet	Europe	NA	NA	HGDP-CEPH
	Tuscan1	8	agriculturalist	folate poor-diet	Europe	NA	NA	HGDP-CEPH
	Tuscan2	89	agriculturalist	folate poor-diet	Europe	NA	NA	HapMap Ph3
	Adygei	17	agriculturalist	folate poor-diet	Middle East	NA	NA	HGDP-CEPH
	Druze	48	agriculturalist	folate poor-diet	Middle East	NA	NA	HGDP-CEPH
	Palestinian	51	agriculturalist	folate poor-diet	Middle East	NA	NA	HGDP-CEPH
	Burusho	25	agriculturalist	folate poor-diet	South Asia	NA	NA	HGDP-CEPH

Table S3 (continued)

	Kalash	25	agriculturalist	folate poor-diet	South Asia	NA	NA	HGDP-CEPH
	Pathan	25	agriculturalist	folate poor-diet	South Asia	NA	NA	HGDP-CEPH
	Sindhi	25	agriculturalist	folate poor-diet	South Asia	NA	NA	HGDP-CEPH
	Xibo	9	agriculturalist	folate poor-diet	South Asia	NA	NA	HGDP-CEPH
	Gujarati	88	agriculturalist	folate poor-diet	South Asia	NA	NA	HapMap Ph3
	Cambodian	11	agriculturalist	folate poor-diet	East Asia	NA	NA	HGDP-CEPH
	Dai	10	agriculturalist	folate poor-diet	East Asia	NA	NA	HGDP-CEPH
	Han	45	agriculturalist	folate poor-diet	East Asia	NA	NA	HGDP-CEPH
	Japanese	31	agriculturalist	folate poor-diet	East Asia	NA	NA	HGDP-CEPH
	Lahu	10	agriculturalist	folate poor-diet	East Asia	NA	NA	HGDP-CEPH
	Miaozu	10	agriculturalist	folate poor-diet	East Asia	NA	NA	HGDP-CEPH
	Naxi	10	agriculturalist	folate poor-diet	East Asia	NA	NA	HGDP-CEPH
	She	10	agriculturalist	folate poor-diet	East Asia	NA	NA	HGDP-CEPH
	Tujia	10	agriculturalist	folate poor-diet	East Asia	NA	NA	HGDP-CEPH
	Yizu	10	agriculturalist	folate poor-diet	East Asia	NA	NA	HGDP-CEPH
	Daur	10	agriculturalist	folate poor-diet	East Asia	NA	NA	HGDP-CEPH
	Afar (Ethiopia)	12	agriculturalist	folate poor-diet	Africa	NA	NA	Pagani, et al. (2012)
	Amhara1	NA	agriculturalist	folate poor-diet	Africa	NA	NA	HapMap Ph3
	Amhara2 (Ethiopia)	26	agriculturalist	folate poor-diet	Africa	NA	NA	Pagani, et al. (2012)
	Anuak (Ethiopia)	23	agriculturalist	folate poor-diet	Africa	NA	NA	Pagani, et al. (2012)
	Ari Blacksmith (Ethiopia)	17	agriculturalist	folate poor-diet	Africa	NA	NA	Pagani, et al. (2012)
	Ari cultivator (Ethiopia)	24	agriculturalist	folate poor-diet	Africa	NA	NA	Pagani, et al. (2012)
	Tigray (Ethiopia)	21	agriculturalist	folate poor-diet	Africa	NA	NA	Pagani, et al. (2012)
	Wolayta (Ethiopia)	8	agriculturalist	folate poor-diet	Africa	NA	NA	Pagani, et al. (2012)
	SEBantu (Herero Namibia)	12	agriculturalist	folate poor-diet	Africa	NA	NA	Schlebusch, et al. (2012)
	SWBantu (South Africa)	20	agriculturalist	folate poor-diet	Africa	NA	NA	Schlebusch, et al. (2012)
	Ugandan	110	pastoralist	folate poor-diet	Africa	NA	NA	*this study*

Table S3 (continued)

	Masaai	143	pastoralist	folate poor-diet	Africa	NA	NA	HapMap Ph3
	Bedouin	49	pastoralist	folate poor-diet	South Asia	NA	NA	HGDP-CEPH
	Brahui	25	pastoralist	folate poor-diet	South Asia	NA	NA	HGDP-CEPH
	Hazara	25	pastoralist	folate poor-diet	South Asia	NA	NA	HGDP-CEPH
	Makrani	25	pastoralist	folate poor-diet	South Asia	NA	NA	HGDP-CEPH
	Uygur	10	pastoralist	folate poor-diet	South Asia	NA	NA	HGDP-CEPH
	Tu	10	pastoralist	folate poor-diet	South Asia	NA	NA	HGDP-CEPH
	Mongola	10	pastoralist	folate poor-diet	East Asia	NA	NA	HGDP-CEPH
	Balochi	25	pastoralist	folate poor-diet	East Asia	NA	NA	HGDP-CEPH
	Yakut	25	pastoralist	folate poor-diet	East Asia	NA	NA	HGDP-CEPH
	Somali (Ethiopia)	17	pastoralist	folate poor-diet	Africa	NA	NA	Pagani, et al. (2012)
	Gumuz (Ethiopia)	19	pastoralist	folate poor-diet	Africa	NA	NA	Pagani, et al. (2012)
	Nama (Namibia)	20	pastoralist	folate poor-diet	Africa	NA	NA	Schlebusch, et al. (2012)
	Uganda	110	pastoralist	folate poor-diet	Africa	NA	NA	*this study*
	Masaai	143	pastoralist	folate poor-diet	Africa	NA	NA	HapMap Ph3
	Oroqen	10	hunter-gatherer	folate rich-diet	East Asia	NA	NA	HGDP-CEPH
	Hezhen	10	hunter-gatherer	folate rich-diet	East Asia	NA	NA	HGDP-CEPH
	Maritime Chukchee	*NA*	hunter-gatherer	folate rich-diet	East Asia	NA	NA	HGDP-CEPH
	Naukan Yup'ik	*NA*	hunter-gatherer	folate rich-diet	East Asia	NA	NA	HGDP-CEPH
	Khoisan Namibia	62	hunter-gatherer	folate rich-diet	Africa	NA	NA	*this study*
	San	7	hunter-gatherer	folate rich-diet	Africa	NA	NA	HGDP-CEPH
	Baka Pygmies	39	hunter-gatherer	folate rich-diet	Africa	NA	NA	*this study*
	Biaka Pygmy	36	hunter-gatherer	folate rich-diet	Africa	NA	NA	HGDP-CEPH
	Mbuti Pygmy	15	hunter-gatherer	folate rich-diet	Africa	NA	NA	HGDP-CEPH
	Khomani1 (South Africa)	35	hunter-gatherers	folate rich-diet	Africa	NA	NA	Henn, et al. (2011)
	Hadza (Tanzania)	20	hunter-gatherers	folate rich-diet	Africa	NA	NA	Henn, et al. (2011)
	Sandawe (Tanzania)	35	hunter-gatherers	folate rich-diet	Africa	NA	NA	Henn, et al. (2011)

Table S3 (continued)

	Khwe (Angola)	17	hunter-gatherers	folate rich-diet	Africa	NA	NA	Schlebusch, et al. (2012)
	!Xun (Angola)	19	hunter-gatherers	folate rich-diet	Africa	NA	NA	Schlebusch, et al. (2012)
	Gui and Ghana (Botswana)	15	hunter-gatherers	folate rich-diet	Africa	NA	NA	Schlebusch, et al. (2012)
	Ju/'hoansi (Namibia)	18	hunter-gatherers	folate rich-diet	Africa	NA	NA	Schlebusch, et al. (2012)
	Karretjie (South Africa)	20	hunter-gatherers	folate rich-diet	Africa	NA	NA	Schlebusch, et al. (2012)
	Khomani2 (South Africa)	39	hunter-gatherers	folate rich-diet	Africa	NA	NA	Schlebusch, et al. (2012)
	Kung Vasekela	NA	hunter-gatherers	folate rich-diet	Africa	NA	NA	HGDP-CEPH

NA - no data available.

Table S4. Hardy-Weinberg Equilibrium test.

	c.32C>T (*AGXT*)			c.1074G>A (*PLRP2*)			c.1130A>G (*MTRR*)			c.191G>A (*NAT2*14*)			c.341T>C (*NAT2*5*)			c.590G>A (*NAT2*6*)			c.857G>A (*NAT2*7*)			c.219-237G>A (*CYP3A5*)		
POP	H_{OB}	H_{EX}	***P*-value**	H_{OB}	H_{EX}	***P*-value**	H_{OB}	H_{EX}	***P*-value**	H_{OB}	H_{EX}	***P*-value**	H_{OB}	H_{EX}	***P*-value**	H_{OB}	H_{EX}	***P*-value**	H_{OB}	H_{EX}	***P*-value**	H_{OB}	H_{EX}	***P*-value**
ANG (n=32)		*		0.5652	0.4493	0.3462	0.7059	0.5134	0.1616	0.0435	0.2638	0.0014	0.1364	0.3330	0.0171	0.1818	0.47355	0.0055		*		0.4000	0.3722	1.0000
EQG (n=82)	0.0964	0.0923	1.0000	0.4048	0.4388	0.6165	0.5287	0.4614	0.2423	0.1954	0.1773	1.0000	0.5529	0.4629	0.0972	0.3333	0.2951	0.4487	0.0465	0.0457	1.0000	0.2143	0.2464	0.3581
MOZ (n=30)	0.0741	0.0727	1.0000	0.3333	0.3638	0.6319	0.4333	0.5034	0.4818	0.2857	0.2509	1.0000	0.4000	0.3846	1.0000	0.4762	0.4181	0.6322		*		0.2333	0.2096	1.0000
UGN (n=116)	0.0909	0.1355	0.0099	0.4404	0.4799	0.4249	0.5534	0.4752	0.0994	0.1183	0.1307	0.3621	0.3902	0.4788	0.1077	0.4255	0.4290	1	0.0110	0.0110	1.0000	0.4112	0.3598	0.1783
BPY (n=39)	0.0294	0.0294	1.0000	0.3784	0.3110	0.3089	0.5128	0.4795	0.7428	0.0526	0.0519	1.0000	0.2105	0.3046	0.0838	0.4595	0.3869	0.3926		*		0.2895	0.2509	1.0000
KNA (n=62)		*		0.0161	0.0476	0.0241	0.2742	0.2385	0.5856		*		0.1312	0.1236	1.0000	0.0968	0.0928	1	0.1613	0.1762	0.4448	0.3871	0.3341	0.2703
PTG (n=48)	0.2979	0.3130	0.6605	0.4681	0.5052	0.7704	0.2340	0.2409	1.0000		*		0.39130	0.5055	0.1465	0.4043	0.3260	0.1708	0.0638	0.1018	0.1051	0.1522	0.1785	0.3510

*H*OB – Heterozygozity observed; *H*EX – Heterozygozity expected; * – this locus is monomorphic. Significant differences, after Bonferroni's correction for multiple tests are highlighted in **bold**; Populations' abbreviations as referred in Material and Methods.

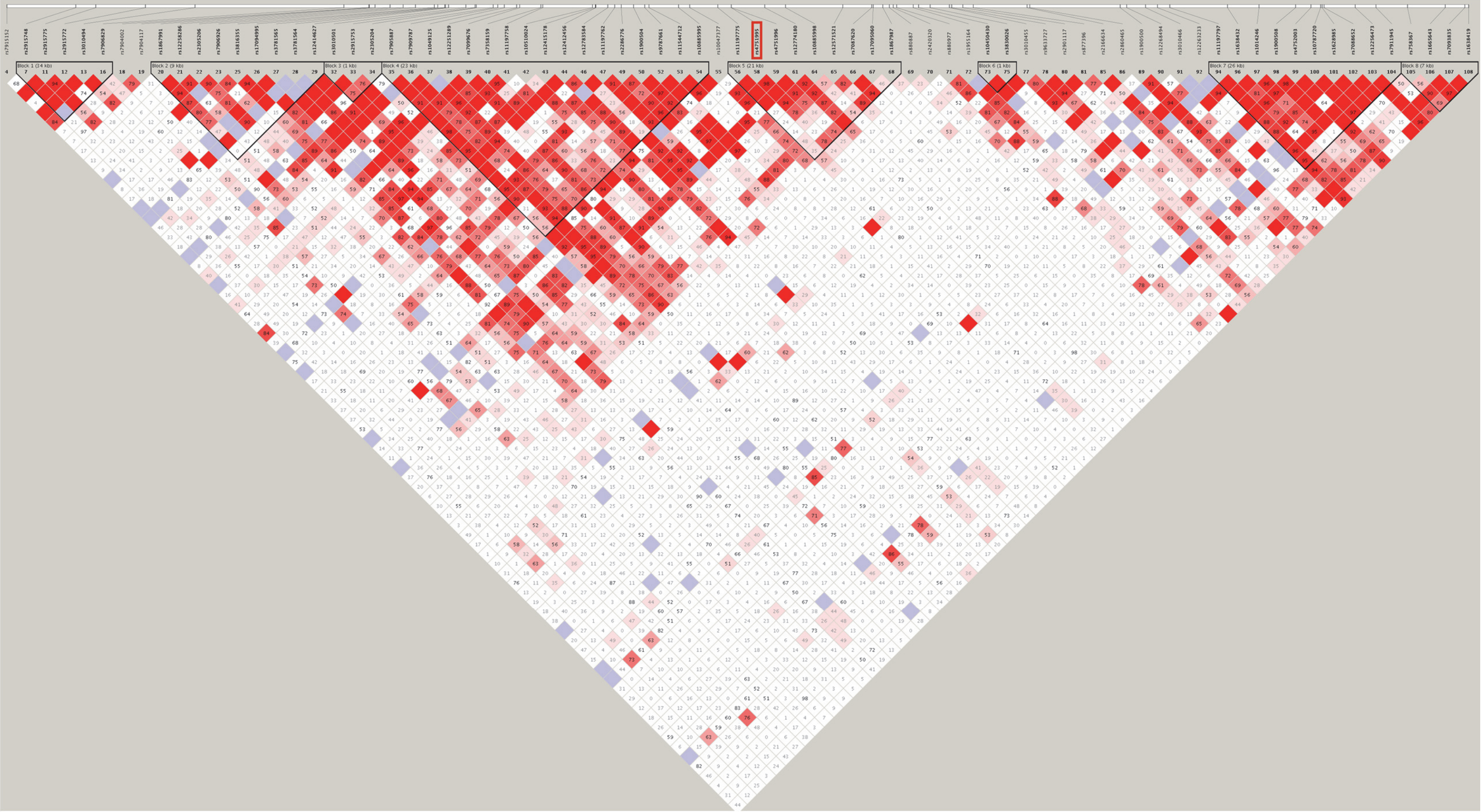


Figure S1. LD patterns for *PLRP2* (Pagani, et al. 2012).
The studied SNP is indicated by a red rectangle. In black triangles are represented the LD blocks. The degree of LD between pairs of markers is indicated by the |D'| statistic (|D'| = 1, red; |D'|<1, shades of red).

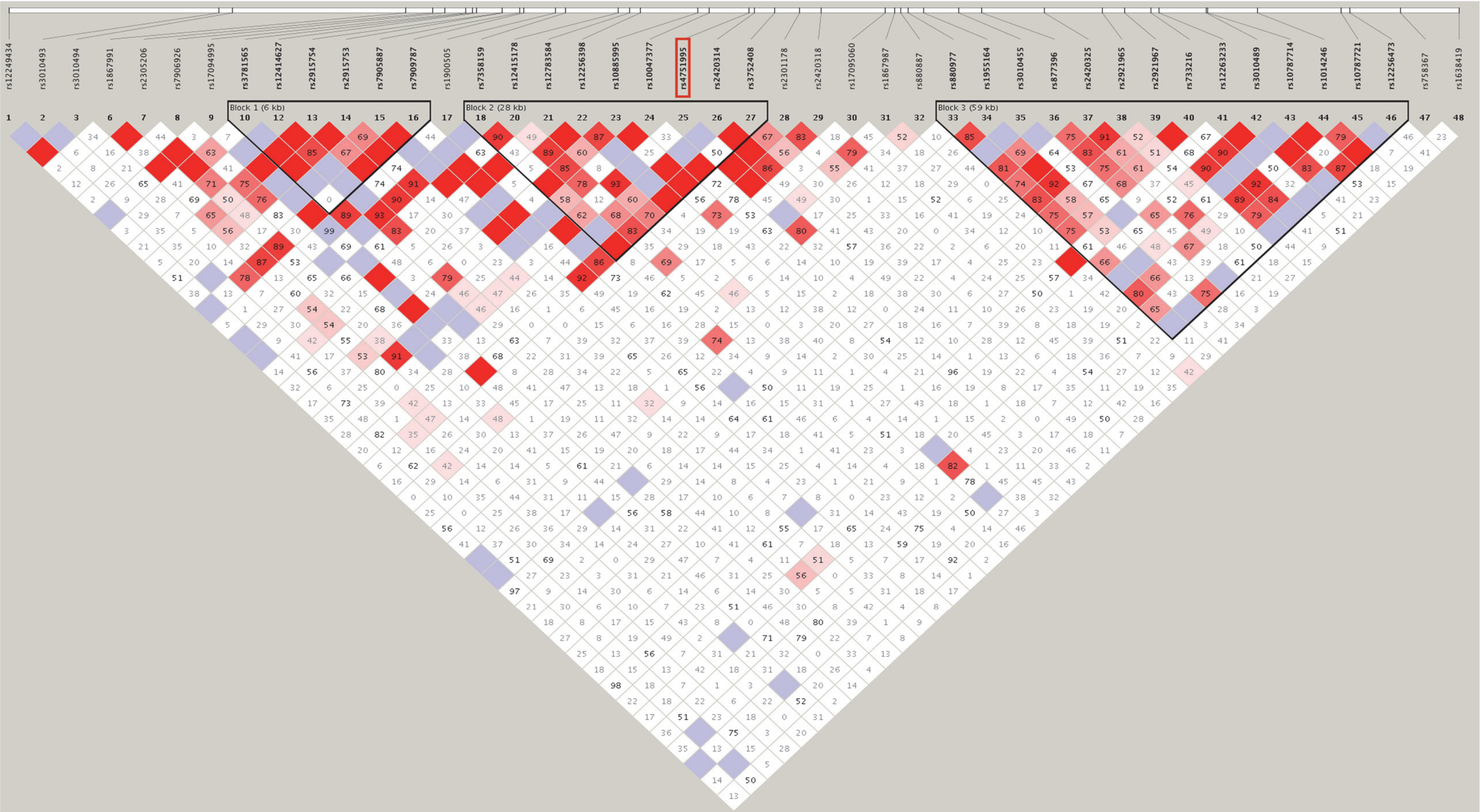


Figure S2. LD patterns for *PLRP2* (Henn, et al. 2011).
The studied SNP is indicated by a red rectangle. In black triangles are represented the LD blocks. The degree of LD between pairs of markers is indicated by the |D'| statistic (|D'| = 1, red; |D'|<1, shades of red).

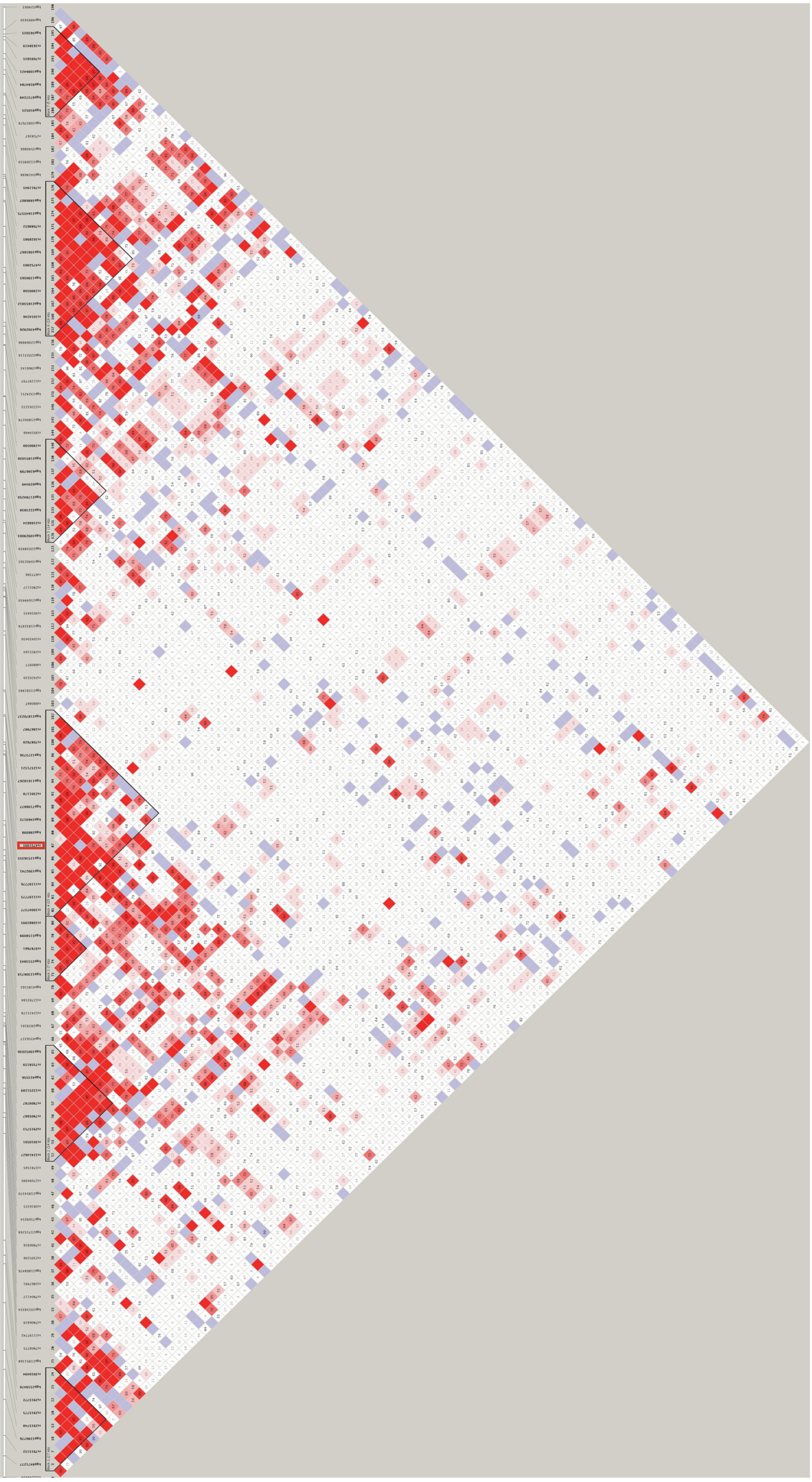


Figure S3. LD patterns for *PLRP2* (Schlebusch, et al. 2012).
The studied SNP is indicated by a red rectangle. In black triangles are represented the LD blocks. The degree of LD between pairs of markers is indicated by the |D'| statistic (|D'| = 1, red; |D'|<1, shades of red).

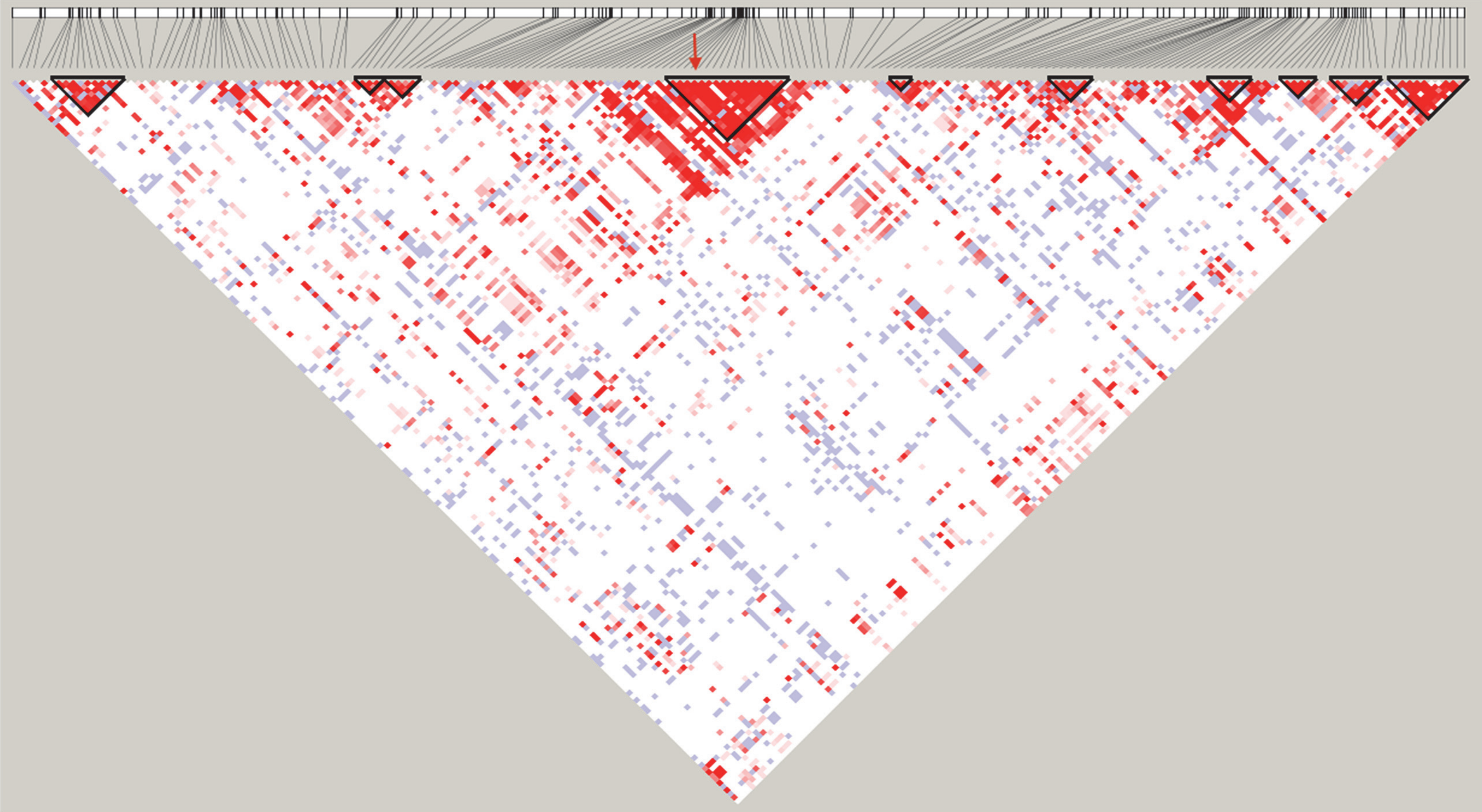


Figure S4. LD patterns for *NAT2* (Henn, et al. 2011).
A red arrow indicates the studied SNP. In black triangles are represented the LD blocks. The degree of LD between pairs of markers is indicated by the |D'| statistic (|D'| = 1, red; |D'|<1, shades of red).

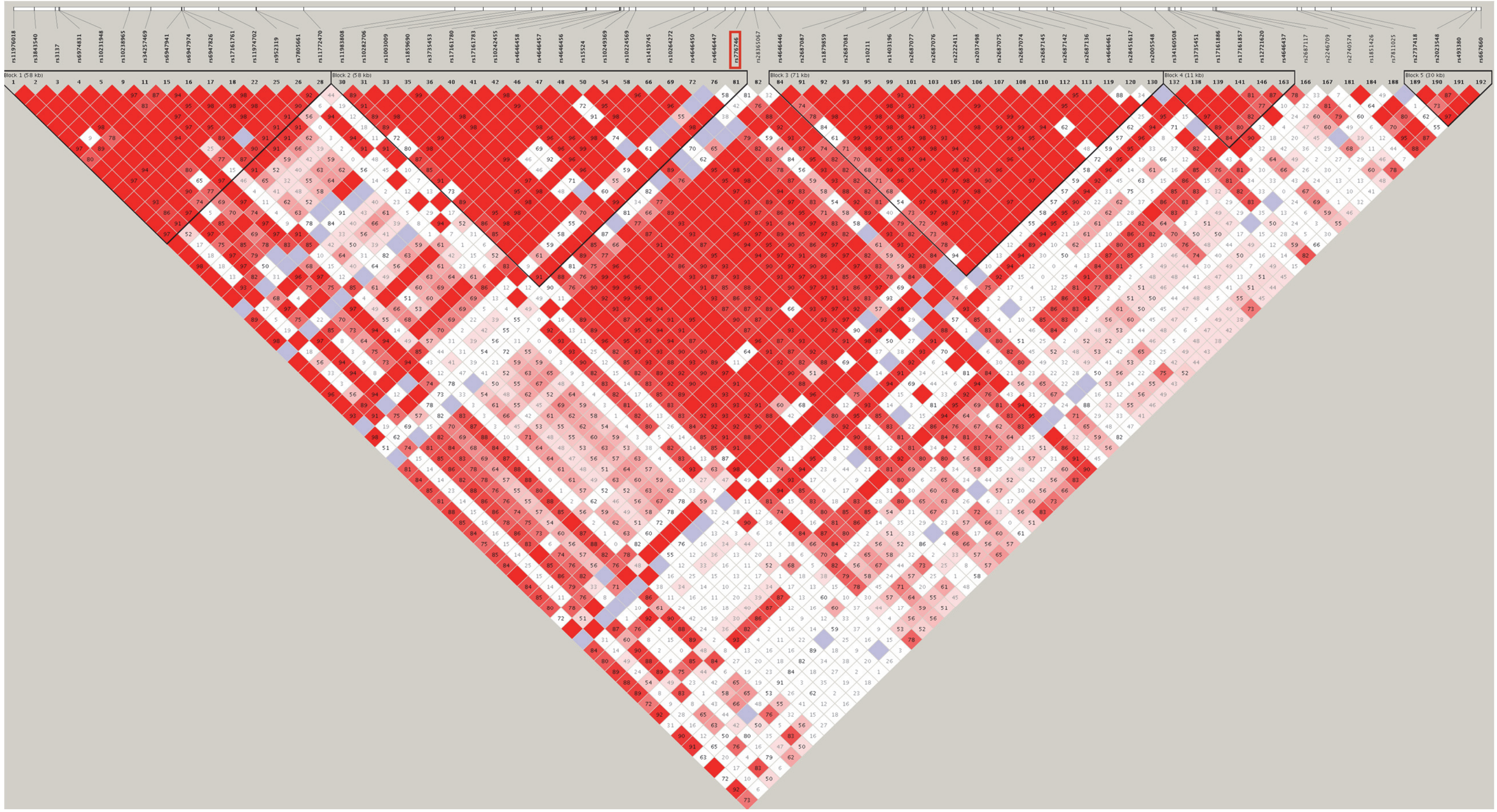


Figure S5. LD patterns for *CYP3A5* (Pagani, et al. 2012).
The studied SNP is indicated by a red rectangle. In black triangles are represented the LD blocks. The degree of LD between pairs of markers is indicated by the |D'| statistic (|D'| = 1, red; |D'|<1, shades of red).

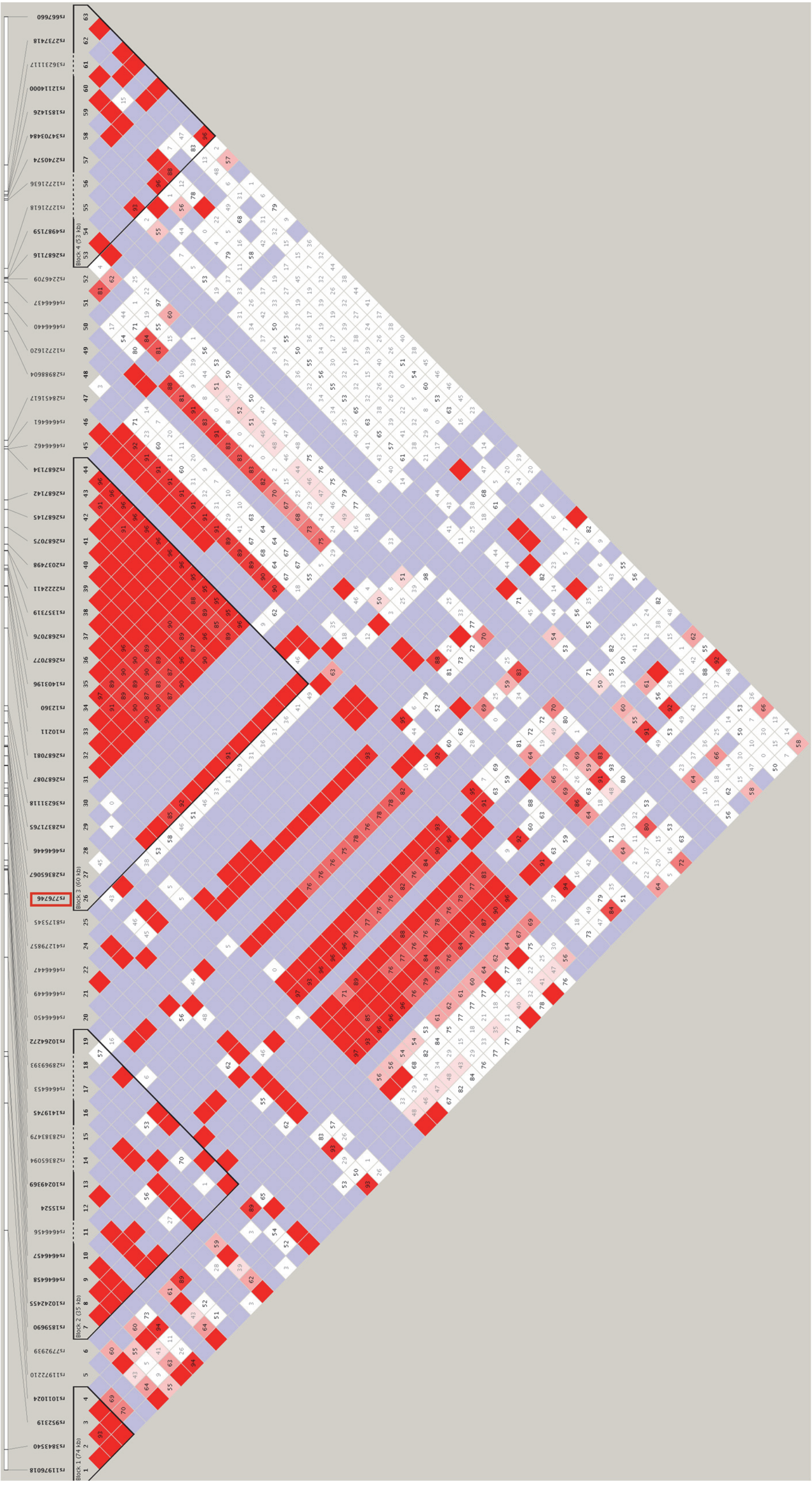


Figure S6. LD patterns for *CYP3A5* (Henn, et al. 2011).
The studied SNP is indicated by a red rectangle. In black triangles are represented the LD blocks. The degree of LD between pairs of markers is indicated by the |D'| statistic (|D'| = 1, red; |D'|<1, shades of red).

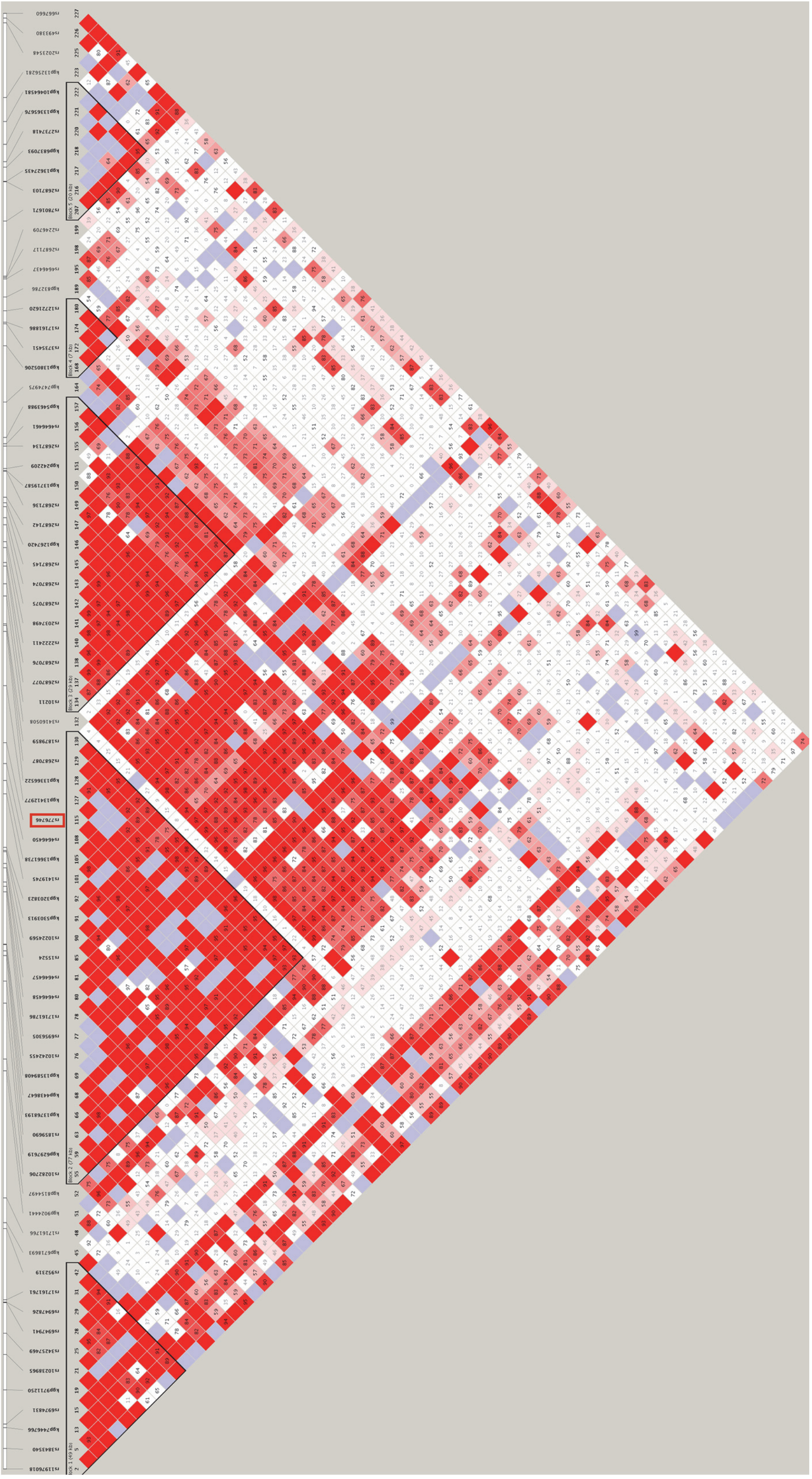


Figure S7. LD patterns for *CYP3A5* (Schlebusch, et al. 2012).
The studied SNP is indicated by a red rectangle. In black triangles are represented the LD blocks. The degree of LD between pairs of markers is indicated by the |D'| statistic (|D'| = 1, red; |D'|<1, shades of red).

3.2 "Assessing the relationship between variations in taste receptors genes from the TAS1R and TAS2R families and lifestyle of human populations”

3.2.1 Introduction

The human capacity to perceive flavors begins *in utero* with the development of the gustatory and olfactory systems, when also starts the learning about flavors in foods (Ventura and Worobey 2013). This early experience represents the first stage for the development of food preferences, which continues across the lifetime under a complex interplay between biological and environmental factors (Ventura and Worobey 2013).

Among the biological determinants of food preferences it is included the taste perception elicited by the stimulation of the gustatory receptors by food constituents, perception that varies widely among individuals in part due to the influence of genetic factors.

Taste receptors are proteins primarily found in taste receptor cells on the tongue and the upper digestive system that recognize ligands belonging to one of the five basic taste modalities - salty, sweet, bitter, sour and *umami*. Sweet, bitter and *umami* receptors are seven transmembrane G-protein-coupled receptors belonging to the so-called taste-1 (TAS1R) and taste-2 (TAS2R) families. Members of TAS1R might combine in different ways to form dimers that interact with sweet or *umami* tastings, while TAS2R receptors are structurally monomers that can detect natural and artificial compounds causing bitter responses. One of the best-characterized bitter receptors is TAS2R38, which is sensitive to phenylthiocarbamide (PTC), 6-n-propylthiouracil (PROP) and related compounds that contain a thiourea (N-C=S) moiety (Bufe, et al. 2005); besides it also affects perception of bitterness of glucosinolate-containing plants, such as broccoli, turnip, and horseradish (reviewed in Bachmanov and Beauchamp (2007)). It is now clearly demonstrated that a remarkable fraction of the phenotypic differences in PTC/PROP sensitivity is explained by genetic variability at the encoding gene, *TAS2R38* (Wooding, et al. 2004).

Another well-studied human bitter receptor is TAS2R16, which mediates response to β-D-glucopyranosides like salicin, for instance. It was also shown that a polymorphic variation at the encoding gene confers a strong differential response to salicin and other bitter ligands, including harmful cyanogenic glycosides (Soranzo, et al. 2005), suggesting that such variation influence preference for foods or beverages containing these naturally occurring bitter compounds and may have been advantageous for avoidance of toxic compounds in human diets (Imai, et al. 2012; Campbell, et al. 2013).

TAS1R1 and TAS1R3 combine in a heteromer that detects all 20 amino acids found in nature, sensing a taste described as *umami* (Nelson, et al. 2002), a Japanese word that means “good taste”. In humans the response is preferential to L-glutamate, the most common amino acid, a cleavage product of all proteins, which is present in food such as meat, fish, cheese, mushrooms, potatoes, among others (Garcia-Bailo, et al. 2009)

There is substantial interindividual variability in taste sensitivity to *umami*, for which it was already proven to account genetic variations in the encoding regions of both genes *TAS1R1* and *TAS1R3* (Chen, et al. 2009; Raliou, et al. 2009; Shigemura, et al. 2009; Bae, et al. 2010; Rawal, et al. 2013).

TAS1R3 also dimerizes with TAS1R2 forming a receptor that detects natural and artificial sweeteners. Although recognized since long that the perception of sweetness varies among humans, it was recently shown that only a moderate amount of the variation in the perceived intensity of high-potency sweeteners is due to genetic factors that appear to involve a single set of genes (Hwang, et al. 2015). Up to now, the understanding of the genetic basis of variability in human sweet perception still remains largely undetermined (Bachmanov, et al. 2014), with only two genetic variants located in the promoter of TAS1R3 having been identified as strongly correlated with human taste sensitivity to sucrose, though explaining only 16% of population variability in the perception (Fushan, et al. 2009).

It is widely assumed that genetic differences in taste perception substantially contribute to interindividual variability in food preferences and dietary habits. Individual differences in taste perception may influence dietary habits, affecting nutritional status and nutrition-related chronic disease risk (Garcia-Bailo, et al. 2009).

Over the past ~6 million years of hominin evolution, a number of major dietary transitions have occurred, including substantial increases in the consumption of meat and starch, the cooking of food, and the domestication of plants and animals (Luca, et al. 2010; Perry, et al. 2015). These transitions, in turn, played critical roles in many aspects of hominin evolution (Perry, et al. 2015), being often presumed that dietary shifts have triggered important genetic adaptations to cope with the new diets.

The agricultural revolution brought along many profound changes in human demography culture and technology. It was a transition lasting several thousand years that initially took place in the Middle East around 10-12,000 years, followed by the independent development in Northern and Southern China, Africa’ Sahel, Americas, New Guinea and

Sub-Saharan Africa, mainly tropical West Africa, and Ethiopia (Diamond 2002). It was in the Neolithic that nomadic hunters-gatherers began to turn into sedentary agrarian communities while other become herders, representing a shift in lifestyle that implied dramatic changes in the food consumed. Among other dietary changes, agriculturist populations increased highly the intake of cereals, and herder populations adopted a meat-rich and high diary input diet. Mounting evidence exist that diversity at number of genes involved in the processing of food was shaped by the strong selection pressures arising at that time, being believed that the resulting nutritional and metabolic adaptations were very important (Balaresque, et al. 2007; Heyer and Quintana-Murci 2009).

In this study, we planned to investigate whether there was any association between mode of subsistence in which populations rely and genetic variations in TAS1R1, TAS1R3, TAS2R16 and TAS2R38 known to have phenotypic effects in the perception of bitter, sweet and *umami* tastes. Viewing that, we first screened the target variations in 5 African populations practicing distinct subsistence strategies, to enrich the information available for populations from Africa, since the continent still remained largely understudied for the polymorphisms of interest. We also recruited previously published data from other African as well as European and Asian populations to obtain a broader panorama of the distribution of the referred variations.

Taking into account modes of subsistence of populations, we have assumed that the corresponding diets varied in the amounts of food derived from foraging, farming, or livestock. Then, that variable was tested against the pattern of frequency distribution of the genetic variants examined.

Given previous reports on clear signatures of selection at *TAS1R1*, *TAS1R*3, *TAS2R16* and *TAS2R38* genes (Kim, et al. 2006; Campbell, et al. 2012; Campbell, et al. 2013), the aim of this study was to evaluate if selective pressures were exerted over variations modeling taste perception prompting adaptations specific of population' lifestyle that could thus represent likely dietary adaptions.

3.2.2 Material and Methods

Samples

A total of 226 individuals from five sub-Saharan African populations were characterized: 37 from Angola (ANG), 31 from Mozambique (MOZ); 75 from Equatorial Guinea (EQG); 47 Karimojong from Uganda (UGN); and 36 Baka Pygmies from Gabon (BPY). ANG, MOZ and EQG are 3 agriculturalist Bantu speaker groups; UGN is a group of Nilotic herders and BPY are hunter-gatherers pygmies. A sample of 24 Portuguese (PTG) was also screened to use as reference for a non-African population. Samples management and DNA isolation were performed as previously referred in section 3.1 (Valente, et al. 2015).

SNPs and Genotyping

A battery of 10 SNPs located across four *TASR* genes were genotyped. In the *TAS1R* gene family, the selected SNPs were: i) C329T (rs41278020) and G1114A (rs34160967) in *TAS1R1* involved in *umami* perception (Kim, et al. 2006; Raliou, et al. 2009); ii) G13A (rs76755863) and C2269T (rs307377), located in exon 1 and 6 of *TAS1R3*, respectively, associated with *umami* perception (Kim, et al. 2006; Raliou, et al. 2009); and iii) -T1572C (rs307355) and -T1266C (rs35744813) located in the promoter region of *TAS1R3,* both correlated with sweet perception (Bachmanov, et al. 2002; Fushan, et al. 2009).

Concerning the *TAS2R* family, the selected SNPs influencing bitter perception were: G516T (rs846664) in *TAS2R16* (Bufe, et al. 2005; Soranzo, et al. 2005; Hinrichs, et al. 2006; Rajeevan, et al. 2012; Campbell, et al. 2013), and C145G (rs713598), C785T (rs1726866) and G886A (rs10246939) in *TAS2R38*, (Wooding, et al. 2004; Bufe, et al. 2005; Campbell, et al. 2012).

Primers were designed for the co-amplification of two set of SNPs using Primer3 software ver. 4.0 and possible secondary structures or interactions between primers were checked with AutoDimer software ver. 1.0 (http://www.cstl.nist.gov/biotech/strbase/AutoDimerHomepage/AutoDimerProgramHomepage.htm). The genotyping methodology was based in a multiplex minisequencing reaction, using the Single Base Extension (SBE) reaction kit (Applied Biosystems). SBE primers were designed and tested for secondary structures with the software described above.

Poly tails of varied lengths were attached to the 5' end of each primer in order to avoid identical fragment sizes. SBE products were run on ABI 3130 Genetic Analyser (AB Applied Biosystems) and the electropherograms were analyzed using GeneMapper software ver. 4.0. (Applied Biosystems, Foster City, USA), based on fragment size measured on GeneScan-120 size standard. The sequencing data were examined in the Sequence Scanner^{TM} v1.0 (AB Applied Biosystems). Information about the primers used for amplification and extension are summarized in Table S5 and S6, respectively, in supplementary material, from this chapter.

Statistical analyses

The ARLEQUIN software ver. 3.5.1 (Excoffier and Lischer 2010) was used to estimate allele and haplotype frequencies; to test for Hardy-Weinberg Equilibrium (HWE); to calculate genetic distances (F_{ST}s) and to perform AMOVA. Statistical significance of the *P* values for HWE tests and F_{ST}s values was evaluated after applying the Bonferroni's correction for multiple tests.

Multidimensional Scaling Analysis (MDS) of F_{ST}s values was performed using the SPSS software (SPSS 2013).

The contour map of frequencies of the T516 allele at *TAS2R16,* according to geography, was constructed with Surfer v8.05 (Golden Software).

Population data

Besides the data produced in this work for 5 African populations, ANG, MOZ, EQG, UGN and BPY, and that from Portugal, PTG, the following population' samples were used (Table S7):

i) for *TAS1R1* and *TAS1R3* coding region, we retrieved data from Kim, et al. (2006) for CAM (Cameroonian), NEU (North European), HUN (Hungarian), RUS (Russian), PAK (Pakistani), CHI (Chinese) and JPT (Japanese);

ii) for *TAS1R3* promoter, we obtained from Fushan, et al. (2009) data for AFR (African), EUR (European) and ASI (Asian);

iii) for *TAS2R16* we considered the populations a) studied by Soranzo, et al. (2005) - TUN (Tunisian), BST (Bantu from South Africa), AMH (Amhara from Ethiopia) and CAM

(Cameroonian); b) from the Hapmap database (International HapMap, et al. 2007) - YOR (Yoruba in Ibadan, Nigeria), LUH (Luhya in Webuye, Kenya), MAS1 (Maasai in Kinyawa, Kenya), CEU (Utah Residents (CEPH) with Northern and Western European Ancestry), CHB (Han Chinese in Beijing, China), JPT1 (Japanese in Tokyo, Japan); c) analysed by Hinrichs, et al. (2006) - BKE (Bantu from Kenya), BSE (Bantu from Southeast Africa), BIP (Biaka Pigmies from West Africa), MAN (Mandenka from West Africa), MPY (Mbuti Pigmies from Central African Republic), BNE (Bantu from Northeast Africa), SAN (San from South Africa), MOB (Mozabite), PAL (Palestine), BED (Bedouin), BRA (Brahui), ADY (Adygei), BAL (Balochi), MAK (Makrani), SIN (Sindhi), BUR (Burusho), HAZ (Hazara), KAL (Kalash), PAT (Pathan), BSQ (Basque), FRA (French), BER (Bergamo), SAR (Sardinian), TUS (Tuscan), CAB (Cambodian), DAI (Dai), LAH (Lahu), NAX (Naxi), DAU (Daur), DRU (Druze), HAN (Han), HEZ (Hezhen), ORO (Orogen), TU (Tu), JPT2 (Japanese), MIA (Miaozu), UYG (Uygur), XIB (Xibo), MON (Mongolian), ORC (Orcadian), RUS (Russian), SHE (She), TUJ (Tujia), YIZ (Yizu), YAK (Yakut) and JAP2 (Japanese); and d) contained in the ALFRED database (Rajeevan, et al. 2012) - HAU (Hausa from Nigeria), IBO (Ibo from Nigeria), LIS (Lisongo from Central African Republic), CHA (Chagga from Tanzania), MAS2 (Maasai from Tanzania-Kenya border), SAD (Sandawe from Tanzania), ZAR (Zaramo from Tanzania), KUW (Kuwait) and TSA (Tsaatan from lake Baikal region, Siberia);

iv) for *TAS2R38*, the population data was from a) Wooding, et al. (2004) – EUR1 (European from Hungary and European resident in Utah), EUR2 (European resident in Texas), and AFR (African from Cameroon); b) the ALFRED database (Rajeevan, et al. 2012) - PCA (Pygmies from Central African Republic) and c) Campbell, et al. (2012) - AEA (Africoasiatic from Kenya), AWC (Afroasiatic from Cameroon), NWC (Nilo-Saharan from Cameroon), NGC (Niger-Kordofanian from Cameroon), FWC (Fulani from Cameroon), PWC (Baka Pygmies from Cameroon), NEA (Nilo-Saharan from Kenya), NGA (Niger-Kordofanian from Kenya), LEA (Luo from Kenya), HAD (Hadza from Tanzania), PAL (Palestinian), BRA (Brahui) and EAS (East Asian from Japan and China).

Most populations were classified in different lifestyles (Table S7 from supplementary material), either based on the original reports, or according to Murdock Ethnographic Atlas (http://lucy.ukc.ac.uk/cgi-bin/uncgi/Ethnoatlas/atlas.vopts) and Encyclopedia of World Cultures, Africa: An Encyclopedia for Students (John Haley 2002).

3.2.3 Results and Discussion

Allele and haplotype frequencies

The observed genotypic distributions in populations here registered for the different polymorphisms did not showed significant departures from the Hardy-Weinberg expectations after applying the Bonferroni´s correction for multiple tests (Tables S8 and S9 from supplementary material). The estimates of allele and haplotypes' frequencies are shown and discussed in the following sections.

TAS1R1

In *TAS1R1* we screened the two non-synonymous variations known to influence *umami* perception, specifically C329T, in which the derived T allele reportedly decreases glutamate sensitivity (free glutamate induces the *umami* flavor); and G1114A, where the derived A allele increases glutamate sensitivity (Raliou, et al. 2009; Shigemura, et al. 2009). Since the two polymorphisms are in strong LD, they can define 4 haplotypes, out of which 3 were detected in our samples: CG the ancestral haplotype, which was assumed to confer intermediate *umami* sensitivity, CA assumed to be a high sensibility haplotype, and TG presumed to lead to decreased umami perception. The frequencies of those haplotypes in the populations screened are presented in Table 5.

The ancestral CG combination was by far the most frequent in all populations, reaching frequencies up to 90% in the African ones. The high sensitivity CA haplotype was present but at relatively low or moderate frequency in Africans, among whom it attained the peak value of ~9% in the Baka Pigmies from Gabon. In the Portuguese the haplotype was detected at ~10%, which was the highest value found in this work. It was also in the Portuguese that the low sensitivity haplotype, TG, was most well represented at ~5%. In Africa, it was only rarely found in the samples from EQG (~1.9%) and UGN (~2.3%), being absent in the remaining three population groups.

Overall, the distributions here observed fitted well the scarce data previously published on *TAS1R1* haplotypes that, to the best of our knowledge, were only provided by Kim, et al. (2006), as also fitted the haplotype distributions inspected in 1000 Genomes database (http://www.1000genomes.org/), that is going to be presented in the next section.

Remarkably, in the African sequences extracted from the 1000 Genomes database that we are going to show the TG haplotypes were absent, whereas in this study it was detected in the samples from Equatorial Guinea and Gabon, and in the study of Kim, et al. (2006) it was also found at 5% in the a sample from Cameroon. This means, and without devaluating the wealth of genome data provided by the 1000 Genomes Project, that the information contained in the public database is far from encompassing the extent of genetic diversity that can be found in African populations.

In relation to levels of haplotype diversity, in general they were reduced in all populations, especially from Africa and particularly in the agriculturalists from Angola and Mozambique where diversity only reached ~3%. Contrary to the most common pattern observed across major worldwide human populations, haplotype diversity in *TAS1R1* was more elevated outside then inside Africa, since it attained the value of 0.263 in the Portuguese, almost twice the highest value found in Africa, which was in the Baka Pygmies (0.155).

Table 5. Frequencies of haplotypes defined by variants C329T and G114A in *TAS1R1* and haplotype diversities in different populations.

Population	CG	CA	TG	TA	Haplotype diversity
EQG (n=75)	0.92105	0.05921	0.01974	-	0.1276 ± 0.0512
BPY (n=36)	0.91177	0.08824	-	-	0.1549 ± 0.0544
UGN (n=46)	0.94318	0.03409	0.02273	-	0.1054 ± 0.0434
ANG (n=32)	0.98214	0.01786	-	-	0.0312 ± 0.0300
MOZ (n=31)	0.98214	0.01786	-	-	0.0323 ± 0.0310
PTG (n=24)	0.85000	0.10000	0.05000	-	0.2633 ± 0.0785

Populations' abbreviations are referred in Material and Methods section.

Pairwise F_{ST} distances between the Portuguese plus the five African populations here studied were very low and statistically non-significant, meaning thus that no minor signals emerged that diversity at *TAS1R1* could be related with lifestyle of populations.

Comparisons were also performed with the populations studied by Kim, et al. (2006), showing up the reduced differentiation of *TAS1R1* at a broader geographical scale (Table 6). Exceptionally, the sample from Cameroon presented significant distances with all populations considering the conventional *P*=0.05, and with ANG, MOZ and UGN using the significance level corrected for multiple tests. Kim, et al. (2006) did not provide details on this sample, which probably could explain why it deviates from the overall panorama for *TAS1R1,* where sharp dissimilarities across populations are absent.

Table 6. Pairwise diferences between different populations, based on haplotype frequencies at *TAS1R1* gene.

	EQG	**BPY**	**UGN**	**ANG**	**MOZ**	**PTG**	**CAM**	**NEU**	**CHI**	**HUN**	**JAP**	**PAK**	**RUS**
EQG	*												
BPY	-0.00888	*											
UGN	-0.00936	0.00290	*										
ANG	0.01065	0.03158	0.00181	*									
MOZ	0.00953	0.03003	0.00100	-0.01612	*								
PTG	0.00960	-0.00203	0.02703	0.07706	0.07457	*							
CAM	0.09685	0.06597	**0.13334**	**0.19491**	**0.19063**	0.00771	*						
NEU	0.00509	0.02427	-0.00470	-0.02317	-0.02301	0.05201	0.13437	*					
CHI	0.00509	0.02427	-0.00470	-0.02317	-0.02301	0.05201	0.13437	0.00000	*				
HUN	0.00509	0.02427	-0.00470	-0.02317	-0.02301	0.05201	0.13437	0.00000	0.00000	*			
JAP	0.00509	0.02427	-0.00470	-0.02317	-0.02301	0.05201	0.13437	0.00000	0.00000	0.00000	*		
PAK	-0.00273	0.01584	-0.01201	-0.03028	-0.03015	0.04136	0.11947	0.00000	0.00000	0.00000	0.00000	*	
RUS	0.00509	0.02427	-0.00470	-0.02317	-0.02301	0.05201	0.13437	0.00000	0.00000	0.00000	0.00000	0.00000	*

values in **bold** are statistically significant after correction for multiple tests; Populations' abbreviations are referred in Material and Methods.

This has become clear in the MDS plot of F_{ST}s values shown in Figure 14, where most populations tend to cluster together regardless of their geographical location or lifestyle. The sample of Cameroonians occupies an evident outsider position in the plot, illustrating the before mentioned atypical high level of differentiation for *TAS1R1* that the sample exhibited.

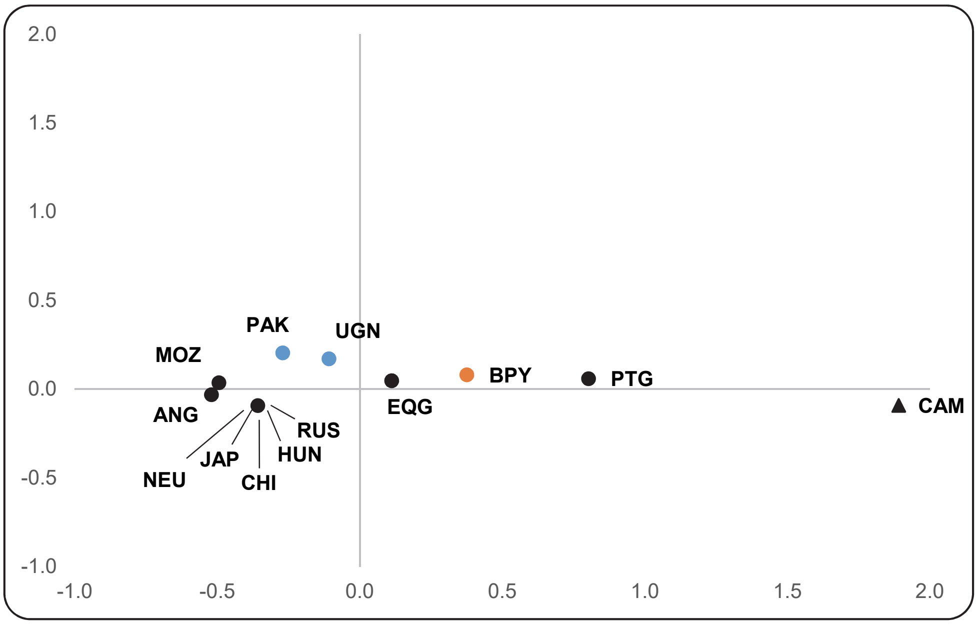


Figure 14. MDS plot based on F_{ST}s for *TAS1R1*.
In the MDS plot different colors represent distinct lifestyles: hunter-gatherers (orange), herders (blue) and farmers (black). Populations' abbreviations are referred in Material and Methods.

TAS1R3

In the genic region of *TAS1R3* the variations studied were G13A and C2269T, two non-synonymous substitutions that also influences sensibility to *umami* taste (Raliou, et al. 2009; Shigemura, et al. 2009). They define the haplotypes GC, the ancestral configuration associated with high *umami* sensibility, AC and GT, two intermediate haplotypes, and AT associated with low *umami* sensibility.

The frequencies depicted in Table 7 show that the high sensibility haplotype GC is very well represented in all populations, especially in the Baka Pygmies where it reaches ~94%. In this population group, it was only additionally found the intermediate AC haplotype, which was the second most frequent haplotype in all populations. So the low AT was the rarest one, solely detected in the Africans from Equatorial Guinea and Mozambique. GT, the intermediate sensibility haplotype, was also rarely detected in Africans, while it was moderately present in the Portuguese (~7%).

Table 7. Frequencies of haplotypes defined by variants G13A and C2269T in *TAS1R3* genic region and haplotype diversities in different populations.

Population	GC	GT	AC	AT	Haplotype diversity
EQG (n=75)	0.77333	-	0.20667	0.02000	0.3613 ± 0.0402
BPY (n=36)	0.94444	-	0.05556	-	0.1064 ± 0.0480
UGN (n=44)	0.86667	0.01111	0.12222	-	0.2411 ± 0.0549
ANG (n=36)	0.76667	0.01667	0.21667		0.3721 ± 0.0562
MOZ (n=28)	0.66071	-	0.32143	0.01786	0.4682 ± 0.0475
PTG (n=21)	0.76191	0.071423	0.16667	-	0.3961 ± 0.0829

Populations' abbreviations are referred in Material and Methods.

In the African populations, the level of haplotype diversity was lowest in the Baka Pigmies (~0.106), which it is understandable since they only showed two distinct haplotypes, followed by, Ugandans, Equatorial Guineans and Angolans, whereas it attained the highest value of 46.8% in the Mozambicans. The Portuguese was the second most diverse population out of the six here studied (39.6%).

Comparing through pairwise F_{ST}s the population groups addressed in this work, significant distances were detected between the Baka Pygmies and the two samples of agriculturalists from Equatorial Guinea and Mozambique, which can be mainly explained by the very high frequency of the haplotype that confers increased *umami* sensibility in the group of hunter-gatherers (see Table 8). When comparisons were extended to the populations studied by Kim, et al. (2006), the unique with data available for *TAS1R3*, once more genetic distances involving the Cameroonian population were often statistically significant. Besides, significant genetic differentiations were only observed in pairs involving the sample from Mozambique and different non-African populations.

Table 8. Pairwise differences between different populations, based on haplotype frequencies at *TAS1R3* gene.

	EQG	**BPY**	**UGN**	**ANG**	**MOZ**	**PTG**	**CAM**	**NEU**	**CHI**	**HUN**	**JAP**	**PAK**	**RUS**
EQG	*												
BPY	**0.07735**	*											
UGN	0.01498	0.01880	*										
ANG	-0.00909	0.10129	0.01934	*									
MOZ	0.02046	**0.21304**	**0.09687**	0.00940	*								
PTG	-0.00546	0.09144	0.00931	-0.01056	0.02387	*							
CAM	**0.11352**	**0.19523**	**0.12079**	**0.10569**	**0.14377**	0.04856	*						
NEU	0.10644	0.00480	0.05312	0.12852	**0.21899**	0.11393	0.17157	*					
CHI	0.10644	0.00480	0.05312	0.12852	**0.21899**	0.11393	0.17157	0.00000	*				
HUN	0.10644	0.00480	0.05312	0.12852	**0.21899**	0.11393	0.17157	0.00000	0.00000	*			
JAP	0.01221	-0.01627	-0.02691	0.01818	0.09129	0.00633	0.10117	0.05263	0.05263	0.05263	*		
PAK	0.09858	-0.00307	0.04451	0.11758	0.20455	0.10015	0.15540	0.00000	0.00000	0.00000	0.03679	*	
RUS	0.10644	0.00480	0.05312	0.12852	0.21899	0.11393	0.17157	0.00000	0.00000	0.00000	0.05263	0.00000	*

values in **bold** are statistically significant after correction for multiple tests; Populations' abbreviations are referred in Material and Methods.

The delivered pairwise F_{ST}s were used to construct the MDS presented in Figure 15. In this plot it is evident the outsider position of the Cameroonians, which is again difficult to explain unknowing more details about that sample. Excepting it, the remaining African populations do not form a clear cluster, though they tend to be discriminated from the non-Africans by the slightly negative values in the 2nd dimension of the plot. Most of them occupy the right lower quadrant and only the Baka Pygmies stand in the left lower quadrant of the plot.

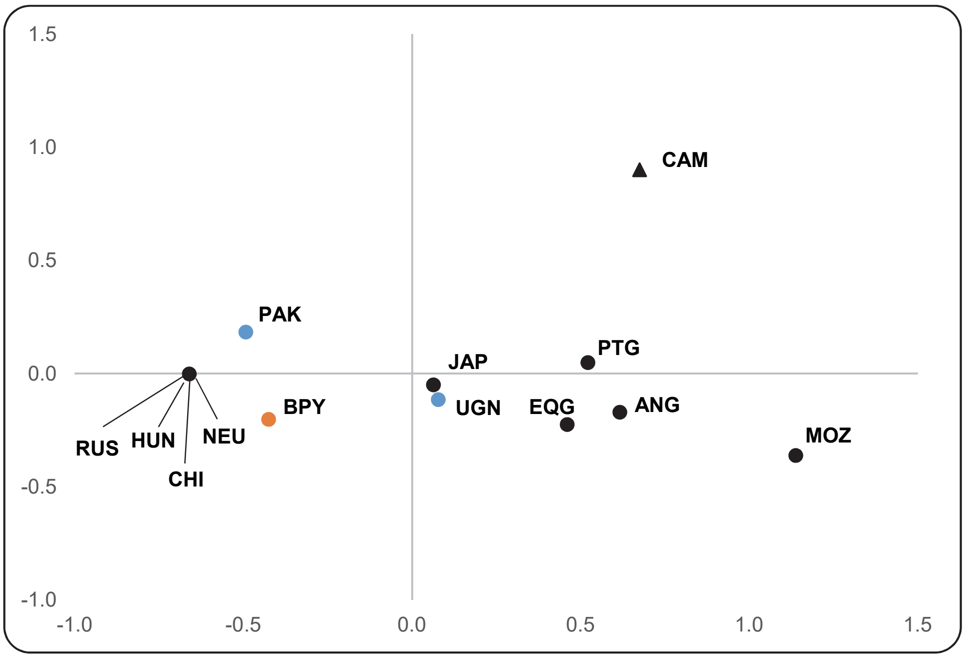


Figure 15. MDS plot based on F_{ST}s for *TAS1R3*.
In the MDS plot different colors represent distinct lifestyles: hunter-gatherers (orange), herders (blue) and farmers (black). Populations' abbreviations are referred in Material and Methods.

The relative position of the African populations could suggest that in respect to *TAS1R3* haplotypes accounting to *umami* perception, populations that do not produce food tend to be slightly differentiated from those that are food producers. Still, the analysis relies in a too much-reduced number of populations, which makes it impossible to draw any meaningful inference.

TAS1R3 promoter

In the promoter of *TAS1R3* lie the two variations -T1572C and -T1266C, that were here examined because they were strongly correlated with the ability of people to correctly sort ascending concentrations of sucrose (Fushan, et al. 2009). We have verified that the two SNPs were in significant LD, defining four haplotypes: the ancestral TT, associated with low sensibility to sweet, CT and TC, two “intermediate” haplotypes, and the double derived CC haplotype that confers augmented sweet sensibility. Estimates of their frequencies in the different populations are presented in Table 9.

Table 9. Frequencies of haplotypes defined by variants -T1572C and -T1266C in the promoter of *TAS1R3* and haplotype diversities in different populations.

Population	CC	CT	TT	TC	Haplotype diversity
EQG (n=75)	0.36577	0.17423	0.45244	0.00756	0.6342 ± 0.0175
BPY (n=36)	0.43702	0.34076	0.20090	0.02132	0.6647 ± 0.0256
UGN (n=47)	0.25438	0.18179	0.55225	0.01158	0.6024 ± 0.0355
ANG (n=37)	0.33670	0.11918	0.51317	0.03095	0.6150 ± 0.0348
MOZ (n=31)	0.30343	0.22882	0.44860	0.01915	0.6611 ± 0.0258
PTG (n=24)	0.85417	0.02083	0.12500	-	0.2525 ± 0.0531

Populations’ abbreviations are referred in Material and Methods.

It stands out the remarkable difference between the Portuguese and the African samples: whereas in Portugal the high sweet sensibility CC haplotype reaches ~85.4% and the low sensibility TT haplotype is presented at only 12.5%, in Africa the frequency of CC is usually more than half lower and TT is very well represented in most populations, where it is present at frequencies varying between 45.2 and 55.2%, except in the Baka Pygmies who carry it at 20.1%. Regarding haplotype diversity, the African populations presented at least two times higher values than the Portuguese, fitting so the more usual scenario concerning levels of diversity in African versus non-African populations.

Among the African groups, only the Baka Pygmies revealed to be highly and significantly differentiated from the agriculturalists from Angola and Equatorial Guinea, and from the pastoralists from Uganda (F_{ST} values in Table 10), differentiation for which likely accounted the discrepancy in the frequency of TT haplotype that was more than two times higher in Angolans, Guineans and Ugandans compared to the Pygmies’ group.

The trend that can be drawn from these data apparently indicates that haplotypes conferring high or intermediate sensibility to sweet are more common in hunter-gatherers than in agriculturalists or herders.

Comparisons were also performed with the few previously published data by Fushan, et al. (2009), whose study relied in general samples from Asia, Africa and Europe. Concerning the last two, the reported frequencies were in good conformity with those here obtained, but in addition the results of Fushan, et al. (2009) revealed that Asians are also remarkably differentiated from Africans. The F_{ST} distances between Asian and African groups were very high and still higher between European and African groups, with all those values (ranging from 18.9% to 51.5%) being statistically significant (Table 10).

Table 10. Pairwise differences between different populations based on haplotype frequencies at the promoter of *TAS1R3*.

	EQG	BPY	UGN	ANG	MOZ	PTG	AFR	EUR	ASI
EQG	*								
BPY	**0.06432**	*							
UGN	0.00930	**0.12018**	*						
ANG	-0.00410	**0.10105**	-0.00208	*					
MOZ	-0.00485	0.05389	0.00005	-0.00022	*				
PTG	**0.27413**	**0.25000**	**0.39990**	**0.34181**	**0.34681**	*			
AFR	-0.01054	0.00952	0.02469	0.00711	-0.01188	0.29949	*		
EUR	**0.37334**	**0.36284**	**0.51456**	**0.46483**	**0.47172**	0.00234	**0.44247**	*	
ASI	**0.18801**	**0.18827**	**0.29752**	**0.23216**	**0.24607**	0.01299	**0.18901**	0.05987	*

values in **bold** are statistically significant after correction for multiple tests; Populations' abbreviations are referred in Material and Methods.

The tight differentiation between African and non-African populations is patent in the MDS plot shown in Figure 16, which further shows some demarcation of the Baka Pygmies in the cluster of African populations.

The outstanding elevated prevalence of high sweet sensitivity haplotypes in Eurasian populations is suggestive that some kind of selective pressure was exerted favoring different genetic compositions in Africans and non-Africans. Fushan, et al. (2009) hypothesized that during human evolution such pressure could be related with the differential geographical distribution of sugar-rich aliments. Assuming that initially carbohydrate-rich plants, like sugar cane and sugar beets, grew mainly in tropical latitudes, the authors posited that the ability to taste sugars at lower concentrations might has been an important factor for human survival in cold geographical regions. Under this scenario, high sweet sensitivity haplotypes were favored in the non-tropical geographical regions, eventually driving their frequency to high values in populations.

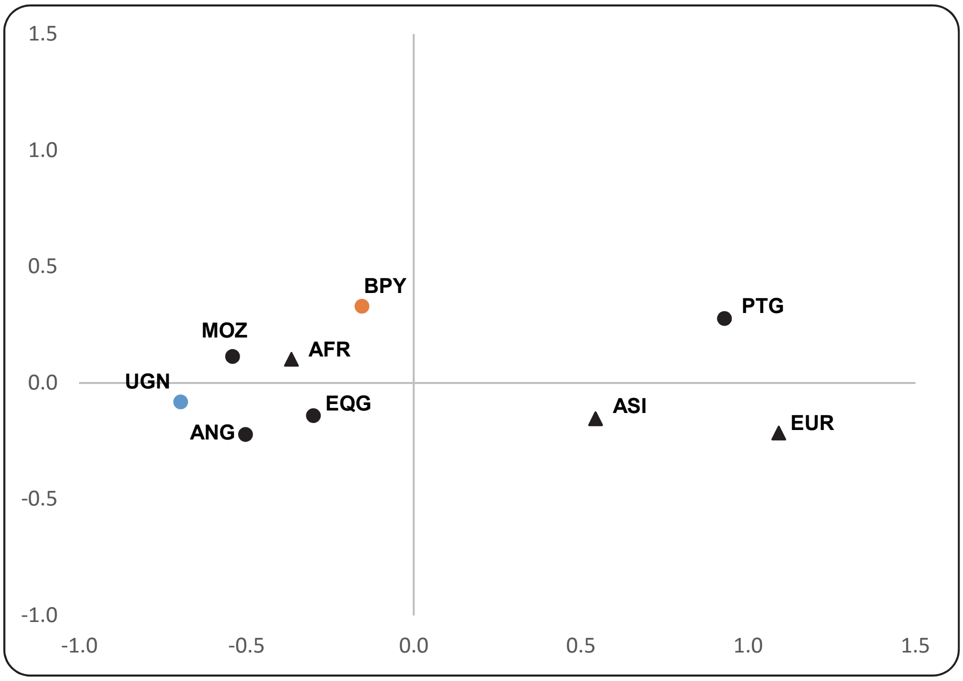


Figure 16. MDS plot based on F_{ST}s for the promoter region of *TAS1R3*.
In the MDS plot different colors represent distinct lifestyles: hunter-gatherers (orange), herders (blue) and farmers (black). Populations' abbreviations are referred in Material and Methods.

TAS2R16

Among the proteins belonging to the TAS2R family of taste receptors, which function in the recognition of bitter compounds, is included *TAS2R16*, a receptor mediating response to various β-D-glucopyranosides commonly found in nature, such as salicin. Within *TAS2R16*, we screened the variant G516T, concerning which previous transfection studies have demonstrated that the ancestral G and the derived T alleles were associated with lowered and increased sensitivity, respectively, to different cyanogenic glycosides (Soranzo et al, 2005).

Estimates of their frequencies in the six populations examined are presented in Table 11.

While in the African populations the two alleles were always present, the derived T allele was much more common than the ancestral form, peaking in frequency in the pastoralists from Uganda (86.2%). In the Portuguese the derived high sensibility allele was fixed.

These results are in full accordance with previous works (Soranzo, et al. 2005; Hinrichs, et al. 2006; International HapMap, et al. 2007; Rajeevan, et al. 2012) revealing that the ancestral G516 variant is absent or very rarely found outside of Africa, where it was only reported in populations from the Middle East, like Palestinians (6%) or Bedouins (2%), occurring also residually in a few ethnic groups from Asia such as the Brahui from Pakistan and the Tsaatan from the Baikal region (Hinrichs, et al. 2006; (Rajeevan, et al. 2012).

Table 11. Allele frequencies of the variant G516T located in *TAS2R16* gene.

Population	G	T	Gene diversity
EQG (n=75)	0.28667	0.71333	0.4117 ± 0.0317
BPY (n=36)	0.25714	0.74286	0.3940 ± 0.0497
UGN (n=44)	0.13830	0.86170	0.2382 ± 0.0533
ANG (n=36)	0.29412	0.70588	0.4190 ± 0.0454
MOZ (n=28)	0.29032	0.70968	0.4186 ± 0.0543
PTG (n=21)	0	1	0.0000 ± 0.0000

Populations' abbreviations are referred in Material and Methods.

In Africa, the ancestral low sensibility allele was found in all populations up to now studied, with the highest values having been reported in agriculturalist groups from Cameroon (33.0%) and from Nigeria, namely in the Ibo (42.4%) and the Yoruba (44.0%).

Noteworthy, in the Lisongo the allele was also very common (38%). The Lisongo belong to an ethnic group from Central African Republic that coexists with the Baka Pygmies from the forests of the region. However, despite being mainly foragers, the Lisongo are Bantu speakers who are believed to have resulted from recent admixture with other groups (Jin, et al. 2000; Knight, et al. 2003).

Pairwise genetic distances involving the six populations here screened are presented in a triangular matrix in Table 12. No significant distances were observed between the 5 African groups, while they significantly differed from the Portuguese, with exception of the herders from Uganda.

Table 12. Pairwise differences between the six populations addressed in this study based on allele frequencies for G516T in *TAS2R16*.

	EQG	BPY	UGN	ANG	MOZ	PTG
EQG	*					
BPY	-0.00908	*				
UGN	0.05298	0.03859	*			
ANG	-0.01032	-0.01214	0.05915	*		
MOZ	-0.01311	-0.01530	0.05712	-0.01681	*	
PTG	**0.19308**	**0.20633**	0.08863	**0.23160**	**0.25212**	*

values in **bold** are statistically significant after correction for multiple tests; Populations' abbreviations are referred in Material and Methods.

The pairwise F_{ST} values including the entire set of populations with data available for G516T are shown in Table S10 from supplementary material.

Differences between non-African and Africans populations were usually very large and most of them statistically significant.

Among Sub-Saharan Africans, some structure was also found, with the most significant differentiations having been observed between the Sandawe, a group of hunter-gatherers from Tanzania, and the Ibo (agriculturalists from Nigeria) or the Lisongo (admixed hunter-gatherers from Central African Republic).

In the MDS plot shown in Figure 17, which was constructed with pairwise F_{ST} distances, it was clearly captured in dimension 1 the strong differentiation between African and non-African populations.

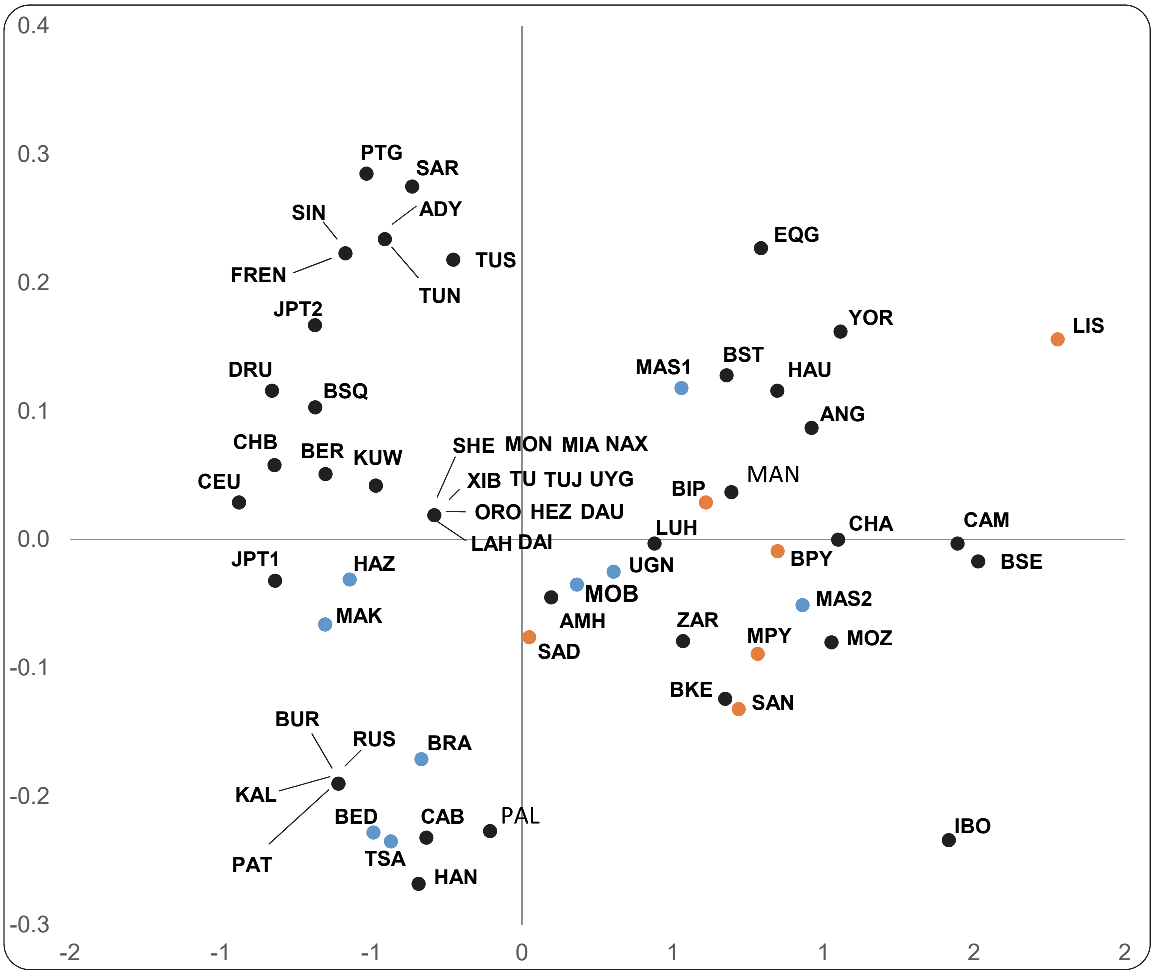


Figure 17. MDS plot based on F_{TS}s for *TAS2R16*.
In the MDS plot different colors represent distinct lifestyles: hunter-gatherers (orange), herders (blue) and farmers (black). Populations' abbreviations are referred in Material and Methods.

Within the cluster of Africans, which occupies the 2nd and 4th quadrants of the plot, populations are dispersed without showing a pattern of substructure sharply defined by the subsistence economy of populations. Still, pastoralists groups like UGN, MAS1 and MOB, as well as the hunter-gatherers SAD, BIP, BPY, MPY and SAN localize in the left side of those quadrants, reflecting the fact of having been among African populations relying in both lifestyles that low frequencies of the G516 allele associated with decreased sensitivity to bitter taste compounds were more commonly found than in the farmer groups.

The recognition of bitter natural toxins is supposed to have worked as a selective advantage alerting humans to noxious foods, especially in hunter-gatherers who depend much in plants that might be toxic (Li et al, 2011). Likewise, it is tempting to speculate that the variant increasing the bitter perception might have been advantageous not only in hunter-gatherer but also in herder populations, for reasons difficult to discern for the moment. But seemingly in Africa, and excluding the Lisongo, it was mainly among agriculturalist groups that the low bitter sensitivity allele has succeeded to increase in frequency, a finding that might reflect a relaxation of selective pressures somehow related to the farmers' economy.

Remarkably, the distribution of T516 allele in Sub-Saharan farmers brings to mind the Bantu expansion (Figure 18). As before mentioned, it was in agriculturalist groups from Nigeria and Cameroon that were found the highest frequencies of the G516 allele. Linguistic and archaeological inferences suggest that the region of modern Cameroon and Eastern Nigeria corresponds to the core area from which began, approximately 4 Kya ago, the expansion of the Bantu languages, which was largely a consequence of the movement

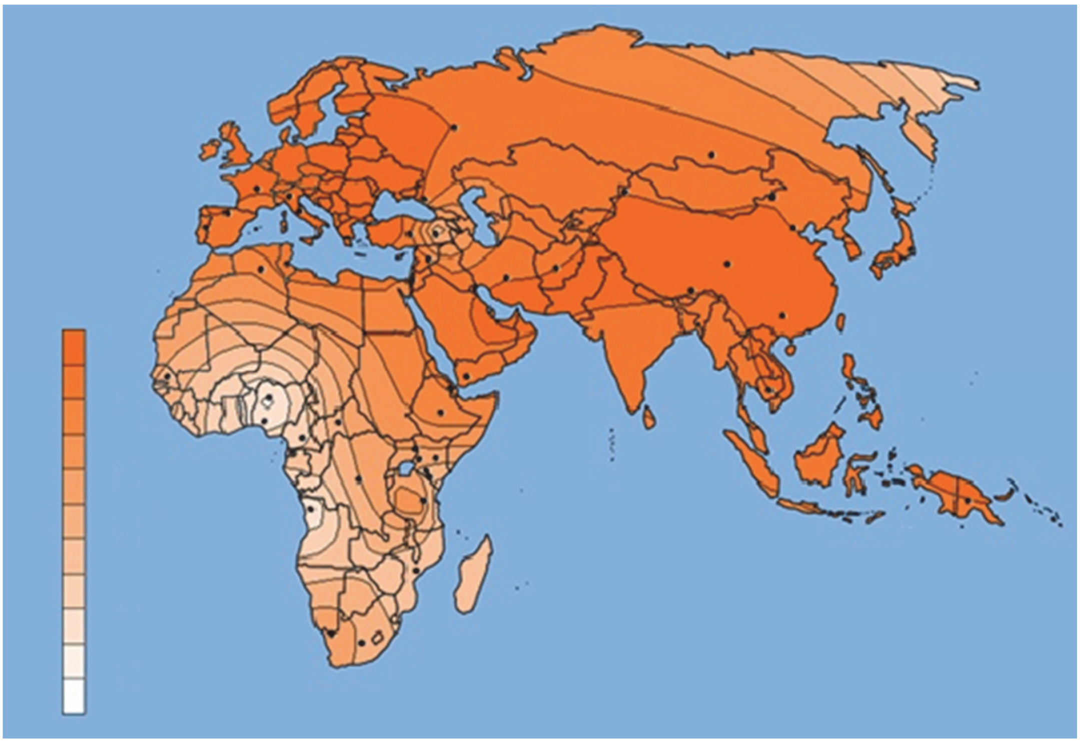


Figure 18. T516 frequency contour map in Africa and Eurasia.

of people (Diamond 1997; de Filippo, et al. 2012). The Bantu diffusion was strongly associated with the spread of agriculture, as well as with the replacement (at least at large extent) and/or assimilation by the Bantu-speakers of groups that existed all over sub-Saharan Africa, before the Bantu expansion.

The elevated frequency of the low bitter sensitivity allele in agrarian populations seems to indicate that in farmers the detection of compounds containing cyanogenic glycosides was less critical than in hunter-gatherers and pastoralists.

To explain the near-fixation of allele T516 outside Africa, we can recall the hypothesis of Campbell, et al. (2013), holding that when modern humans started the migration out of Africa, facing new environments, some pressure was exerted favoring carriers of this allele associated with increased sensitivity to bitter taste compounds. However, we believe that it is not possible to rule out a simple demographical explanation for such distribution, since it seems equally like that it had resulted from the bottleneck effects associated with the migration of modern humans out of Africa.

TAS2R38

In *TAS2R38* we have examined the distribution of the haplotypes previously identified as responsible for a large fraction of the phenotypic variance in PTC/PROP bitter perception (Wooding, et al. 2004). They are defined by the allelic status of three non-synonymous variants C145G, C785T and G886A, which result in the amino acid substitutions P49A, A262V and V296I, respectively. Haplotypes are usually described by the first 3 letters referring to the encoded amino acids. It has been demonstrated that the ancestral configuration PAV is the “taster” haplotype whereas the derived AVI configuration is the “non-taster” haplotype (Wooding, et al. 2004).

The results obtained in the 6 populations analyzed are present in Table 13, showing that 8 different haplotypes were detected, out of which only two were shared by all populations: the ancestral PAV haplotype and the triple derived AVI. The first, the taster haplotype, was well represented in all populations, attaining the maximum value in the Mozambicans (54.8%) followed by the Equatorial Guineans (50.7%); its lowest frequency was detected in the Baka Pigmies from Gabon (28.6 %). The non-taster AVI haplotype was much less frequent in any of the African groups comparatively to the Portuguese, where its frequency was only a little bit lower (0.46) than PAV (0.50).

Table 13. Haplotype frequencies based on three variants (P49A, A262V and V296I) located in *TAS2R38* gene.

Population	PAV	AAV	AAI	AVI	PAI	PVV	PVI	AVV	Haplotype diversity
EQG (n=75)	0.50655	0.01345	0.17989	0.29333	0.00678	-	0.29333	-	0.7279 ± 0.0147
BPY (n=35)	0.28571	-	0.54286	0.17143	-	-	-	-	0.6029 ± 0.0379
UGN (n=46)	0.38167	0.03542	0.29222	0.24213	-	0.03641	0.01216	-	0.7111 ± 0.0212
ANG (n=32)	0.30498	0.08856	0.25325	0.25788	0.01988	0.02281	0.03566	0.01699	0.7708 ± 0.0236
MOZ (n=31)	0.54781	-	0.26254	0.16850	-	0.02116	-	-	0.6033 ± 0.0457
PTG (n=24)	0.50000	0.04167	-	0.45833	-	-	-	-	0.5496 ± .02970

Populations' abbreviations are referred in Material and Methods.

Apart from these two haplotypes, six other were detected, AAV, AAI, PAI, PVV, PVI, and AVV, usually considered to be intermediate haplotypes in terms of taste perception. In the Portuguese, however, only one such haplotype was present at very low frequency (AAV, 4.2%), explaining so the lowest level of haplotype diversity for the locus estimated in this work.

In contrast, African populations presented consistently high number and frequency of intermediate haplotypes, deserving note haplotype AAI, whose frequency even exceeded that of the non-taster haplotype in the samples from Uganda, Mozambique and especially in the Baka Pygmies (BPY), among whom it was by far the most prevalent haplotype (54.3%). As such, the BPY were the less diverse of the African groups, among which haplotype diversity was typically higher than in the Portuguese.

Most of the intermediate haplotypes here detected had been already described in other populations from different geographical regions, mainly from Africa, but also from Eurasia, where at least AAI, AVI, AAV and PAI were before reported at very low frequencies (Wooding et al, 2004, Wooding et al, 2010, Campbell, et al. 2012). However, to the extent of our knowledge, the intermediate PVV and AVV haplotypes were here firstly described. Even in the work of Campbell, et al. (2012), which represents the most comprehensive pan-Sub Saharan African study up to now performed on *TAS2R38*, those two haplotypes escaped detection, likely due to the small size of each individual population sample.

As a whole, the distribution of *TAS2R38* haplotypes observed in this work fitted very well previously reported data for other African and Eurasian populations (references in Table S7).

Regarding the African groups here studied, significant differentiations were observed between the farmers from Mozambique and those from Equatorial Guinea, who also differed significantly from the pastoralist from Uganda as well as from the Baka Pygmies; distance between the Baka Pygmies and the Mozambicans was also significant (Table 14). Although these results pinpoint to some structure in Africa, seemingly it was not related with lifestyle of populations.

Table 14. Pairwise differences between the six populations addressed in this study based on haplotype frequencies defined by P49A, A262V and V296I in *TAS2R38*.

	EQG	**BPY**	**UGN**	**ANG**	**MOZ**	**PTG**
EQG	*					
BPY	**0.13414**	*				
UGN	**0.04090**	0.04379	*			
ANG	0.03308	0.05885	-0.00460	*		
MOZ	**0.05631**	**0.10022**	0.01495	0.04082	*	
PTG	**0.07915**	**0.25799**	**0.08935**	0.07808	**0.10718**	*

values in **bold** are statistically significant after correction for multiple tests; Populations' abbreviations are referred in Material and Methods.

This was further reinforced when the analysis was extended to other African populations.

In Table S11 from supplementary material are listed the pairwise F_{ST} distances based on *TAS2R38* haplotypes that were used to construct the MDS plot shown in Figure 19.

The majority of significant differentiations involved two groups of Baka Pygmies: the sample here screened from Gabon (BPY) and that retrieved from the Alfred database consisting of Baka Pygmies from Central African Republic (PCA). Their differentiation in the panorama of African diversity is quite visible in the plot of Figure 19, yet such differentiation cannot be justified by a shared economy of subsistence (hunter-gathering), since both groups significantly differed from other African hunter-gatherers like the Hadza from Tanzania (HAD) and the Baka Pygmies from Cameroon (PWC).

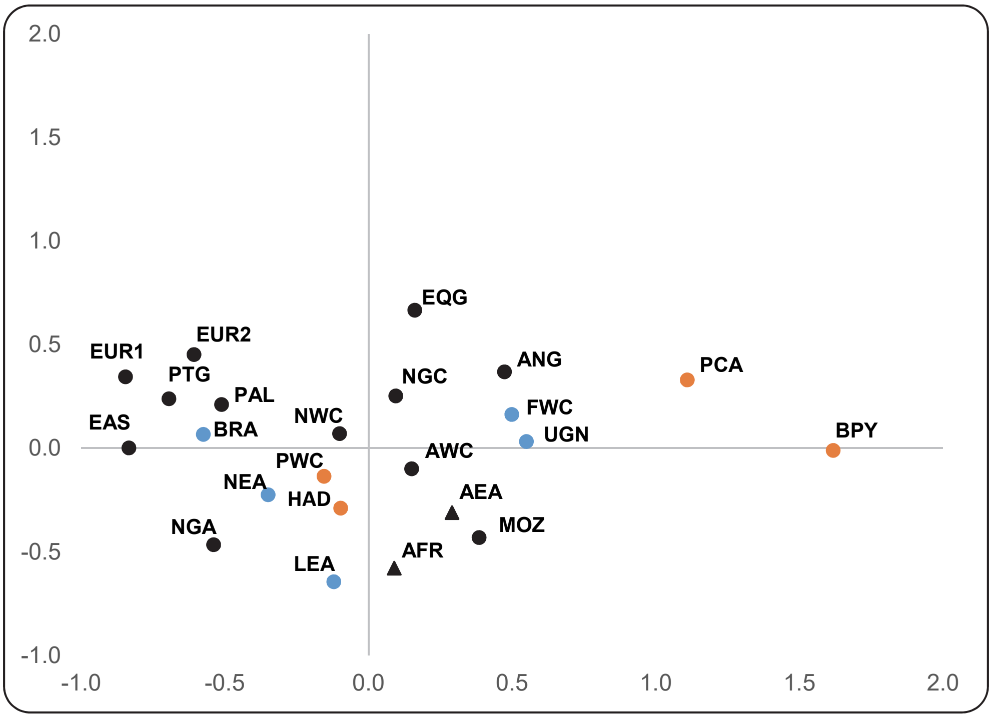


Figure 19. MDS plot based on F_{ST}s for *TAS2R38*.
In the MDS plot different colors represent distinct lifestyles: hunter-gatherers (orange), herders (blue) and farmers (black). Populations' abbreviations are referred in Material and Methods.

In the MDS plot of Figure 19, it is further shown that Eurasian populations tend to cluster near each other in the first quadrant of the plot, only slightly differentiated from the African populations. Actually, at *TAS2R38* locus it was registered one of the less pronounced differentiation between African and non-African populations out of the five examined *TAS* regions. This small differentiation at *TAS2R38* can be explained because both in Africans and Eurasians the two opposite haplotypes PAV and AVI occur at moderate/high frequencies. In turn, what seems to account more for the discrimination of Eurasians relatively to Africans, is the nearly absence of haplotypes associated with intermediate bitter taste sensitivity outside of Africa, whereas contrarily in Africa they can be found at frequencies comparatively much higher across different groups, regardless of mode of subsistence.

A last aspect that deserves to be commented is that within the cluster of African populations (spread across all quadrants of the plot), they are considerably more dispersed than populations in the cluster of non-African ones (restricted to the plot' 1st quadrant), meaning thus that Africa harbors much more population diversity at *TAS2R38* comparatively to other continents. Although this interpretation might appear compromised by the unbalanced number of African versus non-African populations used in the analysis (which in fact was only determined by the data available), it is in agreement with the recent study of Campbell et al, (2012) demonstrating that African populations have a wider range of bitter taste sensitivity than is observed outside the continent.

Contrary to the usual, in the case of *TAS2R38* there are currently more populations screened inside Africa than in other regions. The few Eurasian samples available *a*re scattered throughout entire Europe and Asia, but even so they do not encompass all different lifestyles that still are practiced at least in Asia, meaning that future studies are still needed to obtain a better assessment of the worldwide distribution of diversity at this locus.

Relationship between taste and geography/lifestyle: clues from AMOVA

In order to evaluate the relative contribution of geography and lifestyle to the apportionment of the diversity in each *TAS* region, we have performed AMOVA assuming different hierarchical levels for clustering populations. These analyses, however, were highly conditioned by the scarcity of data for *TAS1R1*, *TAS1R3* and the promoter region of *TAS1R3*. Concerning the last, it was not possible to test lifestyle with the small set of samples, since additionally to ours only three more were available respecting general samples from Africans, Asians and Europeans (see Table S7).

In the first runs of AMOVA, we considered all populations from Africa, Europe and Asia using the continent of origin as the criterion to assess the role of geography. The results are shown in Table 15A, revealing that geography does not influence distribution of diversity at *TAS1R1* or at *TAS1R3*. It significantly accounts to 4.58% of the diversity at *TAS2R38*, and explains as much as 19.37% and 28.24% of the total diversity at *TAS2R16* and *TAS1R3* promoter, respectively. In this broad geographical context, the contribution of lifestyle was only found to be marginally significant in *TAS2R38*, yet explaining merely 1.85% of total diversity at the *locus*. The meaning of this is result, raises, however, serious reservations since it might be highly biased due to the underrepresentation on non-African populations (only six samples, including the Portuguese), among which hunter-gatherers were lacking.

Table 15A. Proportion of among groups (F_{CT}) variance provided by AMOVA using geography and lifestyle to group African, Asian and European populations, and associated *P*-values.

	GEOGRAPHY		**LIFESTYLE**	
	F_{CT}	*P-value*	F_{CT}	*P-value*
TAS1R1	-0.68	0.49792±0.00504	-0.54	0.56574±0.00471
TAS1R3*	**28.24**	**0.00426±0.00066**	NA	-
TAS1R3**	2.36	0.09911±0.00262	-1.08	0.54426±0.00518
TAS2R16	**19.37**	**0.00000±0.00000**	-0.92	0.74733±0.00392
TAS2R38	**4.58**	**0.00139±0.00045**	**1.85**	**0.04297±0.00180**

* promoter and ** genic region of *TAS1R3*; values in **bold** are statistically significant after correction for multiple tests

Then we performed a 2nd series of AMOVA runs restrictively in Africa, assuming the following 5 sub-regions in the continent: Northern, Western, Central, Eastern and Southern Africa, as before discriminated in Table S7. The results shown in Table 15B, indicate that the geographical proximity of populations in Africa was broadly irrelevant to the patterns of genetic diversity at the *TAS loci*, excepting in *TAS2R16* concerning which geography was found to play a significant role, though explaining only 3.25% of the locus total diversity in Africa.

Table 15B. Proportion of among groups (F_{CT}) variance provided by AMOVA using geography and lifestyle to group African populations, and associated *P*-values.

	GEOGRAPHY		**LIFESTYLE**	
	F_{CT}	*P-value*	F_{CT}	*P-value*
TAS1R1	0.10	0.30010±0.00476	0.56	0.49644±0.00491
TAS1R3*	-0.84	0.52475±0.00496	4.97	0.10416±0.00326
TAS1R3**	-5.79	0.88396±0.00347	4.81	0.19950±0.00435
TAS2R16	**3.25**	**0.00139±0.00035**	1.21	0.10851±0.00297
TAS2R38	1.19	0.11079±0.00253	1.15	0.13554±0.00301

* promoter; ** genic region of *TAS1R3*; values in **bold** are statistically significant after correction for multiple tests

Similarly irrelevant appeared to be sharing a lifestyle, factor that was never associated to any significant difference among groups. It must be stressed, that only for *TAS2R16* and

TAS2R38 the number of African populations available was enough to produce reliable AMOVA results, which in fact failed to capture the trend between lifestyle and genetic diversity at *TAS2R16,* that appeared to emerge when a less objective analysis of the data was performed and discussed in the previous point referring to *TAS2R16*.

Genetic differentiation among global populations

Global genetic distances (F_{ST}s) between populations from Africa, Europe and Asia calculated with allele or haplotypes frequencies for each gene/promoter are presented in Figure 20, in which are also shown the F_{ST}s values obtained when we used full sequences for the same genomic regions retrieved from the 1000 Genomes Project database, as will be presented in the next section of results (section 3.3).

In general, the measures of differentiation across Africa, Europe and Asia were rather similar when considering specific SNPs in each region or the 1000 Genomes full sequences. Whatever the approach, *TAS2R16* and especially the promoter of *TAS1R3* were characterized by global levels of differentiation substantially more elevated than in *TAS1R1*, *TAS1R3* and *TAS2R38*. The last *locus* was the one revealing the slightest structure at a global continental level.

The great discrepancy in F_{ST}s was observed in *TAS1R1*, since the differentiation decreased almost half when based on the haplotypes defined by the two studied SNPs, (F_{ST}=0.057) comparatively to the value estimated with full sequences (F_{ST}=0.103). This can be largely explained by the fact of having dealt with two SNPs at the locus that were in strong LD and, most importantly, were characterized by very reduced haplotype diversity.

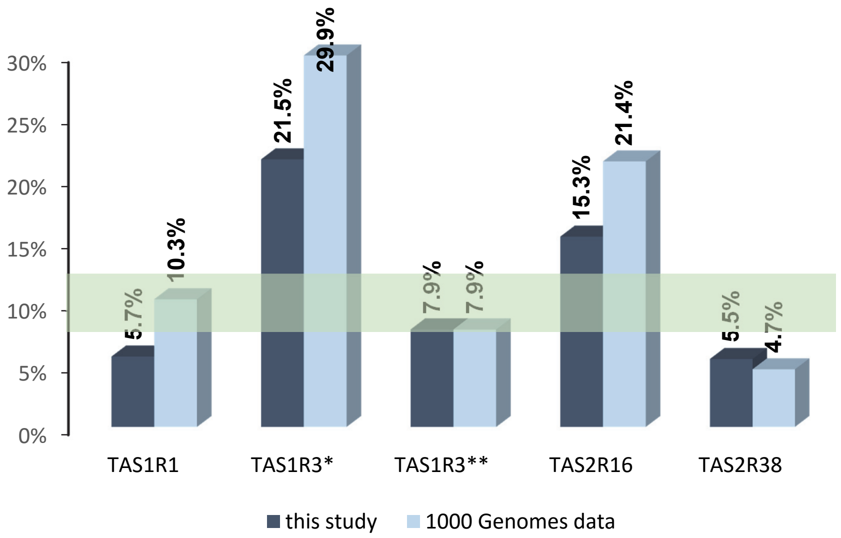


Figure 20. Genetic variation among gene taste receptors.
* promoter; ** genic region of *TAS1R3*; the green shaded area corresponds to F_{ST}=8-13%

Previous studies based on the analysis of wide numbers of randomly selected SNPs in the genome have indicated that the average level of population structure across Africa, Europe and Asia is around ~8-13% (Wright's fixation index, F_{ST}) (Akey, et al. 2002; Shriver, et al. 2004; Elhaik 2012). Such value range is also represented in Figure 20, functioning as a referential to analyze the differentiations registered at different *TAS* regions (green shaded bar). The most remarkable departures from the average range were observed at the promoter of *TAS1R3* implied in sweet perception, and at *TAS2R16*, influencing bitter taste sensitivity, both revealing an unusual high level of overall differentiation. In the opposite direction, was the gene *TAS2R38,* also associated with bitter perception, which disclosed a lower structure across Africa, Europe and Asia than commonly found.

3.2.4 Conclusions

This work provided a contribution to enlarge the knowledge on a fraction of genetic diversity that influences taste perception. It mainly focused on African populations with different lifestyles, assumed as proxies for different dietary regimens, having as ultimate goal to investigate whether any relationship existed between patterns of diversity at five *TAS* regions and main diets of populations. The study was motivated on the one hand because it is well demonstrated that genetic differences in taste perception substantially account for individual differences in food preferences, and on the other because signatures of selection have been identified in several *TAS* genes (Kim, et al. 2006; Campbell, et al. 2012; Campbell, et al. 2013).

Unfortunately, the scarcity of data available for the different TAS regions, in general and not only for Africa was far from enough to obtain the global scenario needed to reach more clear answers on the matter. Even so, no evidence arose that the subsistence economies, or, indirectly, diet-related variables in populations, could be associated with the distribution of diversity at *TAS1R1*, *TAS1R3* genic/promoter regions and *TAS2R38*.

Concerning *TAS2R16*, the strong structure observed in Africa, raises the question on the main factors underling such pattern. Although lifestyle failed to produce significant values in AMOVA, it remains to be clarified why the prevalence of an ancestral variant, associated with diminished sensibility to cyanogenic glycosides, peaks in farmer groups, particularly from the region that corresponds to the center of origin of the Bantu dispersion. The pattern of distribution of that variant In Sub-Saharan Africa suggests that its spread was mediated by the expansion of Bantu-speakers and their agricultural way of life. Assuming this reconstruction, the decrease in allele frequency from the core region of the Bantu dispersion, can likely be explained by the demographical history of the farmer populations. What still needs to be elucidated is whether the observation of high frequencies of the ancestral allele restrictively in farmer groups from West Central Africa, might be a signal of local adaptation guided by a selective factor of nature difficult to discern for now.

Also in *TAS2R16*, another finding that called attention was the unexpectedly high differentiation across Africa, Asia and Europe, feature that was even exceeded by the level of substructure detected in the case of the promoter of *TAS1R3*. For these two *loci*, the excessive differentiation across continental populations does not easily comply with the ancient population structure shaping contemporary genetic variation that can be explained by the past demographical history of human populations, during which modern humans first migrated out of Africa to colonize entire Asia and Europe. Instead, the sharp structure observed at *TAS2R16* and the promoter of *TAS1R3*, are patterns that *per se* fit better the expectations of population models incorporating the action of positive selection.

Contrariwise, the level of differentiation registered at *TAS2R38* was too low comparatively to the average genome level, which is an effect that in terms of worldwide population structure is predicted for *loci* evolving under balancing selection.

In summary, and safeguarding any misinterpretation induced by the scarcity of data, no connection emerged between the distribution of genetic diversity at five *TAS* regions and mode of subsistence in populations. Extrapolating the result to differences in main food consumption in which populations rely, no sign emerged that the *TAS* variations examined could represent dietary adaptations. Nevertheless, most of the *TAS* regions showed intriguing patterns of geographical distribution, which in reality are difficult to reconcile with pure demographic explanations.

It is now demonstrated that a number of TAS1R and TAS2R proteins exert functions unsuspected until recently, that largely extend the well establish role as taste receptors.

Genes from the TAS1R/TAS2R families besides expressed in organs of the gastrointestinal system are also expressed in several organs and tissues not implicated in food intake or digestion, such as the lung, the airway smooth muscle, the nose, the testis and the hypothalamus (Campa, et al. 2012; Welcome, et al. 2015).

In addition, there is growing evidence indicating that taste genes play a much broader role in human health. Suggestive associations are being disclosed between allelic variants in taste receptor genes and body mass index, complex diseases such as cancer, alcohol consumption, smoking and nicotine dependence, longevity (reviewed in Campa, et al. (2012).

In the future, a better understanding of the involvement of taste genes in the many aspects that are crucial to health and survival will be certainly achieved, which hopefully will bring new clues on the selective factors that might have shaped diversity at TAS genes during human evolution.

3.2.5 Supplementary Material

Table S5. Amplification primers sequences.

Gene	SNPs	Primer forward	Primer reverse	Size (bp)
TAS1R1	C329T (rs41278020)	TCAATGAGCATGGCTACCAC	CACCGTAGGGGAATAGTGGA	218
	G1114T (rs34160967)	CCTGAAGGCGTTTGAAGAAG	GGCAGAACTCATGGAGAAGG	160
TAS1R3	G13A (rs76755863)	CCTGTTGGAAGTTGCCTCTG	ACGTAGTCCCCCTTCATCCT	129
	C2269T (rs307377)	CTGGCCTTTCTCTGCTTCCT	CAAAGGAGACCCAGGTGATG	119
	-T1572C (rs307355); -T1266C (rs35744813)	CGTGTGTGCTGTGAGCGTA	AATATGGCGCACATGCGAA	520
TAS2R16	G516T (rs846664)	GGCTGAGGTGGAGAATTTTG	CCAGGAACAGGATGAAAGGA	231
TAS2R38	C145G (rs713598)	CAATGCCTTCGTTTTCTTGGTG	GATGGCTTGGTAGCTGTGGT	190
	C785T (rs1726866); G886A (rs10246939)	CCCACATTAAAGCCCTCAAG	TCTCCTCAACTTGGCATTGC	197

Table S6. Mini-sequencing primers sequences.

Gene	SNPs	Minisequensing primer seq 5'>3'	Size (bp)	Detection	Mutation	Size (bp)
TAS1R1	C329T (rs41278020)	TATGATGTGTGTTCTGACTCTG	22	C/T	C/T	22
	G1114A (rs34160967)	AATGCCAAGCTTTCATG	17	G/A	G/A	17
TAS1R3	G13A (rs76755863)	aaagtctgacaaCTGAGGCCCAGGACAG	16	C/T	G/A	28
	C2269T (rs307377)	tgccacgtcgtgaaagtctgacaaCGGGCACGGTTGTAGC	16	G/A	C/T	40
	-T1572C (rs307355)	aactgactaaactaggtgccacgtcgtgaaagtctgacaaACATGGTACACGCAAAGC	18	A/G	T/C	58
	-T1266C (rs35744813)	actaaactaggtgccacgtcgtgaaagtctgacaaCGGCACACATGCATGCC	17	A/G	T/C	52
TAS2R16	G516T (rs846664)	agtctgacaaaactgactaaactaggtgccacgtcgtgaaagtctgacaaTGGAACTGATACTGATGAAA	20	C/A	G/T	70
TAS2R38	C145G (rs713598)	gacaaaactgactaaactaggtgccacgtcgtgaaagtctgacaaGGATGTAGTGAAGAGGCAG	19	C/G	C/G	64
	C785T (rs1726866)	cgtgaaagtctgacaaaactgactaaactaggtgccacgtcgtgaaagtctgacaaCTTTGTGATATCATCCTGTG	20	C/T	C/T	76
	G886A (rs10246939)	actaggtgccacgtcgtgaaagtctgacaaCTCTGGGCATGCAGCC	16	G/A	G/A	46

aactgactaaactaggtgccacgtcgtgaaagtctgacaa is the tale used by convention

Table S7. Populations used in the comparative analyses.

			AMOVA groups classification		
	Population	**n**	***mode of subsistence***	***geography***	**reference**
TAS1R1/TAS1R3*	Equatorial Guinean	75	agriculturalist	Central Africa	*this study*
	Baka Pygmies	36	hunter-gatherer	Central Africa	*this study*
	Ugandan	44	pastoralist	East Africa	*this study*
	Angolan	36	agriculturalist	West Africa	*this study*
	Mozambican	28	agriculturalist	East Africa	*this study*
	Portuguese	21	agriculturalist	Europe	*this study*
	Cameroonian	20	NA	Central Africa	Kim et al. 2006
	North European	10	agriculturalist	Europe	Kim et al. 2006
	Chinese	10	agriculturalist	East Asia	Kim et al. 2006
	Hungarian	10	agriculturalist	Europe	Kim et al. 2006
	Japanese	10	agriculturalist	East Asia	Kim et al. 2006
	Pakistanis	8	pastoralist	Middle East	Kim et al. 2006
	Russian	10	agriculturalist	Europe	Kim et al. 2006
TAS1R3*	Equatorial Guinean	75	agriculturalist	Central Africa	*this study*
	Baka Pygmies	36	hunter-gatherer	Central Africa	*this study*
	Ugandan	44	pastoralist	East Africa	*this study*
	Angolan	36	agriculturalist	West Africa	*this study*
	Mozambican	28	agriculturalist	East Africa	*this study*
	Portuguese	21	agriculturalist	Europe	*this study*
	African	15	NA	Africa	Fushan et al. 2009
	European	92	NA	Europe	Fushan et al. 2009
	Asian	37	NA	Asia	Fushan et al. 2009
TAS2R16	Equatorial Guinean	75	agriculturalist	Central Africa	*this study*
	Baka Pygmies	36	hunter-gatherer	Central Africa	*this study*
	Ugandan	44	pastoralist	East Africa	*this study*
	Angolan	36	agriculturalist	West Africa	*this study*
	Mozambican	28	agriculturalist	East Africa	*this study*
	Portuguese	21	agriculturalist	Europe	*this study*
	Tunisian	17	pastoralist	North Africa	Soranzo et al. 2005
	Bantu (SA)	28	agriculturalist	South Africa	Soranzo et al. 2005
	Amhara (Ethiopia)	20	agriculturalist	East Africa	Soranzo et al. 2005
	Cameroonian	12	agriculturalist	Central Africa	Soranzo et al. 2005
	Yoruba (Nigeria)	147	agriculturalist	West Africa	Hapmap
	Luhya (Kenya)	109	agriculturalist	East Africa	Hapmap
	Masaai1 (Kenya)	156	pastoralist	East Africa	Hapmap
	CEU	65	agriculturalist	Europe	Hapmap
	Chinese	45	agriculturalist	East Asia	Hapmap
	Japanese1	44	agriculturalist	East Asia	Hapmap
	Bantu (Kenya)	12	agriculturalist	East Africa	Hinrichs et al. 2006
	Bantu (SE)	8	agriculturalist	South Africa	Hinrichs et al. 2006
	Biaka Pygmies	36	hunter-gatherer	Central Africa	Hinrichs et al. 2006

Table S7 (continued)

	Mandenka	24	agriculturalist	West Africa	Hinrichs et al. 2006
	Mbuti Pygmies	15	hunter-gatherer	Central Africa	Hinrichs et al. 2006
	Bantu (NE)	12	agriculturalist	North Africa	Hinrichs et al. 2006
	San	7	hunter-gatherer	South Africa	Hinrichs et al. 2006
	Mozabite	30	pastoralist	North Africa	Hinrichs et al. 2006
	Palestinian	50	agriculturalist	Middle East	Hinrichs et al. 2006
	Bedouin	47	pastoralist	South Asia	Hinrichs et al. 2006
	Brahui	24	pastoralist	South Asia	Hinrichs et al. 2006
	Adygei	17	agriculturalist	Middle East	Hinrichs et al. 2006
	Balochi	26	pastoralist	East Asia	Hinrichs et al. 2006
	Makrani	26	pastoralist	South Asia	Hinrichs et al. 2006
	Sindhi	24	agriculturalist	South Asia	Hinrichs et al. 2006
	Burusho	25	agriculturalist	South Asia	Hinrichs et al. 2006
	Hazara	22	pastoralist	South Asia	Hinrichs et al. 2006
	Kalash	25	agriculturalist	South Asia	Hinrichs et al. 2006
	Pathan	25	agriculturalist	South Asia	Hinrichs et al. 2006
	Basque	30	agriculturalist	Europe	Hinrichs et al. 2006
	French	24	agriculturalist	Europe	Hinrichs et al. 2006
	Bergamo	27	agriculturalist	Europe	Hinrichs et al. 2006
	Sardinian	14	agriculturalist	Europe	Hinrichs et al. 2006
	Tuscan	8	agriculturalist	Europe	Hinrichs et al. 2006
	Cambodian	11	agriculturalist	East Asia	Hinrichs et al. 2006
	Dai	10	agriculturalist	East Asia	Hinrichs et al. 2006
	Lahu	10	agriculturalist	East Asia	Hinrichs et al. 2006
	Naxi	10	agriculturalist	East Asia	Hinrichs et al. 2006
	Daur	10	agriculturalist	East Asia	Hinrichs et al. 2006
	Druze	47	agriculturalist	Middle East	Hinrichs et al. 2006
	Han	45	agriculturalist	East Asia	Hinrichs et al. 2006
	Hezhen	10	hunter-gatherer	East Asia	Hinrichs et al. 2006
	Orogen	10	hunter-gatherer	East Asia	Hinrichs et al. 2006
	Tu	10	pastoralist	South Asia	Hinrichs et al. 2006
	Miaozu	10	agriculturalist	East Asia	Hinrichs et al. 2006
	Uygur	10	pastoralist	South Asia	Hinrichs et al. 2006
	Xibo	10	agriculturalist	South Asia	Hinrichs et al. 2006
	Mongola	10	pastoralist	East Asia	Hinrichs et al. 2006
	Orcadian	16	agriculturalist	Europe	Hinrichs et al. 2006
	Russian	25	agriculturalist	Europe	Hinrichs et al. 2006
	She	10	agriculturalist	East Asia	Hinrichs et al. 2006
	Tujia	10	agriculturalist	East Asia	Hinrichs et al. 2006
	Yizu	10	agriculturalist	East Asia	Hinrichs et al. 2006
	Yakut	25	pastoralist	East Asia	Hinrichs et al. 2006
	Japanese2	31	agriculturalist	East Asia	Hinrichs et al. 2006
	Hausa (Nigeria)	39	agriculturalist	West Africa	ALFRED
	Ibo (Nigeria)	48	agriculturalist	West Africa	ALFRED
	Lisongo (CAR)	8	hunter-gatherer	Central Africa	ALFRED

	Chagga (Tanzania)	45	agriculturalist	East Africa	ALFRED
	Masaai2 (Tanz-Ken)	20	pastoralist	East Africa	ALFRED
	Sandawe (Tanzania)	40	hunter-gatherer	East Africa	ALFRED
	Zaramo (Tanzania)	39	agriculturalist	East Africa	ALFRED
	Kuwait	16	agriculturalist	Middle East	ALFRED
	Tsaatan	42	pastoralist	East Asia	ALFRED
TAS2R38	Equatorial Guinean	75	agriculturalist	Central Africa	*this study*
	Baka Pygmies	36	hunter-gatherer	Central Africa	*this study*
	Ugandan	44	pastoralist	East Africa	*this study*
	Angolan	36	agriculturalist	West Africa	*this study*
	Mozambican	28	agriculturalist	East Africa	*this study*
	Portuguese	21	agriculturalist	Europe	*this study*
	African	31	NA	Africa	Wooding et al. 2004
	European1	55	agriculturalist	Europe	Wooding et al. 2004
	European2	50	agriculturalist	Europe	Wooding et al. 2010
	Pigmies (CAR)	67	hunter-gatherer	Central Africa	ALFRED
	Afroasiatic (Cam.)	73	agriculturalist	Central Africa	Campbell et al. 2012
	Nilosaharan (Cam.)	26	agriculturalist	Central Africa	Campbell et al. 2012
	NigerKordofanian (Cam.)	62	agriculturalist	Central Africa	Campbell et al. 2012
	Fulani (Cam.)	48	pastoralist	Central Africa	Campbell et al. 2012
	Pigmies (Cam.)	62	hunter-gatherer	Central Africa	Campbell et al. 2012
	Afroasiatic (Kenya)	132	NA	East Africa	Campbell et al. 2012
	Nilosaharan (Kenya)	130	pastoralist	East Africa	Campbell et al. 2012
	NigerKordofanian (Kenya)	34	agriculturalist	East Africa	Campbell et al. 2012
	Luo (Kenya)	21	pastoralist	East Africa	Campbell et al. 2012
	Hadza (Tanzania)	23	hunter-gatherer	East Africa	Campbell et al. 2012
	Palestinians	20	agriculturalist	Middle East	Campbell et al. 2012
	Brahui	8	pastoralist	South Asia	Campbell et al. 2012
	East Asian	71	agriculturalist	East Asia	Campbell et al. 2012

NA - data not available

Table S8. Hardy-Weinberg Equilibrium test for *TAS1R* family genes.

	C329T (*TAS1R1*)			**G114A (*TAS1R1*)**			**G13A (*TAS1R3*)**			**C2269T (*TAS1R3*)**			**-T1572C (*TAS1R3* promoter*)***			**-T1266C (*TAS1R3* promoter)**		
POP	H_{OB}	H_{EX}	***P*-value**	H_{OB}	H_{EX}	***P*-value**	H_{OB}	H_{EX}	***P*-value**	H_{OB}	H_{EX}	***P*-value**	H_{OB}	H_{EX}	***P*-value**	H_{OB}	H_{EX}	***P*-value**
EQG (n=75)	0.0400	0.0395	1.0000	0.1067	0.1017	1.0000	0.3467	0.3529	1.0000	0.0400	0.0395	1.0000	0.0400	0.0395	1.0000	0.3200	0.4711	0.0071
BPY (n=36)		*		0.1765	0.1633	1.0000	0.1177	0.1124	1.0000		*			*		0.6177	0.5070	0.3010
UGN (n=47)	0.0000	0.0449	0.0117	0.0909	0.1285	0.1652	0.0909	0.2832	0.0000	0.0227	0.0227	1.0000	0.0455	0.1285	0.0058	0.2273	0.3848	0.0067
ANG (n=37)		*		0.0357	0.0357	1.0000	0.3929	0.3630	1	0.0357	0.0357	1.0000	0.0385	0.0385	1.0000	0.2308	0.4615	0.0236
MOZ (n=31)		*		0.0357	0.0357	1.0000	0.3929	0.4565	0.6719	0.0357	0.0357	1.0000	0.0357	0.0357	1.0000	0.4286	0.4156	1
PTG (n=24)	0.1000	0.0974	1.0000	0.2000	0.1846	1.0000	0.04348	0.26377	0.00151	0.1500	0.1423	1.0000	0.1429	0.1359	1.0000	0.1429	0.2149	0.2348

*H*OB – Heterozygozity observed; *H*EX – Heterozygozity expected; * – this locus is monomorphic. Significant differences, after Bonferroni's correction for multiple tests are highlighted in **bold**; populations' abbreviations as referred in Material and Methods.

Table S9. Hardy-Weinberg Equilibrium test for *TAS2R* family genes.

	G516T (*TAS2R16*)			**P49A (*TAS2R38*)**			**A262V (*TAS2R38*)**			**V296I (*TAS2R38*)**		
POP	H_{OB}	H_{EX}	***P*-value**	H_{OB}	H_{EX}	***P*-value**	H_{OB}	H_{EX}	***P*-value**	H_{OB}	H_{EX}	***P*-value**
EQG (n=75)	0.2533	0.4117	0.0016	0.5467	0.5030	0.4921	0.4000	0.4174	0.7813	0.5333	0.5026	0.6477
BPY (n=36)	0.3939	0.3883	1.0000	0.5294	0.4214	0.2142	0.2647	0.2752	1.0000	0.5294	0.4214	0.2147
UGN (n=47)	0.23404	0.24091	1.0000	0.5349	0.5092	0.5262	0.37778	0.40574	0.71294	0.5581	0.5034	0.5460
ANG (n=37)	0.4000	0.4702	0.6620	0.5000	0.4857	1.0000	0.2857	0.4442	0.0841	0.5357	0.5033	1.0000
MOZ (n=31)	0.3214	0.3994	0.3457	0.5714	0.4987	0.4681	0.1786	0.2747	0.1128	0.5714	0.4987	0.4693
PTG (n=24)		*		0.4762	0.5110	1.0000	0.4762	0.5110	1.0000	0.4762	0.5110	1.0000

*H*OB – Heterozygozity observed; *H*EX – Heterozygozity expected; * – this locus is monomorphic. Significant differences, after Bonferroni's correction for multiple tests are highlighted in **bold**; populations' abbreviations as referred in Material and Methods.

Table S10. Pairwise differences between the six populations addressed in this study and other comparisons populations, based on allele frequencies of G516T from *TAS2R16* gene.

	EQG	BPY	UGN	ANG	MOZ	PTG	TUN	MOB	BST	AMH	BSE	CAM	MPY	MAN	BIP	MAS1	CEU	CHB	JPT1	YOR	LUH	BSE	SAN	PAL	BED	BRA	ADY	BAL	MAK	SIN	BUR	HAZ	KAL	PAT	BSQ
EQG	*																																		
BPY	-0.00908	*																																	
UGN	0.05298	0.03859	*																																
ANG	-0.01032	-0.01214	0.05915	*																															
MOZ	-0.01311	-0.01530	0.05712	-0.01681	*																														
PTG	**0.19308**	**0.20633**	0.08863	**0.23160**	**0.25212**	*																													
TUN	0.18418	0.19176	0.08026	**0.21614**	**0.23261**	0.00000	*																												
MOB	0.08155	0.06958	-0.00795	0.09297	0.09387	0.06747	0.05851	*																											
BST	0.00106	-0.00934	0.00712	-0.00035	-0.00405	0.17555	0.16012	0.03189	*																										
AMH	0.09916	0.08895	-0.00013	0.11256	0.11532	0.05364	0.04376	-0.01738	0.05059	*																									
BSE	0.01210	-0.00213	-0.02336	0.01240	0.00866	0.18958	0.16513	-0.00837	-0.02337	0.00923	*																								
CAM	-0.01933	-0.01631	0.09863	-0.02432	-0.02648	**0.38798**	**0.35392**	0.15168	0.00762	0.18154	0.03106	*																							
MPY	-0.00116	-0.01291	-0.00691	-0.00226	-0.00609	0.20829	0.18498	0.01781	-0.02563	0.03935	-0.03514	0.00805	*																						
MAN	0.00183	-0.00913	0.00288	0.00062	-0.00312	0.17959	0.16260	0.02720	-0.01962	0.04633	-0.02653	0.00992	-0.02762	*																					
BIP	0.00555	-0.00544	0.00574	0.00452	0.00081	0.15671	0.14413	0.02807	-0.01602	0.04516	-0.02293	0.01468	-0.02396	-0.01767	*																				
MAS1	0.00459	-0.00483	0.01749	0.00288	-0.00081	0.13238	0.12733	0.03819	-0.01022	0.05324	-0.01216	0.00887	-0.01650	-0.01119	-0.00765	*																			
CEU	**0.26752**	**0.32756**	**0.15559**	**0.35943**	**0.40765**	0.00000	0.00000	**0.13934**	**0.30327**	0.13339	**0.37836**	**0.60917**	**0.38943**	**0.31889**	**0.26183**	**0.16884**	*																		
CHB	**0.23660**	**0.27841**	**0.12790**	**0.30797**	**0.34665**	0.00000	0.00000	0.10965	**0.25159**	0.10046	0.30569	**0.53277**	**0.31947**	**0.26299**	**0.21869**	**0.15405**	0.00000	*																	
JPT1	**0.23496**	**0.27574**	**0.12644**	**0.30516**	**0.34325**	0.00000	0.00000	0.10807	**0.24878**	0.09871	0.30159	**0.52815**	**0.31553**	**0.25993**	**0.21637**	**0.15327**	0.00000	0.00000	*																
YOR	0.00941	0.01480	**0.10829**	0.00387	0.00223	**0.24025**	**0.23402**	**0.13837**	0.04012	**0.15636**	0.06214	-0.02046	0.04149	0.04239	0.04647	**0.04299**	**0.29137**	**0.27005**	**0.26893**	*															
LUH	0.01904	0.00598	0.00290	0.01932	0.01544	0.11270	0.10701	0.01914	-0.00985	0.03243	-0.02140	0.03604	-0.01915	-0.01208	-0.00854	-0.00043	**0.15486**	**0.13766**	**0.13675**	**0.06727**	*														
BSE	-0.03410	-0.03345	0.07211	-0.03861	-0.04114	0.42932	0.38870	0.13004	-0.01414	0.16763	0.00754	-0.05390	-0.01432	-0.01219	-0.00778	-0.01257	**0.66781**	**0.59024**	**0.58539**	-0.02722	0.01094	*													
SAN	0.01252	-0.00348	-0.04299	0.01323	0.00935	0.21495	0.18260	-0.03600	-0.02995	-0.02202	-0.05767	0.03486	-0.04368	-0.03389	-0.03030	**-0.01735**	0.44619	0.36138	0.35643	0.06775	-0.03093	0.01112	*												
PAL	**0.14318**	**0.14391**	0.02268	**0.17384**	**0.18391**	0.02703	0.02193	-0.00189	0.09724	-0.01590	0.04899	**0.28031**	0.08966	0.09343	0.08662	**0.08120**	0.06008	0.04711	0.04642	**0.19589**	0.05928	0.26691	0.01143	*											
BED	**0.20161**	**0.22509**	0.07898	**0.25590**	**0.28161**	-0.00173	-0.00549	0.04795	**0.18558**	0.02194	0.17052	**0.42809**	0.21033	**0.18852**	**0.16329**	**0.12629**	0.01624	0.01007	0.00972	**0.24349**	0.10660	0.44823	0.14700	0.00855	*										
BRA	**0.17016**	**0.17732**	0.05904	**0.20357**	**0.21946**	-0.00283	-0.00745	0.03225	0.14092	0.01058	0.12385	**0.33784**	0.15540	0.14156	0.12588	0.11099	0.02415	0.01397	0.01342	**0.22022**	0.09008	0.35709	0.10691	0.00150	-0.01598	*									
ADY	0.18418	0.19176	0.08026	**0.21614**	0.23261	0.00000	0.00000	0.05851	0.16012	0.04376	0.16513	**0.35392**	0.18498	0.16260	0.14413	0.12733	0.00000	0.00000	0.00000	**0.23402**	0.10701	0.38870	0.18260	0.02193	-0.00549	-0.00745	*								
BAL	**0.20316**	**0.22310**	0.09789	**0.24942**	**0.27451**	0.00000	0.00000	0.07741	**0.19327**	0.06463	0.21746	**0.42541**	0.23492	**0.19911**	**0.17108**	**0.13770**	0.00000	0.00000	0.00000	**0.24720**	**0.11877**	**0.47275**	0.25129	0.03219	0.00171	0.00168	0.00000	*							
MAK	**0.20316**	**0.22310**	0.09789	**0.24942**	**0.27451**	0.00000	0.00000	0.07741	**0.19327**	0.06463	0.21746	**0.42541**	0.23492	**0.19911**	**0.17108**	**0.13770**	0.00000	0.00000	0.00000	**0.24720**	0.11877	**0.47275**	0.25129	0.03219	0.00171	0.00168	0.00000	0.00000	*						
SIN	**0.19923**	**0.21655**	0.09430	**0.24246**	**0.26578**	0.00000	0.00000	0.07356	**0.18635**	0.06037	0.20661	**0.41104**	0.22456	0.19149	0.16547	**0.13567**	0.00000	0.00000	0.00000	**0.24450**	0.11644	**0.45623**	0.23723	0.03024	0.00045	0.00000	0.00000	0.00000	0.00000	*					
BUR	**0.20121**	**0.21985**	0.09611	**0.24596**	**0.27018**	0.00000	0.00000	0.07550	0.18984	0.06252	0.21208	**0.41832**	0.22978	**0.19533**	**0.16830**	**0.13670**	0.00000	0.00000	0.00000	**0.24586**	0.11762	**0.46462**	0.24433	0.03123	0.00110	0.00086	0.00000	0.00000	0.00000	0.00000	*				
HAZ	**0.19517**	**0.20980**	0.09056	**0.23528**	**0.25675**	0.00000	0.00000	0.06955	0.17921	0.05593	0.19537	**0.39588**	0.21382	**0.18363**	**0.15968**	**0.13351**	0.00000	0.00000	0.00000	**0.24170**	0.11398	0.43860	0.22255	0.02814	-0.00096	-0.00184	0.00000	0.00000	0.00000	0.00000	0.00000	*			
KAL	**0.20121**	**0.21985**	0.09611	**0.24596**	**0.27018**	0.00000	0.00000	0.07550	**0.18984**	0.06252	0.21208	**0.41832**	0.22978	**0.19533**	**0.16830**	**0.13670**	0.00000	0.00000	0.00000	**0.24586**	0.11762	**0.46462**	0.24433	0.03123	0.00110	0.00086	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	*		
PAT	**0.20121**	**0.21985**	0.09611	**0.24596**	**0.27018**	0.00000	0.00000	0.07550	**0.18984**	0.06252	0.21208	**0.41832**	0.22978	0.19533	**0.16830**	**0.13670**	0.00000	0.00000	0.00000	**0.24586**	0.11762	**0.46462**	0.24433	0.03123	0.00110	0.00086	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	*	
BSQ	**0.21071**	**0.23568**	0.10472	**0.26278**	**0.29121**	0.00000	0.00000	0.08475	**0.20656**	0.07278	0.23810	**0.45205**	0.25466	**0.21373**	**0.18187**	**0.14151**	0.00000	0.00000	0.00000	**0.25237**	0.12314	**0.50287**	0.27773	0.03577	0.00391	0.00472	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	*
FRE	**0.19923**	**0.21655**	0.09430	**0.24246**	**0.26578**	0.00000	0.00000	0.07356	**0.18635**	0.06037	0.20661	**0.41104**	0.22456	0.19149	0.16547	**0.13567**	0.00000	0.00000	0.00000	**0.24450**	0.11644	**0.45623**	0.23723	0.03024	0.00045	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
BER	**0.20509**	**0.22630**	0.09963	**0.25282**	**0.27878**	0.00000	0.00000	0.07928	**0.19666**	0.06671	0.22274	**0.43233**	0.23997	**0.20284**	**0.17383**	**0.13869**	0.00000	0.00000	0.00000	**0.24852**	**0.11989**	**0.48063**	0.25810	0.03312	0.00230	0.00248	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
SAR	0.17531	0.17765	0.07165	0.20122	0.21385	0.00000	0.00000	0.04930	0.14515	0.03369	0.14149	0.32022	0.16247	0.14616	0.13180	0.12182	0.00000	0.00000	0.00000	**0.22767**	0.10087	0.34752	0.15104	0.01611	-0.01018	-0.01292	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
TUS	0.15619	0.14959	0.05175	0.17199	0.17791	0.00000	0.00000	0.02812	0.11526	0.01086	0.09605	0.25581	0.11927	0.11370	0.10632	0.10768	0.00000	0.00000	0.00000	0.21334	0.08545	0.26667	0.09066	0.00013	-0.02467	-0.02873	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
CAB	0.16782	0.16622	0.06409	0.18923	0.19893	0.00000	0.00000	0.04124	0.13300	0.02495	0.12263	0.29317	0.14452	0.13289	0.12162	0.11668	0.00000	0.00000	0.00000	**0.22218**	0.09522	0.31384	0.12583	0.01046	-0.01509	-0.01841	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
DAI	0.16438	0.16115	0.06052	0.18395	0.19241	0.00000	0.00000	0.03743	0.12760	0.02084	0.11439	0.28144	0.13669	0.12702	0.11702	0.11415	0.00000	0.00000	0.00000	0.21960	0.09246	0.29908	0.11486	0.00760	-0.01768	-0.02123	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
LAH	0.16438	0.16115	0.06052	0.18395	0.19241	0.00000	0.00000	0.03743	0.12760	0.02084	0.11439	0.28144	0.13669	0.12702	0.11702	0.11415	0.00000	0.00000	0.00000	0.21960	0.09246	0.29908	0.11486	0.00760	-0.01768	-0.02123	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
NAX	0.16438	0.16115	0.06052	0.18395	0.19241	0.00000	0.00000	0.03743	0.12760	0.02084	0.11439	0.28144	0.13669	0.12702	0.11702	0.11415	0.00000	0.00000	0.00000	**0.21960**	0.09246	0.29908	0.11486	0.00760	-0.01768	-0.02123	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
DAU	0.16438	0.16115	0.06052	0.18395	0.19241	0.00000	0.00000	0.03743	0.12760	0.02084	0.11439	0.28144	0.13669	0.12702	0.11702	0.11415	0.00000	0.00000	0.00000	0.21960	0.09246	0.29908	0.11486	0.00760	-0.01768	-0.02123	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
DRU	**0.23985**	**0.28368**	**0.13080**	**0.31351**	**0.35334**	0.00000	0.00000	0.11276	**0.25714**	0.10391	0.31374	**0.54174**	**0.32721**	**0.26903**	**0.22327**	**0.15560**	0.00000	0.00000	0.00000	**0.27228**	**0.13946**	**0.59960**	0.37104	0.04849	0.01075	0.01506	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
HEZ	0.16438	0.16115	0.06052	0.18395	0.19241	0.00000	0.00000	0.03743	0.12760	0.02084	0.11439	0.28144	0.13669	0.12702	0.11702	0.11415	0.00000	0.00000	0.00000	0.21960	0.09246	0.29908	0.11486	0.00760	-0.01768	-0.02123	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
ORO	0.16438	0.16115	0.06052	0.18395	0.19241	0.00000	0.00000	0.03743	0.12760	0.02084	0.11439	0.28144	0.13669	0.12702	0.11702	0.11415	0.00000	0.00000	0.00000	0.21960	0.09246	0.29908	0.11486	0.00760	-0.01768	-0.02123	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
TU	0.16438	0.16115	0.06052	0.18395	0.19241	0.00000	0.00000	0.03743	0.12760	0.02084	0.11439	0.28144	0.13669	0.12702	0.11702	0.11415	0.00000	0.00000	0.00000	0.21960	0.09246	0.29908	0.11486	0.00760	-0.01768	-0.02123	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
MIA	0.16438	0.16115	0.06052	0.18395	0.19241	0.00000	0.00000	0.03743	0.12760	0.02084	0.11439	0.28144	0.13669	0.12702	0.11702	0.11415	0.00000	0.00000	0.00000	0.21960	0.09246	0.29908	0.11486	0.00760	-0.01768	-0.02123	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
UYG	0.16438	0.16115	0.06052	0.18395	0.19241	0.00000	0.00000	0.03743	0.12760	0.02084	0.11439	0.28144	0.13669	0.12702	0.11702	0.11415	0.00000	0.00000	0.00000	0.21960	0.09246	0.29908	0.11486	0.00760	-0.01768	-0.02123	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
XIB	0.16438	0.16115	0.06052	0.18395	0.19241	0.00000	0.00000	0.03743	0.12760	0.02084	0.11439	0.28144	0.13669	0.12702	0.11702	0.11415	0.00000	0.00000	0.00000	0.21960	0.09246	0.29908	0.11486	0.00760	-0.01768	-0.02123	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
MON	0.16438	0.16115	0.06052	0.18395	0.19241	0.00000	0.00000	0.03743	0.12760	0.02084	0.11439	0.28144	0.13669	0.12702	0.11702	0.11415	0.00000	0.00000	0.00000	0.21960	0.09246	0.29908	0.11486	0.00760	-0.01768	-0.02123	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
ORC	**0.18178**	0.18789	0.07796	0.21204	0.22744	0.00000	0.00000	0.05605	0.15602	0.04106	0.15863	0.34471	0.17879	0.15809	0.14077	0.12589	0.00000	0.00000	0.00000	**0.23232**	0.10540	0.37754	0.17394	0.02044	-0.00665	-0.00883	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
RUS	**0.20121**	**0.21985**	0.09611	**0.24596**	**0.27018**	0.00000	0.00000	0.07550	**0.18984**	0.06252	0.21208	**0.41832**	0.22978	0.19533	**0.16830**	**0.13670**	0.00000	0.00000	0.00000	**0.24586**	0.11762	**0.46462**	0.24433	0.03123	0.00110	0.00086	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
SHE	0.16438	0.16115	0.06052	0.18395	0.19241	0.00000	0.00000	0.03743	0.12760	0.02084	0.11439	0.28144	0.13669	0.12702	0.11702	0.11415	0.00000	0.00000	0.00000	0.21960	0.09246	0.29908	0.11486	0.00760	-0.01768	-0.02123	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
TUJ	0.16438	0.16115	0.06052	0.18395	0.19241	0.00000	0.00000	0.03743	0.12760	0.02084	0.11439	0.28144	0.13669	0.12702	0.11702	0.11415	0.00000	0.00000	0.00000	0.21960	0.09246	0.29908	0.11486	0.00760	-0.01768	-0.02123	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
YUZ	0.16438	0.16115	0.06052	0.18395	0.19241	0.00000	0.00000	0.03743	0.12760	0.02084	0.11439	0.28144	0.13669	0.12702	0.11702	0.11415	0.00000	0.00000	0.00000	0.21960	0.09246	0.29908	0.11486	0.00760	-0.01768	-0.02123	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
YAK	**0.20121**	**0.21985**	0.09611	**0.24596**	**0.27018**	0.00000	0.00000	0.07550	**0.18984**	0.06252	0.21208	**0.41832**	0.22978	**0.19533**	**0.16830**	**0.13670**	0.00000	0.00000	0.00000	0.24586	0.11762	0.46462	0.24433	0.03123	0.00110	0.00086	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
HAU	-0.00908	-0.01346	0.04234	-0.01227	-0.01535	**0.20658**	0.19268	0.07333	-0.00735	0.09242	0.00091	-0.01761	-0.01058	-0.00699	-0.00328	-0.00313	**0.32261**	**0.27540**	**0.27284**	0.01300	0.00872	-0.03421	0.00000	**0.14687**	**0.22498**	**0.17865**	**0.19268**	**0.22256**	**0.22256**	**0.21631**	**0.21946**	**0.20988**	**0.21946**	**0.21946**	**0.23456**
IBO	0.03463	0.04474	**0.17783**	0.02678	0.02557	**0.33815**	**0.32298**	**0.21359**	0.08176	**0.22989**	0.10963	-0.00812	0.08353	0.08471	0.09104	0.08831	**0.46255**	**0.41302**	**0.41029**	0.00063	**0.12423**	-0.00973	0.11673	**0.30483**	**0.37413**	**0.31660**	**0.32298**	**0.35568**	**0.35568**	**0.34883**	**0.35228**	**0.34177**	**0.35228**	**0.35228**	**0.36884**
LIS	-0.01626	-0.00833	0.14446	-0.02266	-0.02431	**0.50171**	**0.46064**	0.21025	0.02763	0.24736	0.06045	-0.05092	0.02996	0.03114	0.03684	0.02736	**0.72714**	**0.65673**	**0.65224**	-0.03394	0.06443	-0.05747	0.06639	0.36146	**0.53428**	**0.43838**	**0.46064**	**0.54476**	**0.54476**	**0.52849**	**0.53677**	**0.51098**	**0.53677**	**0.53677**	**0.57409**
SAD	**0.11615**	0.11015	0.00765	**0.13732**	0.14288	0.04091	0.03456	-0.01076	0.06682	-0.01911	0.02072	0.22218	0.05536	0.06235	0.05955	0.06281	0.08702	0.06832	0.06733	0.17166	0.04158	0.20476	-0.01297	-0.00955	0.02104	0.01151	0.03456	0.04763	0.04763	0.04506	0.04636	0.04234	0.04636	0.04636	0.05245
ZAR	0.01358	0.00117	-0.00068	0.01353	0.00984	0.13925	0.12799	0.01801	-0.01405	0.03353	-0.02576	0.02889	-0.02343	-0.01629	-0.01272	-0.00459	**0.23322**	**0.19445**	**0.19238**	0.05869	-0.00878	0.00504	-0.03538	0.07007	**0.14159**	0.10905	0.12799	0.15206	0.15206	0.14707	0.14959	0.14191	0.14959	0.14959	**0.16166**
KUW	**0.18178**	0.18789	0.07796	0.21204	**0.22744**	0.00000	0.00000	0.05605	0.15602	0.04106	0.15863	**0.34471**	0.17879	0.15809	0.14077	0.12589	0.00000	0.00000	0.00000	**0.23232**	0.10540	0.37754	0.17394	0.02044	-0.00665	-0.00883	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
TSA	**0.19118**	**0.20983**	0.07046	**0.23978**	**0.26260**	0.00015	-0.00383	0.04010	**0.17018**	0.01513	0.15099	**0.40125**	0.19017	**0.17211**	**0.15017**	**0.12009**	0.02055	0.01328	0.01288	**0.23550**	0.09992	0.41876	0.12577	0.00472	-0.01125	-0.01644	-0.00383	0.00387	0.00387	0.00250	0.00320	0.00098	0.00320	0.00320	0.00629
CHA	-0.00593	-0.00443	0.08259	-0.01044	-0.01270	**0.24579**	**0.23193**	0.11658	0.01415	0.13520	0.03153	-0.02681	0.01363	0.01576	0.02000	0.01687	**0.36158**	**0.31461**	**0.31206**	-0.00284	0.03855	-0.03798	0.03429	**0.19676**	**0.27032**	**0.22080**	**0.23193**	**0.26177**	**0.26177**	**0.25552**	**0.25867**	**0.24909**	**0.25867**	**0.25867**	**0.27379**
BNE	0.01210	-0.00213	-0.02336	0.01240	0.00866	0.18958	0.16513	-0.00837	-0.02337	0.00923	-0.04348	0.03106	-0.03514	-0.02653	-0.02293	-0.01216	0.37836	0.30569	0.30159	0.06214	-0.02140	0.00754	-0.05767	0.04899	0.17052	0.12385	0.16513	0.21746	0.21746	0.20661	0.21208	0.19537	0.21208	0.21208	0.23810
HAN	0.17098	0.17098	0.06731	0.19421	0.20510	0.00000	0.00000	0.04467	0.13806	0.02866	0.13043	0.30435	0.15194	0.13840	0.12588	0.11891	0.00000	0.00000	0.00000	0.22451	0.09766	0.32782	0.13625	0.01293	-0.01290	-0.01599	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
JPT2	**0.21254**	**0.23874**	0.10637	**0.26602**	**0.29523**	0.00000	0.00000	0.08652	**0.20978**	0.07475	0.24306	**0.45832**	**0.25941**	**0.21726**	**0.18449**	**0.14242**	0.00000	0.00000	0.00000	0.25362	**0.12419**	**0.50986**	0.28403	0.03661	0.00440	0.00543	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
MAS2	-0.01268	-0.01932	0.02663	-0.01541	-0.01880	**0.23602**	0.21433	0.06071	-0.01821	0.08361	-0.01365	-0.01684	-0.02276	-0.01841	-0.01473	-0.01282	**0.40605**	**0.34005**	**0.33635**	0.01570	-0.00465	-0.03531	-0.01608	0.14681	**0.25775**	0.19576	**0.21433**	**0.26084**	**0.26084**	**0.25117**	**0.25604**	**0.24116**	**0.25604**	**0.25604**	**0.27928**

Significant differences, after Bonferroni's correction for multiple tests are highlighted in **bold**

Table S10 (continued)

	FRE	BER	SAR	TUS	CAB	DAI	LAH	NAX	DAU	DRU	HEZ	ORO	TU	MIA	UYG	XIB	MON	ORC	RUS	SHE	TUJ	YUZ	YAK	HAU	IBO	LIS	SAD	ZAR	KUW	TSA	CHA	BNE	HAN	JPT2	MAS2
EQG																																			
BPY																																			
UGN																																			
ANG																																			
MOZ																																			
PTG																																			
TUN																																			
MOB																																			
BST																																			
AMH																																			
BSE																																			
CAM																																			
MPY																																			
MAN																																			
BIP																																			
MAS1																																			
CEU																																			
CHB																																			
JPT1																																			
YOR																																			
LUH																																			
BSE																																			
SAN																																			
PAL																																			
BED																																			
BRA																																			
ADY																																			
BAL																																			
MAK																																			
SIN																																			
BUR																																			
HAZ																																			
KAL																																			
PAT																																			
BSQ																																			
FRE	*																																		
BER	0.00000	*																																	
SAR	0.00000	0.00000	*																																
TUS	0.00000	0.00000	0.00000	*																															
CAB	0.00000	0.00000	0.00000	0.00000	*																														
DAI	0.00000	0.00000	0.00000	0.00000	0.00000	*																													
LAH	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	*																												
NAX	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	*																											
DAU	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	*																										
DRU	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	*																									
HEZ	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	*																								
ORO	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	*																							
TU	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	*																						
MIA	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	*																					
UYG	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	*																				
XIB	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	*																			
MON	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	*																		
ORC	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	*																	
RUS	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	*																
SHE	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	*															
TUJ	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	*														
YUZ	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	*													
YAK	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	*												
HAU	**0.21631**	**0.22562**	0.17920	0.15226	0.16827	0.16340	0.16340	0.16340	0.16340	**0.28045**	0.16340	0.16340	0.16340	0.16340	0.16340	0.16340	0.16340	0.18899	0.21946	0.16340	0.16340	0.16340	**0.21946**	*											
IBO	**0.34883**	**0.35904**	**0.30847**	**0.28090**	**0.29698**	**0.29199**	**0.29199**	**0.29199**	**0.29199**	**0.41840**	**0.29199**	**0.29199**	**0.29199**	**0.29199**	**0.29199**	**0.29199**	**0.29199**	**0.31897**	**0.35228**	**0.29199**	**0.29199**	**0.29199**	**0.35228**	0.04174	*										
LIS	**0.52849**	**0.55247**	0.41826	0.33333	0.38311	0.36759	0.36759	0.36759	0.36759	**0.66538**	0.36759	0.36759	0.36759	0.36759	0.36759	0.36759	0.36759	**0.44923**	**0.53677**	0.36759	0.36759	0.36759	**0.53677**	-0.01083	-0.03197	*									
SAD	0.04506	0.04887	0.02765	0.01003	0.02124	0.01808	0.01808	0.01808	0.01808	0.07028	0.01808	0.01808	0.01808	0.01808	0.01808	0.01808	0.01808	0.03276	0.04636	0.01808	0.01808	0.01808	0.04636	0.11360	**0.26466**	0.29445	*								
ZAR	**0.14707**	**0.15451**	0.11682	0.09318	0.10749	0.10323	0.10323	0.10323	0.10323	**0.19855**	0.10323	0.10323	0.10323	0.10323	0.10323	0.10323	0.10323	0.12496	0.14959	0.10323	0.10323	0.10323	0.14959	0.00376	0.10835	0.05581	0.04589	*							
KUW	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.18899	**0.31897**	**0.44923**	0.03276	0.12496	*						
TSA	0.00250	0.00451	-0.00872	-0.02351	-0.01377	-0.01641	-0.01641	-0.01641	-0.01641	0.01408	-0.01641	-0.01641	-0.01641	-0.01641	-0.01641	-0.01641	-0.01641	-0.00505	0.00320	-0.01641	-0.01641	-0.01641	0.00320	**0.21023**	**0.35805**	**0.50559**	0.01583	**0.12975**	-0.00505	*					
CHA	**0.25552**	**0.26483**	**0.21854**	0.19216	0.20776	0.20299	0.20299	0.20299	0.20299	**0.31964**	0.20299	0.20299	0.20299	0.20299	0.20299	0.20299	0.20299	**0.22825**	**0.25867**	0.20299	0.20299	0.20299	**0.25867**	-0.00534	0.01252	-0.03158	**0.16099**	0.03113	**0.22825**	**0.25535**	*				
BNE	0.20661	0.22274	0.14149	0.09605	0.12263	0.11439	0.11439	0.11439	0.11439	0.31374	0.11439	0.11439	0.11439	0.11439	0.11439	0.11439	0.11439	0.15863	0.21208	0.11439	0.11439	0.11439	0.21208	0.00091	0.10963	0.06045	0.02072	-0.02576	0.15863	0.15099	0.03153	*			
HAN	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.17282	0.30173	0.39774	0.02401	0.11141	0.00000	-0.01153	**0.21224**	0.13043	*		
JPT2	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	**0.23748**	**0.37202**	**0.58083**	**0.05360**	**0.16398**	0.00000	0.00684	**0.27670**	0.24306	0.00000	*	
MAS2	**0.25117**	**0.26556**	0.19343	0.15344	0.17680	0.16955	0.16955	0.16955	0.16955	**0.34734**	0.16955	0.16955	0.16955	0.16955	0.16955	0.16955	0.16955	0.20857	**0.25604**	0.16955	0.16955	0.16955	**0.25604**	-0.01830	0.04770	-0.00629	0.10691	-0.00915	0.20857	**0.23764**	-0.00580	-0.01365	0.18367	0.28372	*

Significant differences, after Bonferroni’s correction for multiple tests are highlighted in **bold**

Table S11. Pairwise differences between the six populations addressed in this study and other comparisons populations, based on haplotype frequencies of P49A, A262V and V296I variants from *TAS2R38* gene.

	EQG	**BPY**	**UGN**	**ANG**	**MOZ**	**PTG**	**AWC**	**NWC**	**NGC**	**FWC**	**PWC**	**AEA**	**NEA**	**NGA**	**LEA**	**HAD**	**PAL**	**BRA**	**EAS**	**EUR1**	**EUR2**	**AFR**	**PCA**
EQG	*																						
BPY	**0.13414**	*																					
UGN	**0.04090**	0.04379	*																				
ANG	0.03308	0.05885	-0.00460	*																			
MOZ	**0.05631**	**0.10022**	0.01495	0.04082	*																		
PTG	**0.07915**	**0.25799**	**0.08935**	0.07808	**0.10718**	*																	
AWC	**0.04498**	**0.10779**	0.00880	0.02392	0.00076	0.05364	*																
NWC	**0.04590**	**0.14183**	0.02251	0.02712	0.02278	0.02155	0.00129	*															
NGC	0.03936	**0.10007**	0.00233	0.00118	0.03408	0.03051	0.00297	-0.00056	*														
FWC	**0.05058**	0.04980	-0.00583	0.00152	0.02510	**0.08255**	0.01013	0.02173	-0.00332	*													
PWC	**0.05038**	**0.16260**	0.03200	0.04536	0.01411	0.02714	-0.00154	-0.00286	0.00991	0.03148	*												
AEA	**0.05368**	**0.10270**	0.01316	0.03627	-0.01001	**0.08561**	-0.00308	0.01402	0.02162	0.01863	0.00702	*											
NEA	**0.06266**	**0.19749**	**0.05394**	**0.06495**	0.03036	0.02045	0.01140	0.00352	0.02446	**0.05387**	-0.00459	0.02221	*										
NGA	**0.08188**	**0.25999**	**0.09174**	**0.10516**	0.04867	0.03513	0.03400	0.02313	0.06500	**0.09545**	0.00707	0.04169	0.00010	*									
LEA	**0.07290**	**0.19972**	0.06190	**0.08657**	0.00390	0.08434	0.01539	0.02458	0.06174	0.07090	0.00549	0.00787	0.00877	-0.00133	*								
HAD	0.04909	**0.16254**	0.03103	0.04804	0.00008	0.04047	-0.00540	-0.00186	0.01775	0.03429	-0.01374	-0.00291	-0.00783	-0.00210	-0.01248	*							
PAL	**0.06669**	**0.22628**	0.06724	0.06228	0.08312	-0.01940	0.03358	0.00888	0.01216	0.05853	0.01127	0.06277	0.01005	0.02919	0.06775	0.02336	*						
BRA	0.06320	**0.25201**	0.07436	0.06975	0.07666	-0.03743	0.02930	0.00150	0.01797	0.06930	-0.00223	0.05735	-0.00921	-0.00541	0.04441	0.00657	-0.03926	*					
EAS	**0.09506**	**0.28070**	**0.10923**	**0.10924**	**0.10936**	-0.00974	**0.05821**	0.02877	0.04773	**0.09913**	0.02701	**0.08720**	0.01971	0.02571	0.07381	0.03764	-0.01199	-0.03575	*				
EUR1	**0.09404**	**0.26838**	**0.10467**	**0.09012**	**0.13197**	-0.01294	**0.07198**	0.03629	0.03943	**0.09426**	0.04691	**0.10767**	**0.04065**	0.06145	**0.11258**	0.06418	-0.01028	-0.02281	0.00309	*			
EUR2	**0.06738**	**0.24220**	**0.08592**	**0.06870**	**0.11056**	-0.01086	**0.05917**	0.02740	0.02954	**0.08027**	0.03724	**0.09109**	0.03274	0.05104	**0.09392**	0.05058	-0.00857	-0.02075	0.00600	-0.00553	*		
AFR	**0.05783**	**0.15964**	0.04103	0.06237	-0.00623	0.09032	0.00893	0.02173	0.04996	0.05310	0.00878	-0.00020	0.01565	0.01673	-0.01450	-0.00812	0.07352	0.05612	**0.08832**	**0.11784**	**0.09633**	*	
PCA	**0.09562**	0.00596	0.01584	0.02239	**0.07549**	**0.17057**	**0.06543**	**0.08754**	0.04183	0.01405	**0.10573**	**0.07116**	**0.13790**	**0.19158**	**0.15239**	**0.10990**	**0.14167**	**0.16468**	**0.19401**	**0.17912**	**0.16127**	**0.12340**	*

Significant differences, after Bonferroni's correction for multiple tests are highlighted in **bold**

3.3 “Genetic diversity in genes belonging to *TAS1R* and *TAS2R* families of taste receptors: looking for their evolutionary trajectories”

3.3.1 Introduction

Several lines of evidence support that the sense of taste exerts a key role in guiding food preferences, dietary habits and many aspects of human health' balance. Like in other mammals, humans are able to differentiate five taste qualities: sweet, *umami*, sour, salty, and bitter, which are commonly referred to as the basic tastes. It has been argued that despite seeming to be a limited repertoire, it is satisfactory to accommodate the evolutionary need to help recognize and distinguish key dietary components (Chandrashekar, et al. 2006). In reality, the perception of tastes is thought to be of evolutionary significance because it might account to avoid potential dietary threats that impact on nutritional or physiological requirements. Thus, bitter taste, which is generally aversive, is assumed to protect against toxic and harmful foods; sweet taste, contrarily associated with a pleasant sensory perception, can signal the presence in food of carbohydrates that are an important energy source; sour, assumedly also helps to avoid the intake of spoiled food and potentially noxious or poisonous chemicals; salt taste might govern the intake of salts that are essential for the body's electrolyte balance; and *umami*, whose prototypical stimulus is the amino acid glutamate, is supposed to assist in the evaluation of a food's protein content (Garcia-Bailo, et al. 2009).

Among these five taste sensations, sweet, bitter and *umami* are considered the most important for food acceptance (Temussi 2009). Similarly to what happens with other taste modalities, their recognition involves complex physiological mechanisms differing from modality to modality, about which many still remains very vague. It is known, however, that the earliest stages of bitter, *umami* and sweet transduction each depends on the activation of different taste receptors, all of them belonging to two classes of the superfamily of seven-transmembrane G protein coupled receptors (GPCRs), known as TAS1R and TAS2R, which are two family of proteins both expressed in subpopulations of taste buds cells. Compounds that elicit bitter taste are recognized by TAS2R receptors that contain 25 members in humans, of which one of the most widely studied is TAS2R38, encoded by *TAS2R38*, the first bitter-taste receptor gene to be identified (Kim, et al. 2003). This receptor was demonstrated to respond to phenylthiocarbamide (PTC) and 6-n-propylthiouracil (PROP), two compounds associated with a bitter taste perception, which represents a highly variable trait in humans as began to be documented in the early 1930s (reviewed in Wooding (2005)). A substantial fraction of variability in PTC/PROP sensibility was attributed to genetic variability at *TAS2R38*, located on chromosome 7q36, where three non-synonymous polymorphisms were identified (C145G, C785T and G886A according to nucleotide position, or p.P49A, p.A262V and p.V296I, according to residue substitution) that define different haplotypes from which the most common in humans are PAV and AVI (by order of the first letter of the amino acid present at positions 49, 262, and 296) associated with "taster" (sensitive) and "non-taster" (insensitive) phenotypes, respectively (Campbell, et al. 2012).

Another well-known bitter receptor is TAS2R16, which mediates the response to β-glucopyranosides compounds such as salicin present in the willow bark (Bufe, et al. 2002; Campbell, et al. 2013). The encoding gene, *TAS2R16,* also located on chromosome 7, contains a common polymorphic variation, G516T, that contributes to the differential response to salicin and other naturally occurring bitter glycosides, having been shown that the derived T allele confers greater sensibility to several bitter ligands (Soranzo, et al. 2005).

As for the *umami* and sweet tastes, the receptors up till now discovered belong to the TAS1R family of GPCRs, a small family that in humans (and many other mammals) includes three proteins: TAS1R1, TAS1R2 and TAS1R3, whose genes are located in a single cluster on chromosome 1p36 with order *TAS1R2*, *TAS1R1* and *TAS1R3,* in a telomere to centromere direction. The three TAS1R proteins function as heterodimers, such as the dimer TA1R1+TAS1R3, which acts as the main receptor for *umami*, and TAS1R2+TAS1R3 that is the receptor essential to sense sugars responding to a broad variety of natural and artificial ligands (Li, et al. 2002; Nelson, et al. 2002). As follows, TAS1R3 is a common subunit shared by sweet and *umami* taste receptors.

Although humans also differ widely in the perception of sweet and *umami* tastes, the genetic factors that account for interindividual variability in sweet and *umami* responsiveness still remain largely undeciphered (Bachmanov and Beauchamp 2007). Still, recent studies have pinpointed a few *TAS1R* genetic variants that partially underlie such diversity. Two Single Nucleotide Polymorphism (SNPs) located in the *TAS1R3* promoter, -T1572C and -T1266C, shown to influence *TAS1R3* promoter activity *in vitro,* were further associated with taste sensitivity to sucrose in humans, explaining together 16% of population variability in sweet perception, with both derived T alleles accounting to diminished sensibility (Fushan, et al. 2009). In *TAS1R2,* which encodes the other subunit of the sweet taste receptor, a candidate variant was also identified that might influence the responsiveness to sweet *stimuli,* namely G571A (rs35874116), a non-synonymous substitution found to be associated with habitual consumption of sugars (Eny, et al. 2010). In respect to *umami*, both the coding regions of *TAS1R1* and *TAS1R3* contain variants that were correlated with *in vitro* dose-response to *umami* substances and/or with the recognition perception thresholds to those substances, specifically C329T and G1114A in *TAS1R1*, and G13A and C2269T in *TAS1R3* (Raliou, et al. 2009; Shigemura, et al. 2009). For all the five mentioned *TAS1R/2R* genes, extensive genetic diversity across human populations has been documented, including the identification of several variants with phenotypic effects in taste perception that so might result in inter-population differences in response to food resources. This raises the question of the extent to which their current patterns of genetic distribution were shaped by dietary changes, or other selective pressure, that humans experienced throughout their history. In terms of dietary changes, the most recent major shift in humans was initiated with the origin and spread of agriculture, which first arose in the Near East around 10 Kya ago, having later independently emerged in other regions of the world. The adoption of farming economies was accompanied by the introduction of new foods in diets, or at least increased quantities of foods, and by a general reduction in dietary diversity (Luca, et al. 2010). This represented a dramatic dietary transition to which post-Neolithic humans reacted making adjustments of different nature, some implying genetic adaptations. Previous studies have already provided compelling evidence on genes whose patterns of diversity were shaped by selective pressures associated with the transition to the new subsistence strategies (agriculture and pastoralism), with the most emblematic cases involving the digestive enzymes lactase (*LCT*) (Tishkoff, et al. 2007) and amylase (*AMY*) (Perry, et al. 2007). But in addition, a number of genes implicated in food metabolism were identified with signatures indicating that they likewise might represent genetic adaptations to dietary specializations in humans (Hancock, et al. 2010; Valente, et al. 2015).

Genes coding for taste receptors are another kind of diet-related genes that are also thought to have played a role in human adaptations to dietary changes. A few previous studies did already brought to light coherent signatures of selection at genes encoding the bitter receptors TAS2R38 (Wooding, et al. 2004; Calo, et al. 2011; Campbell, et al. 2012) and TAS2R16 (Soranzo, et al. 2005; Campbell, et al. 2013). Also regarding the three genes from the *TAS1R* family, intriguing departures from the neutral expectations were reported by Kim, et al. (2006), after performing the first worldwide survey of molecular diversity at those *loci*.

Taking advantage on the unprecedented wealth of genomic data turned available by The 1000 Genomes Project Consortium (www.1000genomes.org), the principal aim of this work was to interrogate the patterns of diversity at *TAS1R1*, *TAS1R2*, *TAS1R3, TAS2R16* and *TAS2R38* in worldwide populations, to obtain renewed insights on their evolution and on whether they were targets of selection in the recent human history.

3.3.2 Material and Methods

Samples and sequence databases

The data used in this work encompassed the following genomic sequences from five genes taste receptors: *TAS1R1*-chr1: 6,615,241-6,639,817; *TAS1R2*-chr1: 19,186,176-19,166,093; *TAS1R3*-chr1: 1,265,000-1,270,686, including the promoter; *TAS2R16*-chr7: 122,634,759-122,635,754; and *TAS2R38*-chr7: 141,671,556-141,674,413, including the 5' and 3' flanking regions. All sequences were retrieved from the panel of The 1000 Genomes Project (www.1000genomes.org). Out of the populations contained in that panel, 10 populations were considered belonging to three major human groups, as below discriminated: African – LUH (Luhya from Kenya) and YOR (Yoruba from Nigeria); European – CEU (Utah residents with Northern and Western European ancestry), IBS (Iberian from Spain), TUS (Tuscan in Italia), GBR (British in England and Scotland) and FIN (Finnish in Finland); Asian – CHB (Chinese from Beijing), CHS (Southern Han Chinese) and JPT (Japanese from Tokyo).

The ancestral allele for each human variant was determined by comparison with the chimpanzee sequence obtained in the PanMap Project website (http://panmap.uchicago.edu/).

Inspection of the Neanderthal and Denisovan sequences was done with the Neanderthal Genome browser (http://neandertal.ensemblgenomes.org/index.html) and the UCSC browser (http://genome.ucsc.edu/), respectively.

Statistical analyses

Based on the unphased genotypic data retrieved from 1000 Genomes Project, haplotypes were reconstructed for each genomic sequence running the SHAPEIT software (Delaneau, et al. 2012), and using as option for the output files the variant call format (vcf). Next, the SPIDER software (Lischer and Excoffier 2012) was used to convert the .vcf files that contained the SHAPEIT haplotypes into .xml files, ignoring Insertions-Deletions (INDELs).

To identify signatures of selection, several statistics based on the allele frequency spectrum were calculated. The DivSat program (Soares, et al. 2015) was used to estimate the number of segregating sites (*S*) and the nucleotide diversity (π), which is based on the average number of pairwise differences between sequences (Nei and Li 1979), and Tajima's *D* for each population. To address Tajima's *D* significance, coalescent simulations were produced using the "ms" program (Hudson 2002), assuming best-fit demographic models according to Gravel, et al. (2011) for YOR, CEU, GBR, FIN CHB, CHS and JPT and according to Voight, et al. (2005) for TUS samples. Viewing that estimates were obtained for the population recombination parameter Rho (ρ), which combines information on effective population size (N_e) and recombination rate (r) as summarized in the equation $\rho=4N_e r$, where the specific value of r for each gene was obtained in HapMap II (McVean, et al. 2004) and the initial sample size considered for all populations was N_e=7,300 individuals, according to Gravel, et al. (2011).

Besides, the constant demographic model was also simulated using the DNAsp software (Rozas 2009). Whatever the demographical model, the null distributions of Tajima's *D* values were obtained to then calculate their 5th (for negative Tajima's *D* values) or 95th (for positive Tajima's *D* values) percentiles. Relationships between haplotypes were assessed through the median-joining algorithm implemented in the NETWORK v.4.6.1.0 program (Bandelt, et al. 1999).

The tagged variations that appear in the different networks are discriminated in Table 16.

Table 16. Polymorphisms addressed in this section.

Gene	SNPs	Mutation	Taste
TAS1R1	C329T (rs41278020)	C/T	*umami*
	G1114A (rs34160967)	G/A	
TAS1R3	G13A (rs76755863)	G/A	*umami*
	C2269T (rs307377)	C/T	
	-T1572C (rs307355)	T/C	sweet
	-T1266C (rs35744813)	T/C	
TAS2R16	G516T (rs846664)	G/T	bitter (salicin)
TAS2R38	C145G (rs713598)	C/G	bitter (PTC/PROP)
	C785T (rs1726866)	C/T	
	G886A (rs10246939)	G/A	
TAS1R2	G571A (rs35874116)	G/A	sweet

Estimation of Time to Most Recent Common Ancestor (TMRCA) as well of the neutral parameter theta (θ) was done using the coalescent method implemented in GENETREE v.9.0 software (Griffiths and Tavare 1994). To take into account the effect of selection at sites in the sequences, the selection parameter beta (β) was also estimated (Coop and Griffiths 2004). Since the used coalescent method does not assume recombination, some incompatibilities appeared due to rare recombinant haplotypes, which were then removed from the analysis. The mutation rate per generation per basepair μ=(D_{xy}/2n) L, was estimated considering D_{xy} as the differences in sequence between human and chimpanzee for each gene, calculated with the DnaSP software (Rozas 2009), 2n as the number of generations elapsed since the divergence human/chimpanzee assuming that it has occurred at 5.4 Mya (Patterson, et al. 2006) and an generation time of 25 years, and L as the length of the genomic sequence of each gene.

Given that the theta parameter computed by the GENETREE is the population mutation rate given by $\theta=4N_e\mu$ (Watterson 1975), it was possible from the θ values to derive time scaled in $2N_e$ generations, converting coalescent units into years ($2N_et$). The number of sequences used for each population was the same used in NETWORK program, for each gene.

F_{ST} measures of genetic distances were calculated in ARLEQUIN software ver. 3.5.1 (Excoffier and Lischer 2010).

3.3.3 Results and Discussion

Nucleotide diversity

In this study we have examined three genes from the *TAS1R* family, *TAS1R1, TAS1R2* and *TAS1R3*, all containing five introns, and two genes from the *TAS2R* family, *TAS2R16* and *TAS2R38*, which are both intronless. Because the evolution rates of intronless and intron-containing mammalian genes might differ (Shabalina, et al. 2010; Yu, et al. 2014), comparisons of sequence summary statistics between members of the two families must be interpreted with caution, reason that lead us to present the results obtained discriminated by *TAS* family.

Concerning the genes from the *TAS1R* family, all the three members exhibited substantially higher levels of nucleotide diversity in African populations compared to non-African ones (see π values in Tables 17, 18 and 19). Indeed, both the Luhya and the Yoruba, the two African groups integrated in this study, presented much more sequence differences than any of the different Eurasian groups, whichever the *TASR1* gene considered. This finding was absolutely not surprising, since it is now very well documented that at a worldwide scale Sub-Saharan African populations have the highest levels of polymorphism across de genome (Elhassan, et al. 2014). Sub-Saharan Africa has consistently revealed to be the most genetically diverse region in the world, reflecting the important role of the region in the demographic history of human populations (Gibbons 2011). In Africa, π values ranged from $9.7\text{x}10^{-4}$, observed at *TAS1R3* in the Luhya, to $19.1\text{x}10^{-4}$ at *TAS1R2* in the Yoruba. Among the non-African populations, *TAS1R3* also showed the lowest diversity, $2.3\text{x}10^{-4}$ registered in the sample from Iberia (IBS), while *TAS1R2* presented again the highest value, $16.4\ \text{x}10^{-4}$ found also in the same population. Globally, among *TAS1R* genes, *TAS1R2* is notably the more diverse, whereas within the two remaining sequences, *TAS1R3* is the less polymorphic*.* This means that *TAS1R1* and *TAS1R3*, which encode two proteins acting as dimers to form an *umami* receptor, display less nucleotide diversity than *TAS1R2,* coding for the protein that dimerizes with *TAS1R3* to form a sweet receptor. The observation that the less variable gene was *TAS1R3,* which encodes a structural subunit of two different taste receptors, seems to indicate that this double role might increase the evolutionary constrains at the locus.

Overall, our data on sequence diversity at the *TAS1R* family are totally consistent with those previously obtained in the study of Kim et al. 2006 that was based on a much more restricted set of populations with low sample sizes.

In addition to Kim, et al. (2006), we assessed the diversity of *TAS1R3* promoter region since it has been associated with interindividual differences in sweet perception (Fushan, et al. 2009). Levels of sequence heterozygosity at the promoter were substantially higher then observed in the corresponding genic region (Table 19), but differences in diversity between African and non-African populations were of the same order of magnitude in both regions of *TAS1R3.*

Table 17. Nucleotide diversity and neutrality tests for *TAS1R1* gene.

	Population	N	*S*	π ($x10^{-4}$)	Tajima's *D*	5% limit	
						Constant	Best fit
AFRICA	YOR	176	216	12.92205	-0.49807	-1.54094	-1.25119
	LUH	194	243	13.47935	-0.64871	-1.55699	-
EUROPE	CEU	170	82	3.65724	-1.16397	-1.54856	-2.22599
	FIN	186	83	3.71923	-1.11615	-1.55441	-2.22405
	GBR	178	90	4.60367	-0.86062	-1.60491	-2.23311
	IBS	28	61	4.70781	-0.99728	-1.73242	-
	TUS	196	124	5.26590	-1.21741	-1.56105	-1.46852
ASIA	CHB	194	140	6.88424	-0.92418	-1.51554	-2.48828
	CHS	200	142	6.70253	-1.00006	-1.50223	-2.49010
	JPT	178	138	5.25126	-1.46026*	-1.37787	-2.48846

Pop= populations studied.
N= number of chromosomes.
S= number of segregating sites.
π= nucleotide diversity.
Tajima's *D*= values observed of Tajima's *D.*
5% limit= value obtained for the limit of the distribution for Tajima's *D* according to i) constant (Rozas 2009) and ii) best fit model (Voight, et al. 2005; Gravel, et al. 2011).
* values significant for constant model.
**values significant for best fit model.

Table 18. Nucleotide diversity and neutrality tests for *TAS1R2* gene.

	Population	N	*S*	π (x10^{-4})	Tajima's *D*	5% limit	
						Constant	Best fit
AFRICA	**YOR**	176	213	19.10777	0.22388	1.47825	1.25487
	LUH	194	225	18.30847	-0.13553	-1.38089	-
EUROPE	**CEU**	170	140	16.02493	1.08721	1.92769	2.65229
	FIN	186	132	15.37278	1.13890	1.79659	2.62989
	GBR	178	125	16.16456	1.57121*	1.51307	2.67731
	IBS	28	133	16.39161	1.41566*	1.40779	-
	TUS	196	155	15.30505	0.55386	1.55161	1.41307
ASIA	**CHB**	194	137	13.50346	0.52322	1.88733	2.60954
	CHS	200	151	13.64909	0.16442	1.92975	2.61446
	JPT	178	112	12.61982	1.04420	1.96983	2.60980

Pop= populations studied.
N= number of chromosomes.
S= number of segregating sites.
π= nucleotide diversity.
Tajima’s *D*= values observed of Tajima’s *D.*
5% limit= value obtained for the limit of the distribution for Tajima’s *D* according to i) constant (Rozas 2009) and ii) best fit model (Voight, et al. 2005; Gravel, et al. 2011).
* values significant for constant model.
**values significant for best fit model.

Table 19. Nucleotide diversity and neutrality tests for *TAS1R3* gene.

	Population	**region**	***N***	**S**	**π (x10^{-4})**	**Tajima's *D***	**5% limit**	
							Constant	**Best fit**
AFRICA	**YOR**	**promoter**	176	25	16.35743	-0.73586	-1.43241	-1.41304
		genic	176	39	10.02071	-1.09341	-1.35337	-1.37479
		Total (5698 bp)	176	57	11.69465	-1.01861	-1.39237	-1.33867
	LUH	**promoter**	194	26	15.52824	-0.69605	-1.40471	-
		genic	194	34	9.71043	-1.21017	-1.4023	
		Total (5698 bp)	194	58	11.23905	-1.09025	-1.61388	
EUROPE	**CEU**	**promoter**	170	15	5.36745	-1.55059*	-1.45288	-1.97949
		genic	170	15	2.65058	-2.10569**	-1.4488	-1.99024
		Total (5698 bp)	170	37	3.39485	-2.07103*	-1.39528	-2.15002
	FIN	**promoter**	186	14	6.34259	-1.22504	-1.41195	-1.96576
		genic	186	13	3.28416	-1.69170*	-1.34578	-1.94559
		Total (5698 bp)	186	32	4.11715	-1.66284*	-1.34629	-2.12233
	GBR	**promoter**	178	16	6.47084	-1.33438	-1.45053	-1.99861
		genic	178	16	3.14561	-2.04572**	-1.4311	-1.99155
		Total (5698 bp)	178	39	4.05812	-1.94805*	-1.41695	-2.14471
	IBS	**promoter**	28	6	7.40352	-0.98519	-1.51358	-
		genic	28	5	2.31896	-1.91860*	-1.58814	
		Total (5698 bp)	28	16	3.75448	-1.67141*	-1.46974	
	TUS	**promoter**	196	17	4.90216	-1.82615**	-1.42149	-1.56321
		genic	196	22	2.70924	-2.27622**	-1.43049	-1.53598
		Total (5698 bp)	196	47	3.30078	-2.27690**	-1.4163	-1.49878
ASIA	**CHB**	**promoter**	194	17	6.89435	-1.60220*	1.43821	-2.24606
		genic	194	18	3.52234	-1.83274*	-1.45201	-2.25905
		Total (5698 bp)	194	41	4.44233	-1.88537*	-1.41764	-2.41342
	CHS	**promoter**	200	11	6.77792	-1.23707	-1.61632	-2.09584
		genic	200	17	3.99377	-1.64262*	-1.4334	-2.24106
		Total (5698 bp)	200	36	4.73585	-1.62953*	-1.35711	-2.39572
	JPT	**promoter**	178	9	7.14257	-1.08627	-1.4584	-2.02927
		genic	178	18	4.39720	-1.57488*	-1.40219	-2.26781
		Total (5698 bp)	178	35	5.12153	-1.52831*	-1.43198	-2.38990

Pop= populations studied.
N= number of chromosomes.
S= number of segregating sites.
π= nucleotide diversity.
Tajima's *D*= values observed of Tajima's *D*.
5% limit= value obtained for the limit of the distribution for Tajima's *D* according to i) constant (Rozas 2009) and ii) best fit model (Voight, et al. 2005; Gravel, et al. 2011).
* values significant for constant model.
**values significant for best fit model.

In regard to *TAS2R* family, sequence diversity statistics revealed that both genes were characterized by high levels of nucleotide diversity (π values in Table 20 for *TAS2R16* and Table 21 for *TAS2R38*).

Table 20. Nucleotide diversity and neutrality tests for *TAS2R16* gene.

	Population	N	*S*	π (x10^{-4})	Tajima's *D*	5% limit	
						Constant	Best fit
AFRICA	YOR	176	6	22.56477	2.37542**	1.65474	1.64341
	LUH	194	11	18.81798	0.15093	1.65452	-
EUROPE	CEU	170	3	4.68545	-0.18500	-1.38915	-1.45965
	FIN	186	2	5.79766	0.92933	1.82433	2.37902
	GBR	178	2	5.10264	0.64139	2.10933	2.38828
	IBS	28	2	5.26442	0.04015	1.76321	-
	TUS	196	3	4.73482	-0.13001	-1.38553	-1.41655
ASIA	CHB	194	2	5.14994	0.68166	1.82581	2.23145
	CHS	200	3	5.24330	0.03419	1.69522	1.84797
	JPT	178	1	4.97632	1.90786**	1.86867	1.88977

Pop= populations studied.
N= number of chromosomes.
S= number of segregating sites.
Tajima's *D*= values observed of Tajima's *D.*
5% limit= value obtained for the limit of the distribution for Tajima's *D* according to i) constant (Rozas 2009) and ii) best fit model (Voight, et al. 2005; Gravel, et al. 2011).
* values significant for constant model.
**values significant for best fit model.

Table 21. Nucleotide Diversity and neutrality tests for *TAS2R38* genic region, non-coding 3' and 5' flanking regions and the complete region.

	Population	region	*N*	S	π (x10^{-4})	Tajima's *D*	5% limit	
							Constant	Best fit
AFRICA	YOR	genic	176	8	13.30201	0.20363	1.61421	1.67722
		5' non-coding	176	9	5.02030	-1.61103**	-1.49672	-1.43162
		3' non-coding	176	4	2.59881	-0.84944	-1.43814	-1.43611
		Total (2957 bp)	176	21	7.92680	-1.04424	-1.39657	-1.39920
	LUH	genic	194	7	13.49852	0.61144	1.64259	-
		5' non-coding	194	9	5.09054	-1.51004*	-1.44206	
		3' non-coding	194	4	3.25078	-0.68721	-1.44312	
		Total (2957 bp)	194	20	8.37777	-0.79934	-1.45688	
EUROPE	CEU	genic	170	3	12.98764	2.99380**	1.73938	2.69515

		5' non-coding	170	2	0.94971	-0.44057	1.26026	-1.30676
		3' non-coding	170	2	0.11223	-0.95317	-1.26026	-1.30676
		Total (2957 bp)	170	7	5.48923	1.61119*	1.50423	2.61132
	FIN	**genic**	186	5	13.19618	1.44942	1.72684	2.03908
		5' non-coding	186	2	0.40406	-0.77473	-1.25217	-1.29477
		3' non-coding	186	3	0.34601	-1.24926	-1.41887	-1.45219
		Total (2957 bp)	186	10	5.43142	0.62558	1.71352	1.86845
	GBR	**genic**	178	3	12.82051	2.95022**	1.71562	2.09448
		5' non-coding	178	2	1.02883	-0.40743	-1.25616	-1.25616
		3' non-coding	178	2	0.24285	-0.89761	-1.25616	-1.30062
		Total (2957 bp)	178	7	5.55298	1.61955	1.63302	1.94019
	IBS	**genic**	28	3	12.50938	2.04779*	1.55354	-
		5' non-coding	28	4	113.771	-1.24015	-1.56592	
		3' non-coding	28	3	0.44221	-1.20578	-1,5273	
		Total (2957 bp)	28	10	5.87168	-0.03448	-1.47149	
	TUS	**genic**	196	3	13.12752	3.09348**	1.67223	1.68716
		5' non-coding	196	5	1.24152	-1.43237	-1.44656	-1.45021
		3' non-coding	196	4	0.41439	-1.46945*	-1.41858	-1.47076
		Total (2957 bp)	196	12	5.80214	-0.03518	-1.45565	-1.50114
ASIA	**CHB**	**genic**	194	5	11.88248	1.12786	1.62685	2.52282
		5' non-coding	194	2	0.09565	-0.94900	-1.2483	-1.28919
		3' non-coding	194	NA	NA	NA	NA	NA
		Total (2957 bp)	194	7	4.85383	0.73682	1.77051	2.31884
	CHS	**genic**	200	4	11.83501	1.75133*	1.56888	2.31581
		5' non-coding	200	2	0.10594	-0.95981	-1.24547	-1.28516
		3' non-coding	200	2	0.11947	-0.95981	-1.24547	-1.28516
		Total (2957 bp)	200	8	4.76008	0.64529	1.63471	2.32458
	JPT	**genic**	178	5	13.86165	1.60302	1.72952	2.52394
		5' non-coding	178	NA	NA	NA	NA	NA
		3' non-coding	178	4	0.32051	-1.50263*	-1.43514	-1.66321
		Total (2957 bp)	178	9	5.66697	0.43122	1.73881	2.41240

Pop= populations studied.
N= number of chromosomes.
S= number of segregating sites.
π= nucleotide diversity.
Tajima's *D*= values observed of Tajima's *D.*
5% limit= value obtained for the limit of the distribution for Tajima's *D* according to i) constant (Rozas 2009) and ii) best fit model (Voight, et al. 2005; Gravel, et al. 2011).

* values significant for constant model.
** values significant for best fit model.

The π values for African and non-African populations fit remarkably well those previously obtained for *TAS2R16* (Campbell, et al. 2013) or for *TAS2R38* (Campbell, et al. 2012), safeguarding minor differences likely explained by handling with distinct sample sets. Such agreement, besides adding strength to the overall pattern of sequence diversity at *TAS2R16* and *TAS2R38*, also indicates that our data mining approach based on sequences from the 1000 Genomes Project (which offers high coverage for coding regions across the all genome) does not appear to provide biased inferences comparatively to those obtained from resequencing data, as was the case of the two studies of Campbell and collaborators. This was further corroborated when for *TAS2R38* we accessed to sequence diversity for its two noncoding flanking regions, as had been previously done for the same gene by Campbell, et al. (2012) (see Table 21). Levels of diversity obtained were again in agreement with the range reported by Campbell, et al. (2012), and similarly we also detected lower diversity in 3′ noncoding region comparatively to the 5′ noncoding region of *TAS2R38*. Noteworthy, comparing the coding regions of two *TAS2R* genes, *TAS2R16* was strikingly more diverse in Africans (the π value in the Yoruba reached to 22.6x10^{-4}, which was the maximal value of sequence diversity across the five genes here investigated), while in Europeans or Asians it was nearly four times less diverse. Contrarily, *TAS2R38* presented levels of diversity very similar across all populations, not showing a well-defined frontier between African and non-African populations, which remarkably was not characteristic of the two flanking region of the gene, where it was manifest the low diversity in Eurasians. A pattern resembling that registered in the genic region of *TAS2R38*, was further detected in only one gene of the other *TASR* family, namely in *TAS1R2*, where differences in diversity between African and non-African sequences were also much smaller comparatively to the usually observed for the other genes of the family.

Neutrality tests

Among the neutrality tests based on the allele frequency spectra, Tajima’s *D* is a statistic that measures the departures from a neutral model of evolution in an equilibrium population of constant size (Tajima 1989), representing one of the most widely employed tests in population genetics to detect signatures of selection (Wollstein and Stephan 2015).

Because Tajima's *D* is highly influenced by population demographic history, in general Tajima's *D* values tend to be slightly negative in African populations and marginally positive in non-African ones (illustrated, for instance, in the SeattleSNPs Variation Discovery Resource, available at http://pga.gs.washington.edu/, where average Tajima's *D* values across 327 genes are -0.46 and 0.24 for African-Americans and Europeans, respectively). So, to assess the statistical significance of Tajima's *D* values here obtained, we have used distinct best-fit demographic models appropriated for each population (Voight, et al. 2005; Gravel, et al. 2011).

Remarkable differences in patterns of Tajima's *D* were observed in the genic regions of the two distinct *TAS* families, in the sense that positive Tajima's *D*s prevailed in *TAS2R*, whereas they were less common in *TAS1R*.

For the two *TAS1R* genes, *TAS1R1* and *TAS1R3,* which encode the proteins that form the dimeric *umami* receptor*,* both showed predominantly negative Tajima's *D* values, despite only being significantly lower than the neutral expectations for the *TAS1R3* locus in non-African populations. Significant negative Tajima's *D* values reflect an excess of rare alleles, which is a signal associated with positive selection acting on standing variation or with a standard complete selective sweeps leading to a rapid increase in the frequency of advantageous polymorphisms (Przeworski, et al. 2005). Evidence for positive selection at *TAS1R3* was clearly recorded in the patterns of sequence diversity of European populations, where the significance of the negative Tajima's *D*s persisted under the conservative demographic models assumed for those populations (Table 19).

The promoter of *TAS1R3*, reportedly associated with variability in sweet perception, was also explored through Tajima's *D* statistics, which always yielded negative values across populations from the three continents, although only leading to reject the null hypothesis of neutrality under the best-fit demographical model in the Tuscans (Table 19). Looking-alike the observed in the genic region, also for the promoter Tajima's *D*s tended to be less negative in African than in non-African populations.

In *TAS1R2,* the other member of the *TAS1R* family, no significant departures from neutrality were registered when accounting for population-specific demographical models.

Indeed, the vast majority of Tajima's *D*s was positive for the two *TAS2R* genes, with many values significantly departing from the neutral expectations. Significant positive Tajima's *D* indicates an excess of polymorphisms at intermediate frequencies, which is typically associated with balancing selection and/or population bottleneck. For the gene encoding the bitter taste receptor that influences PTC/PROP sensitivity, *TAS2R38*, significant Tajima's *D* values were observed in the five European populations examined and in one Asian, which are all populations that have experienced recent size expansions, meaning thus that demography can be excluded as the interpretation for the significant excess of moderately frequent variants at the gene. Furthermore, in most European populations the Tajima's *D* values remained significantly higher than expected even when the corresponding best-fit demographical model was assumed (Table 21). In Africa, values of Tajima's *D* were also positive but not statistically significant.

We further examined the flanking 5' and 3' non-coding region of *TAS2R38* in order to contrast our data with those provided by Campbell, et al. (2012), who had included those regions in their study. Here, the results revealed that both non-coding regions were characterized by negative Tajima's *D*, tending to be lower in the 5' flanking region than in the 3' one. Still, in most populations the relative excess of rare alleles did not exceed the neutral expectation, except in the Yoruba that presented a significant negative value when assessed under the specific demographical model. Comparatively to the genic region of *TAS2R38,* where positive Tajima's *D* were the rule, the patterns of Tajima's *D* in the flanking regions are overtly distinct, which otherwise is also fully consistent with data before reported by Campbell, et al. (2012). Once more, such comformity reassures that relying in dense, genome-wide data set with high coverage, as we did in this work, instead of full resequencing data, like Campbell, et al. (2012) did, does not seem to compromise the inferences extracted from the frequency spectra of DNA sequences.

Regarding *TAS2R16*, the gene that codes for the bitter receptor that recognizes salicin, most populations exhibited positive Tajima's *D*s, but they were only more positive than expected under the best fit demographical model in the Japanese and the Yoruba.

Haplotype relationships

In order to investigate the phylogenetic relationships across the haplotype spectrum found in each gene, median-joining networks were constructed for *TAS1R1*, *TAS1R3*, *TAS2R16* and *TAS2R38*.

For *TAS1R2* it was impossible to solve the extremely reticulated network that was drawn, for which accounted on the one hand the very high levels of sequence diversity that the gene exhibited in worldwide populations (see Table 18), and on the other hand, and probably most importantly, the presence of a high recombination region located between exons 3 and 4, as illustrated by the pattern of *linkage disequilibrium* (LD) within the gene shown in Figure S8A, where two LD blocks were identified. Download of the deCode recombination map (Kong, et al. 2002) for that region allowed to confirm its uneven recombination rate in the Hapmap YOR and CEU samples (snapshot in Figure S8B) URL deCODE recombination maps (http://www.decode.com/addendum/).

The networks for the remaining four genes are presented in Figures 21-22. For *TAS1R1*, the network showed a double star-like shape, yet rather unbalanced given that most haplotypes were closely related to a very common haplotype that represented the center of one the star structures, from which diverged by a single substitution the second most frequent haplotype, the center of the other star in the network, that also anchored many one-step neighbor haplotypes (Figure 21A). In the network, there are two tagged non-synonymous variations, C329T and G1114A, which are the two polymorphisms that have been implied in *umami* sensitivity. Reportedly, the derived T allele at C329T decreases glutamate sensitivity, whereas the derived A allele at G1114A increases its sensitivity (Chen, et al. 2009; Raliou, et al. 2009; Shigemura, et al. 2009; Bae, et al. 2010; Rawal, et al. 2013). These two polymorphisms, which were found to be in strong LD, can define four different haplotypes: the ancestral CG, assumed to confer intermediate *umami* sensitivity; CA, presumed to be a high sensibility haplotype; TG, supposed to lead to decreased sensitivity; and TA, also an intermediated haplotype that not detected in any sample used in this study.

As the network shows, CG is the central member of the major cluster of haplotypes, being very well represented in all populations. The high sensibility haplotype CA, is the center of the other star structure connecting haplotypes, and reveals more geographical structure because it is essentially confined to non-African populations. TG, the low sensibility haplotype was only found at very low frequency in non-African individuals.

In view of the features of the network, with only one high-frequent and another moderate-frequent haplotype nodes, both connected to a large number of rare lineages few steps away from the modal haplotypes, the signal emerging is that positive selection has operated in *TAS1R1* in such a way that it favored not only the ancestral haplotype configuration, which already harbored an allele associated with increased *umami* perception, but also its one step derived haplotype that additionally contains a second variant that also enhances the perception of the flavor. By other words, the network suggests that two successive, though incomplete, sweeps have occurred, the first involving the intermediate *umami* haplotype CG and the second the high sensibility CA. Actually, this represents a scenario that is quite well reconcilable with the trend observed in *TAS1R1* towards just moderate negative Tajima' *D* values that, consequently, usually do not reach statistical significance.

In the network constructed for the genic region of *TAS1R3*, a star-like shape is also very clear (Figure 21D). The two variants discriminated are those that reportedly also influence the *umami* taste perception, namely G13A and C2269T, for which the derived alleles, A and T, respectively, were both detected at higher frequencies among glutamate non-tasters or hypotasters than in glutamate tasters (Kim, et al. 2006; Shigemura, et al. 2009). Thus, the ancestral GC haplotype was assumed to confer the highest sensitivity to glutamate, followed by the two "intermediate" GT and AC haplotypes, whereas AT was presumed to account for the lowest *umami* perception. The last haplotype was not present in any sample. Contrarily, the high sensitivity GC haplotype was simultaneously the most well represented haplotype across continental populations and the modal member of the start structure that branched in many one step-derived and rare haplotypes. The two "intermediate" haplotypes were also found, mainly in non-African populations, but without phylogenetic signs evoking selection. It appears so, that adaptive pressures were exerted towards maintaining an ancestral configuration that likely augments *umami* perception.

Importantly, the clear star-like shape of this network is entirely consistent with the detection in *TAS1R3* of significant negative Tajima's *D* values in most populations, with both molecular signatures coherently pointing to the role of positive selection is shaping diversity at the *locus*.

In relation to the promoter of *TAS1R3,* the network of haplotypes was also drawn (Figure 21C) focusing attention on the two SNPs known to be correlated with human taste sensitivity to sucrose (Fushan, et al. 2009), -T1572C and -T1266C, which were found to be in significant LD. They define 3 haplotypes observed in our samples: TT, the ancestral configuration presumed to be associated with the lowest sucrose sensitivity; CT, an haplotype originated from the ancestral through a mutation that increases sucrose sensitivity, and CC, which probably arose from the "intermediate" CT haplotype after another mutation that gave rise to a second variant that also increases sucrose sensitivity. The fourth possible haplotype, TC, which could had arose through recombination between CT and TT chromosomes, was not detected in any sample. As can be seen in the network of Figure 21C, in African populations all the three detected haplotypes are quite common, with TT peaking in frequency, whereas in Eurasian populations CC is clearly more prevalent than the second most common haplotype which is TT. The network also shows that the clade of haplotypes from which CC is the central one, and only that clade, exhibits a mild star-like pattern, due to a number of low frequent radiating haplotypes. This seems to indicate that the CC background, which assumedly confer the highest perception of sweet, has been positively selected specially in Eurasian populations, which otherwise is congruent with the detection of the lowest Tajima's *D* values for the promoter in Eurasian populations, including Tuscans where the null hypothesis of neutrality was rejected under the best fit demographical model (see Table 19).

In face of these observations at the promoter of *TAS1R3,* the question arose on whether it could be due to a hitchhiking effect driven by the strong signal of selection present in the genic region of *TAS1R3,* which is associated with *umami* perception instead of sweet as is the case of the promoter. Since the promoter is shorter (~1800 bp) comparatively to the length of the genic region *TAS1R3* length (5700 bp), such an effect could indeed be a very likely possibility.

In addition, population-based patterns of *linkage disequilibrium* available from the 1000 Genomes database, revealed significant LD between -T1572C, in the promoter, and G13A, in the coding region of *TAS1R3*, in the CEU sample (see Figure S9 from supplementary material).

This led us to use full haplotypes encompassing the gene and its promoter to construct the network present in Figure 21B, where are tagged the haplotypes defined by two variations in the promoter and the two in the genic region assumed to have phenotypic effects. In this network, there are two haplotypes showing star-like shapes, one is CCGC (first two letters referring variations from the promoter; last two letters variations from the genic region), globally the most common haplotype showing a sharp radiating structure, and the other is the haplotype bearing the TTGC configuration, which globally is considerably less frequent than CCGC. These haplotypes present a very uneven distribution across continental populations, once CCGC is clearly prevalent out of Africa whereas TTGC is mainly found inside Africa. Remarkably, if these two haplotypes share the same diplotype background concerning the genic region of *TAS1R3* -i.e. GC associated with high *umami* perception, they contain opposite diplotypes for the promoter -i.e. TT and CC associated with low and high, respectively, sweet taste perception. Given this, two scenarios can be posited: i) the target of selection was the GC configuration in the genic region, and diversity at the promoter was essentially dragged by the effect exerted within the gene, or; ii) the selective advantage of the GC background in the genetic region increased even more when it became anchored in a CC background at the promoter.

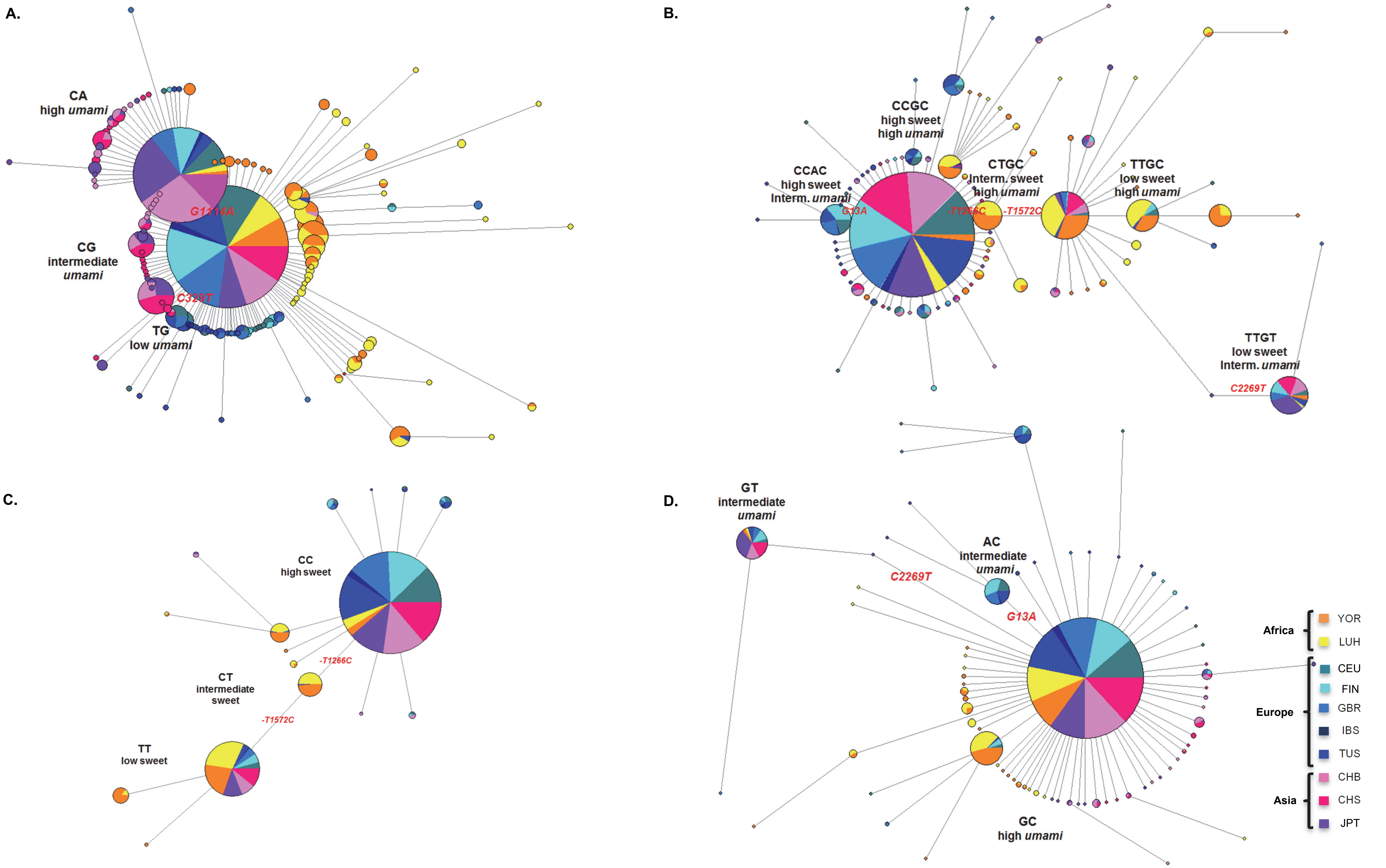


Figure 21. Network based in haplotypes' sequences of (A) *TAS1R1*, (B) *TAS1R3* gene (C) *TAS1R3* promoter and (C) *TAS1R3* genic region.
The each node represents a different haplotype, with the size of circle proportional to frequency. The length of each line is also proportional to the number of mutations that occurred in period of time. Polymorphisms in study are indicated in red.

Concerning the networks for the two genes of the *TAS2R* family, they overtly differed from those obtained for the corresponding regions of the*TAS1R* genes.

In Figure 22A it is presented the network for *TAS2R16,* the gene codifying the receptor that recognizes salicin, where it was shown up the functional variant G516T, because *in vitro* assays have proven to be critically involved in receptor function, with the ancestral G allele conferring "low-sensitivity" to salicin compared to the derived T allele that was associated with "high-sensitivity" (Soranzo, et al. 2005). The network reveals that G516T splits two groups of haplotypes, one defined by the presence of the "low-sensitivity" ancestral allele that contains relatively few lineages, rather divergent and restrictively found in Africa, and another constituted by the haplotypes carrying the "high-sensitivity" derived allele, which are much more frequent and widespread across Africa, Europe and Asia, without showing any noticeable geographical structure. Within this last group, there are two clades, one step distinguished by a common polymorphism that results in the non-synonymous variation A665G (rs860170), and the clade more rich in haplotypes shows a soft star-like shape, which might be interpreted as a sign of positive selection acting specifically over this clade of haplotypes associated with increased sensibility to salicin.

The network for *TAS2R38,* the gene that mediates taste responses to thioamides such the classic markers PTC and PROP, is depicted in Figure 22B. It reveals four main clades of haplotypes phylogenetically related by a chain of mutations leading to the tree amino acid changes defining the two predominant haplotypes in the gene, PAV and AVI, which account for the majority of the interindividual differences in PTC/PROP taste perception; PAV, the ancestral haplotype, considered to be the "taster" form, and AVI, triple derived from PAV, the nontaster form (Wooding, et al. 2004). These two haplotypes are both present at high frequencies in African and non-African populations, whereas the other two haplotypes, AAI and AAV, which in the network are midway between the most divergent PAV and AVI, and in terms of taste sensitivity are considered "intermediate" haplotypes, show striking distributions: AAI occurs globally at moderate frequency, but 65% of the here detected lineages belonged to Africans whereas only 28% and 7% were carried by Europeans and Asians, respectively; in turn, AAV, besides absent from Africa was comparatively much more rare, especially in Asia (of the 13 detected AAV haplotypes, 12 were from European individuals against one Asian). Overall, the structure of the network for *TAS2R38* seems to be compatible with a model of balancing selection maintaining elevated frequencies of the most divergent haplotypes, PAV and AVI, across both African and Eurasian populations.

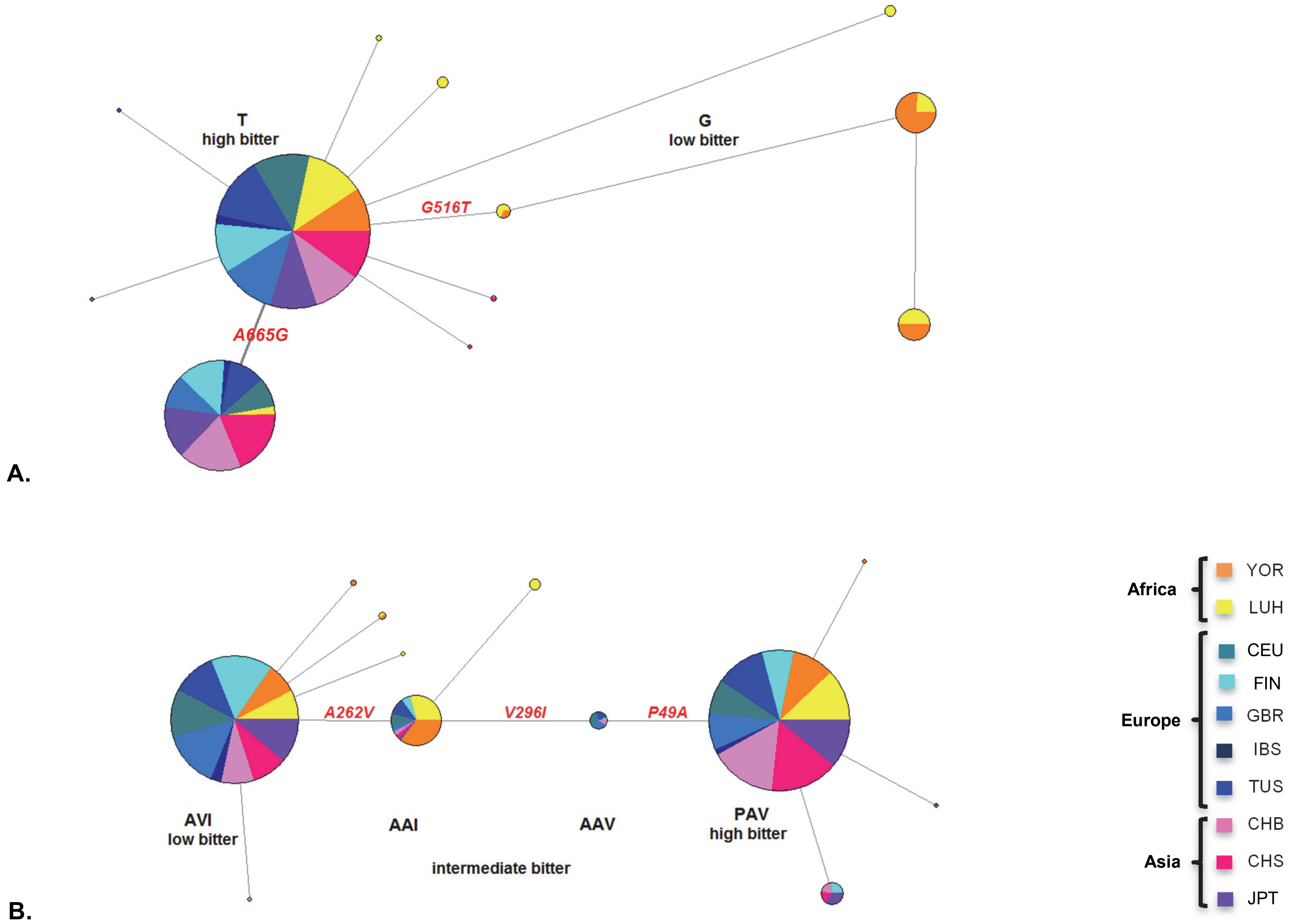


Figure 22. Network based in haplotypes' sequences of (A) *TAS2R16* gene and allele frequencies of (B) *TAS2R38*.
The each node represents a different haplotype, with the size of circle proportional to frequency. The length of each line is also proportional to the number of mutations that occurred in period of time. Polymorphisms in study are indicated in red.

Population differentiation

In order to compare levels of genetic differentiation across African and Eurasian populations for the different *TAS1R* and *TAS2R* genomic regions, estimates of global F_{ST} were computed from complete sequences with values being shown in Table 22.

Table 22. Genetic distance (F_{ST}) values and correspondent *P*-values for all the set of genes addressed in this study.

	F_{ST}	***P*-values**
TAS1R1	0.10264	0.00000
TAS1R2	0.06475	0.00000
TAS1R3*	0.29872	0.00000
TAS1R3**	0.07851	0.00000
TAS2R16	0.21365	0.00000
TAS2R38	0.04791	0.00000

*TAS1R3 promoter
** TAS1R3 genic region

To put those results in a genome wide context, we recall data provided by Elhaik (2012) on the empirical genome wide (2,823,367 autosomal SNPs) distribution of F_{ST} in human populations ascribed to the continents of Africa, Asia and Europe, indicating that the mean F_{ST} was 0.08. Before, Akey, et al. (2002) reported a mean F_{ST}=0.12 using 25,549 autosomal SNPs genotyped in African-American, East Asian, and European-American, and in a later study Shriver, et al. (2004) analyzing 8,525 autosomal SNPs in 84 African-American, European-American, Chinese, and Japanese individuals described a mean F_{ST}=0.13, although these two reports relied on relatively small samples of common SNPs.

In the light of this range of F_{ST}, looking for the estimates here computed there are two values that reveal unusual excessive differentiations in global populations: one concerned the promoter of *TAS1R3* (F_{ST}=0.299; *P*=0.000) and the other concerned *TAS2R16* (F_{ST}=0.214; *P*=0.000). In contrast, *TAS2R38* yielded the F_{ST}=0.048 (*P*=0.000), which is much lower than any of the average F_{ST} assumed as reference for autosomal *loci*.

Comparing F_{ST} for the *TAS1R* genes, the highest differentiation was in *TAS1R1* (F_{ST}=0.103; *P*=0.000), followed by *TAS1R3* (F_{ST}=0.079±; *P*=0.000) and then *TAS1R2* (F_{ST}=0.065; *P*=0.000). Of note that the promoter of *TAS1R3* revealed the F_{ST}=0.299

(*P*=0.000), which was absolutely out of the range of the values computed for the genic regions of the *TAS1R* family.

It also deserves mention, the differences in F_{ST} in the two *TAS2R* genes, once for *TAS2R16* global human populations are four times more differentiated than are for *TAS2R38* (*TAS2R16* F_{ST} =0.214; *P*=0.000; *TAS2R38* F_{ST} =0.048; *P*=0.000).

Gene trees and TMRCA estimates

To further explore the mutation history of the genomic regions under study, gene trees were constructed for *TAS1R1, TAS1R3, TAS2R16* and *TAS2R38* as well as for the promoter of *TAS1R3* (Figures 23-24) using the software GENETREE. Respecting *TAS1R2,* it was impossible to deal with the many sequences incompatibilities reported when the program was running, which stemmed from the reasons before mentioned that hampered the construction of the corresponding network.

Besides the trees, GENETREE also provides maximum likelihood estimates of population mutation rate (θ), which is specific from each genomic sequence, and it further computes the distribution of the time to the most recent common ancestor (TMRCA) and ages of target mutations in each tree (Griffiths and Tavare 1994). These values together with the corresponding trees are present in the Figures 23-24.

Making first a global comment on all trees, it is of note the consistency between the topology and structure of each tree and the haplotypic phylogenetic relationships portrayed in the correspondent networks.

It also stands out the differences in coalescent times obtained for variations in *TAS1R* versus *TAS2R* sequences.

The TMRCAs for *TAS1R* regions were dramatically more recent than for the *TAS2R* genes, pointing to approximately 130±63 Kya and 233±44 Kya for *TAS1R1* and *TAS1R3* genic regions, respectively, and 414±129 Kya for the promoter of *TAS1R3*. All these tree values are below the upper limit of the first quartile of ages estimated for autosomal *loci* by Blum and Jakobsson (2011).

The mean TMRCA values for *TAS2R16* and *TAS2R38* was respectively ~3.8±1.11 and ~1.7±x0.55 Mya, with both values being higher than the average TMRCA estimate for autosomal *loci*, which according to Blum and Jakobsson (2011) is approximately 1.5 Mya (first quartile=950,000 years and third quartile=1,700,000 years).

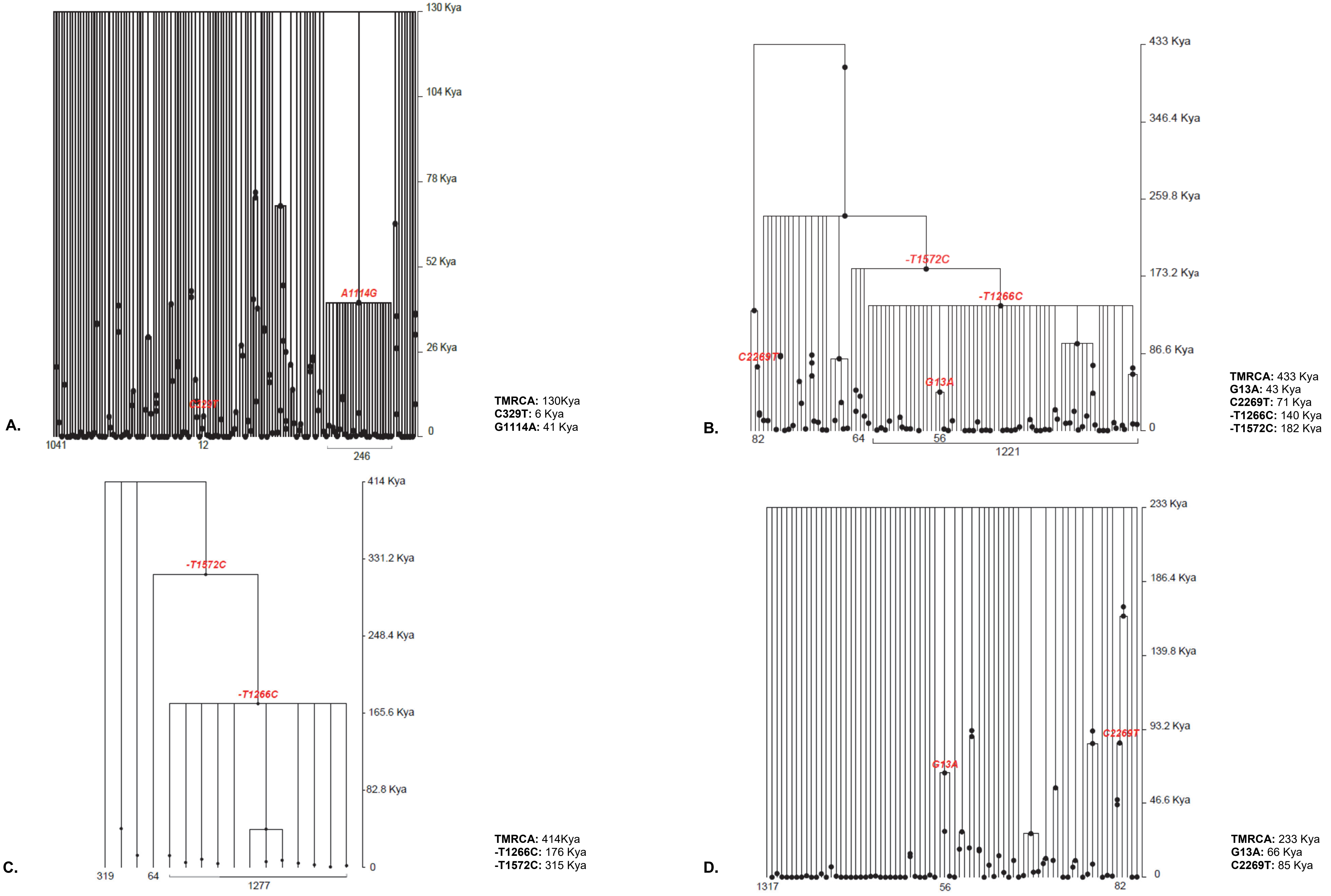
These findings on coalescence times for *TAS1R* versus *TAS2R* sequences here analyzed must be interpreted in the light of the well-established principles i) that genes under balancing selection, maintaining different alleles at a selected locus, have deeper genealogies than expected under neutrality, with long internal branches; while ii) positive selection distorts the structure of gene genealogies expected under neutrality, altering patterns of genetic variation within a population, once beneficial mutations, increasing rapidly in frequency, purge genetic variation at linked sites and shorten coalescence times near the selected locus (Smith and Haigh 1974).

Dissecting now more deeply the trees, the two corresponding to the genes that encode the TAS1R1-TAS1R3 heterodimer acting as an *umami* taste receptor, both denoted signatures that fitted well a model of positive selection accounting for their evolutionary mechanisms in human populations. The genealogies besides showing very recent coalescent times, further revealed signs of remarkable rapid lineage diversification in specific haplotypic backgrounds defined by variations assumed to have phenotypic implications. In the case of *TAS1R1* (Figure 23A), which yielded the youngest TMRCA in this study (~130±6 Kya), an excess of rare variants was clearly associated either to the ancestral CG configuration defined by the functional variations C329T and G1114A, which is an haplotype that confers intermediate *umami* sensitivity; or to its one-step derived CA haplotype, which was originated through a recent mutation at nucleotide 1114, estimated to have occurred only ~41±9 Kya ago, that increases more *umami* sensitivity.

As for *TAS1R3* genic region, whose lineages also coalesce quite recently at ~233±44 Kya, the sharp excess of rare variants was uniquely detected in the GC ancestral haplotype defined by the functional polymorphisms G13A and C2269T (Figure 23D). This is the haplotype that presumably confers the highest *umami* sensitivity, whereas those diminishing sensitivity to the flavor, which arose through mutation at any of the two positions, do not show any signals of rapid lineage diversification. It could be argued that the age of mutations G13A and C2269T, ~66±9 Kya and ~85±10 Kya, respectively, are not old enough for permitting significant accumulation of lineage divergence, but that does not explain the mark of positive selection acting on the ancestral GC haplotypic configuration.

The tree for the promoter of *TAS1R3* (Figure 23C), in which lie the two variations -T1572C and -T1266C correlated with sucrose perception (note that the product of *TAS1R3* is common to both *umami* and sweet receptors), detained a TMRCA deeper than the genic region, pointing to ~414±130 Kya, but even so falling much below 950 Kya, the limit of the first quartile of the estimates for autosomal *loci* (Blum and Jakobsson 2011). The ages of mutations -T1572C and -T1266C were ~315±105 Kya and ~176±49 Kya, respectively, meaning that both roughly trace back to the period during which modern humans emerged, around 200 Kya. Still, signals of diversification were only found in the lineages carrying the derived alleles at the two positions, which appeared approximately ~176 Kya ago, creating the haplotypic background that increases sensitivity to sucrose.

Similarly to that we made in the network analysis, full sequences of the promoter *plus* the genic region of *TAS1R3* were used to construct the tree presented in Figure 23B. It was associated with a TMRCA slightly deeper (~433±101 Kya) than obtained for the promoter, and the ages of the four variations assessed were also a little older than estimated when dealing separately with the two regions. Even though, the broad temporal frame provided by the tree of full haplotypes did not changed much compared to the trees of partial sequences. Moreover, the stronger signal of lineage diversification in the “total” tree involved the CCGC haplotype associated with high sweet/high *umami* sensibility.



TMRCA: 130Kya
C329T: 6 Kya
G1114A: 41 Kya

TMRCA: 433 Kya
G13A: 43 Kya
C2269T: 71 Kya
-T1266C: 140 Kya
-T1572C: 182 Kya

TMRCA: 414Kya
-T1266C: 176 Kya
-T1572C: 315 Kya

TMRCA: 233 Kya
G13A: 66 Kya
C2269T: 85 Kya

Figure 23. Genealogy, TMRCA of global variation and ages of individual mutations of (A) *TAS1R1*, (B) *TAS1R3* gene (C) *TAS1R3* promoter and (D) *TAS1R3* genic region. Mutations are represented by black dots and the number on the bottom corresponds to the number of individuals with that haplotype; SNPs addressed in this study are highlighted in red.

Quite dissimilar were the trees for *TAS2R* genes. *TAS2R16* revealed the very deep TMRCA of ~3.8±1.11 Mya, which is among the oldest time estimates for autosomal *loci* (Blum and Jakobsson 2011). For that accounted the presence in the gene tree of two major clusters of haplotypes rather divergent that are distinguished by alleles at the functional variant G516T that influences salicin recognition. Hence, one cluster contains the haplotypes associated with decreased sensitivity to salicin and the other clusters the haplotypes leading to increased sensitivity. The age yielded for the derived mutation, 516T, was also very ancient, backing in time to ~1.6 Mya, meaning then that enhanced bitter perception arose very early in the *Homo* lineage, much before the emergence of anatomically modern humans ~200 Kya ago. Within the cluster of increased sensitivity haplotypes, there is a sub-cluster of lineages, well represented across African and Eurasian populations, that also carry the derived allele for a more recent non-synonymous mutation (665) that arose 614 Kya (see network from Figure 22A), but only the sub-cluster of high sensitivity haplotypes that maintain the A ancestral allele at 665 shows some signals of lineage diversification.

The gene genealogy for *TAS2R38* (Figure 24B) also yielded a deep coalescence age for the beginning of sequence diversification within the gene, estimated at ~1.7±0.55 Mya. The mean age estimates for the three variations that define the haplotypes associated with variability in bitterness perception of thioamides are accordingly as well very ancient and recapitulate the chronological order of appearance of the mutations: the derived allele leading to the amino acid change in position 49 emerged first, ~1.2±0.55 Mya ago, next arose the mutation altering the amino acid at position 262, approximately ~673±122 Kya, and finally ~345±120 Kya before present occurred the mutation that changed the amino acid 262.

It appears, then, that from the ancestral PAV haplotypic configuration have derived by single mutational steps the remaining haplotypes, by the order: PAV→AAV→AAI→AVI, all in times preceding the emergence of anatomically modern humans. Still, only the two extremes in this chain reach currently high frequencies across worldwide populations: the ancestral, PAV, and its triple derived configuration, AVI, demonstrated to confer, respectively, the lowest and highest responses to thioamides (Bufe, et al. 2005; Behrens, et al. 2013). Concerning these two haplotypes, the gene genealogy of *TAS2R38* remarkably portrays its rather deep antiquity, which is a feature consistent with a model of balancing selection, maintaining PAV and AVI at similar and high frequencies in human populations.

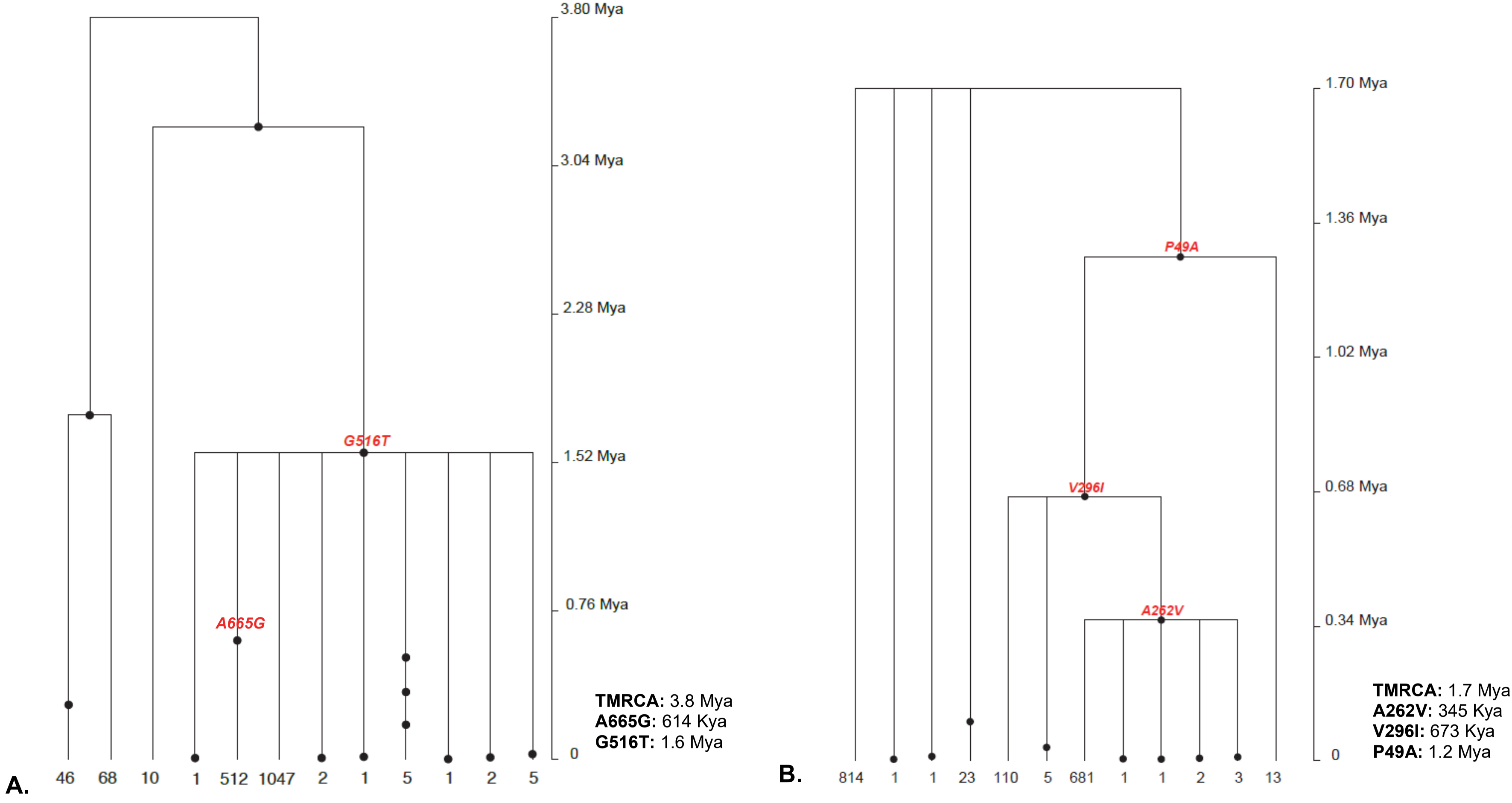


Figure 24. Genealogy, TMRCA of global variation and ages of individual mutations of (A) *TAS2R16* and (B) *TAS2R38* genes.
Mutations are represented by black dots and the number on the bottom corresponds to the number of individuals with that haplotype; SNPs addressed in this study are highlighted in red.

3.3.4 Conclusions

Because the sense of taste is thought to be crucial for monitoring the intake of nutrients, not only warning on environment hazards, but also largely contributing to fulfill the need for nutrients, it has been assumed that it played a significant role throughout the course of human history, during which human taste preferences and aversions have changed (Breslin 2013; Doty 2015).

To better understand the evolution of genes encoding bitter, sweet and *umami* receptors, in the present study we have examined their patterns of sequence diversity across major human populations, viewing which we have mined the public genomic database provided by the 1000 Genome Project. Although previous studies have already provided valuable, although fragmentary, evidence on the issue, this is the first work where i) genes from two different *TAS* families were simultaneously investigated; ii) the analyses performed were the same for all genomic segments, and iii) the set of populations was identical for each of the distinct sequences. Thus, comparisons of results between *loci* can reasonably illuminate differences in evolutionary histories, since the ambiguities and lack of reproducibility resulting from working with heterogeneous approaches and data sets were filtered.

One of our main findings was that the two genes from the *TAS1R* family, *TAS1R1* and *TAS1R3*, have evolved quite dissimilarly from the two genes from the other family we could fully analyze, *TAS2R16* and *TAS2R38*.

Both genes *TAS1R1* or *TAS1R3* involved in coding for the heterodimer TAS1R1-TAS1R3 and that mediates *umami* taste perception, evidenced many signatures expected for regions that were a target of positive selection: their haplotypes networks showed clear “star-like” structures; the sequences coalesced in very recent times, roughly in the late Middle Pleistocene during which early anatomically modern humans emerged; and Tajima’s *D*s were systematically negative in all African, European and Asian populations, even not always reaching statistical significance, as was the case of *TAS1R1*. For this gene the lack of significance of the negative Tajima’s *D*s is still in line with a model of two partial selective sweeps, which in fact can explain the detection in *TAS1R1* of two haplotypes with signals of having experienced rapid expansion and diversification.

Noteworthy, whereas for *TAS1R1* the two haplotypes with signatures of positive selection are the ancestral CG associated with intermediate *umami* perception, and its one step derived CA haplotype that increases sensibility to the perception, this latter originated around 41 Kya ago few after modern humans initiated the dispersion out of Africa; for *TAS1R3* the selected haplotype harbors the ancestral GC background also associated with increased *umami* sensibility. Taken together, these observations indicate that newly arising variations at *TAS1R1* or *TAS1R3* diminishing the *umami* perception might have broken the adaptive role that seemingly had the pre-existing configurations. Thus, the sweeps detected in *TAS1R1* and *TAS1R3*, in which beneficial alleles seem to have arisen from standing variation, keep well with the commonly called soft sweep (reviewed in Wollstein and Stephan (2015)).

The product of *TAS1R3* also participates in the taste receptor TAS1R2-TAS1R3, which largely modulates sweet perception in humans. However, as of now only two variations at the promoter of *TAS1R3* have been directly correlated with human taste sensitivity to sucrose (Fushan, et al. 2009).

Our analyses on this promoter led to identify in the region a signature of a recent selective sweep, although leaving a soft trace then that detected in the genic region of *TAS1R3*. The haplotype of the promoter showing to be under the action of positive directional selection, was the CC derived configuration, associated with increased sensitivity to sucrose, which was estimated to have been originated ~176±50 Kya ago. These molecular signatures are consistent with the conventionally referred to hard sweep, expected under the classical model of positive selection occurring when *de novo* derived alleles have higher fitness than the ancestral ones (reviewed in Wollstein and Stephan (2015)), but in the case of the promoter of *TAS1R3* the body of evidence pointed to an incomplete hard sweep in a scenario of local adaptation. The signal of human local adaptation emerged from the high global differentiation across African, Asian and European populations observed at the promoter, reaching the F_{ST}=0.299, which was the highest value detected in this study. It was striking the uneven distribution of haplotypes conferring high, CC, and low, TT, sucrose sensitivity, with the first being much more prevalent in Eurasia than in Africa. Symptomatically, Tajima's *D*s values were clearly negative in all Eurasian populations despite not lower than expected under the stringency of the best-fit demographical models, except in the Tuscan. Even so, indications seem enough to sustain that the CC haplotype that increases sensitivity to sucrose must have been advantageous for people living in

Europe and Asia, regions where the ancestral TT haplotype it is also present but at moderate frequencies and lacking signals of adaption. This last observation indicates that like CC (or the minority CT) the TT haplotype was also exported from Africa when modern humans migrated from the continent, but in the meantime only CC increased to high frequencies out of Africa, which strongly suggests that this haplotype has become to be positively selected after the environmental shift that humans faced when dispersed out of Africa.

Still and very likely, the evolution of the promoter and the evolution of the genic region of *TAS1R3* are only two sides of the same coin. Up to now, the promoter has only been connected with sugar taste sensibility, and in turn the genic region has only been linked with *umami* perception. To the best of our knowledge it was never examined the role of the variations that influence the promoter activity in *umami* sensibility, or *vice versa* the role of the variations that influence *umami* perception in sugar sensibility.

Given this, it is not possible to exclude that the signals captured in the promoter and in the genic region of *TAS1R3* had been due to the same selective pressure, of nature difficult to decipher for the moment, and not necessarily related with taste perception.

Just as argued for the *TAS2R* genes, the role of *TAS1R1* or *TAS1R3* in *umami* sensing might be only one of the functions performed by the TAS1R1-TAS1R3 taste receptor, as well as the role of the *TAS1R3* promoter in the perception of sweet might not exhaust the physiological function of the sweet receptor TAS1R2-TAS1R3.

The two *TAS1R* genes that code for the heterodimer TAS1R1-TAS1R3 that mediates *umami* taste perception, revealed very different evolutionary histories compared to *TAS2R* genes.

Our analyses on *TAS2R16* uncovered a puzzling evolutionary history of diversity at the locus. The tree yielded for *TAS2R16* revealed a peculiar genealogical structure, with a TMRCA particularly old, ~3.8 Mya, of an order extremely rare for autosomal *loci*, falling in a time frame preceding the emergence of the earliest representatives of the genus *Homo*. The tree also possessed a basal clade and another sub-clade composed uniquely by African sequences (see network in Figure 22A), feature that does not surprise due to the restrictive geographical distribution, but own to the fact of being very deep branches with too few derived polymorphisms. Signals of lineage diversification were limited and solely observed in the subclade defined by the mutation at position 516 associated with increased sensitivity to salicin, which is also a very old mutation, estimated to have arisen

~1.6 Mya ago, explaining not only some diversification but also its current elevated frequency across worldwide populations. The tree topology was fully reconcilable with the atypical high F_{ST} value of 0.214 observed for *TAS2R16* across continental populations. Theoretically, *loci* involved in local adaptations are predicted to show rather high F_{ST} values, but in the case of *TAS2R16* it is difficult to envision the scenario of environmental circumstances accommodating its striking evolutionary dynamics. Actually, it is not easy to explain why very ancient lineages carrying an ancestral allele associated to diminished sensibility to salicin, persisted uniquely in Africa, where as a whole they reach considerable frequencies (> 25%), co-existing with the much more prevalent high sensibility derived lineages, whose defining mutation is too very ancient. In our dataset, the high sensibility allele was fixed in Europe and Asia, suggesting therefore that only inside Africa some selective pressure has acted towards maintaining a large spectrum of phenotypic variability associated of this bitter taste receptor.

Selection at *TAS2R16* had already been investigated in earlier studies by Soranzo, et al. (2005) and more recently Campbell, et al. (2013), using similar methodologies to those applied in this study, yet based on a number of African populations clearly overrepresented comparatively to non-African ones. Campbell, et al. (2013) sustained that *TAS2R16* had evolved under positive selection acting on ancient standing variation, favoring lineages anchored in the high-sensitivity allele at site 516, whereas strong purifying selection had acted in the haplotypes carrying the low-sensitivity ancestral allele. Soranzo, et al. (2005) also reported on a significant signal of positive selection favoring the 516 derived allele, indicated by the excess of rare derived polymorphisms in such lineages. However, in a still another work addressing selection at *TAS2R16,* where a much more equilibrated set of worldwide populations was studied although applying a different methodology for detecting selection, no signals of non-neutral evolution were found in 8 out of 9 African populations examined, while contrarily most of the studied Eurasian populations showed evidence for recent positive selection at the gene (Li, et al. 2011). Despite some conflicting signals, these three studies coincided in postulating positive selection as a main shaper of diversity in sequences carrying the high-sensitivity derived allele at site 516.

Still, our interpretation of the data here yielded, does not led to exclude that pure chance and demography might be the more parsimonious explanation for the distribution observed outside of Africa. First, the mutation creating the high-sensitivity core haplotype was originated long time ago, much before the emergence of anatomic modern humans. Time was enough for allele drifting to increased frequency, eventually rising above the ancestral configuration. Suggestively, inspection of the Neanderthal and the Denisova genomes revealed that both carried the high-sensitivity allele. Second, outside of Africa, the fixation or near-fixation of this allele (in the Middle East it is present at very low frequencies (Soranzo, et al. 2005)), can plausibly be the result of the bottleneck effects associated with the migration of modern humans out of Africa. In addition, in Eurasian populations no excess of rare derived haplotypes anchored in the high-sensitivity variation was detected; the "star-like" shape radiating from the core haplotype was not structured enough to result in negative Tajima's *D* values, as predicted in a model of positive selection, and quite the contrary most Eurasian populations displayed positive Tajima's *D* values, not statistically significant assuming the best proxies of their demographical past. Third, the extremely high F_{ST} values across African, European and Asian population are also at odds with the expectations under a model of recent positive selection acting in the gene, because such high level of inter-population differentiation was essentially due to the presence limited to Africa of remarkably divergent *TAS2R16* lineages carrying the ancestral less-sensitive 516 allele. Importantly, the Yoruba from Africa showed a significant Tajima's *D* under the best fit demographical model, but in fact it was more positive than expected assuming neutrality, apparently favoring a scenario of long-lasting balancing selection. This finding raises the obvious question of why such divergent and ancient lineages were retained till present in African populations, at frequencies far from being residual, resisting so to the erosion afforded by million years of stochastic processes. It might have been pure chance, but as an alternative explanation we reiterate Campbell, et al. (2013) that strong negative selection must have operated maintaining those relic lineages in Africa. If this hypothesis holds true, is still necessary to clarify its biological consistency. Up to now the receptor *TAS2R16* is mainly knew by its influence in the response to bitter tasting β-D-glucopyranosides like salicin, apparently directing towards a diet-related role. Yet, does this environmental-related pressure also include protection against parasites/pathogens, for instance? This is an issue that deserves to be explored in the future, especially given recent research that is deciphering a range of non-gustatory functions for the taste receptors, including in aging, in immune response, among others (Campa, et al. 2012).

Finally, *TAS2R38* gene conciliates the most emblematical signals expected under a model of balancing selection: low level of differentiation across continental populations, Tajima's

D values systematically positive and often statistically significant under the best fit demographical models, and sequences that coalesce in a deep TMRCA indicating the action of long-term balancing selection. Furthermore, the two common and widespread haplotypic configurations that appear to be maintained by balancing selection are associated to opposite phenotypic effects in terms of taste, which also agrees with the expectations under that selective model. One is the ancestral PAV haplotype, the assumed taster form, and the other is the non-taster AVI haplotype, which was estimated to have originated ~345 kya, considerably before the emergence of anatomically modern humans. It follows that the selective pressures underlying diversity at *TAS2R38* might not be directly related with the phenotypic effect in bitterness sensibility, but instead with other functional consequence for which the variations PAV and AVI account that since long promoted their maintenance at high frequencies in human populations.

Recent research has proven that *TAS2R38* as well as other bitter and sweet taste-related receptors, until recently believed to be present only in taste buds, also exist elsewhere in the body, most notably in the gastro-intestinal and respiratory tracts (Shah, et al. 2009). Furthermore, the role of *TAS2R38* in airway physiology and disease susceptibility has just started to be identified in the last few years (reviewed in Lee and Cohen (2015)). So, it is still necessary to understand better the physiological function of *TAS2R38* in order to clarify the likely biological explanation for the clear signature of long-lasting adaptive evolution currently retained in the gene.

In summary, our observations on *TAS* genes from two families, revealed many unconformities with evolutionary neutrality, with signatures of selection fitting different modes of adaptive evolution in each of the TAS families, which in general was consistent with previous observations (Soranzo, et al. 2005; Fushan, et al. 2009; Raliou, et al. 2009; Eny, et al. 2010; Campbell, et al. 2012; Campbell, et al. 2013; Robino, et al. 2014).

The challenge now is to identify the causes of selection acting on a number of *TAS* genes that are very well known as players in taste physiology, but for which it is still very new the knowledge that they are expressed in a plethora of tissues, that many can affect multiple physiological pathways, including in respiratory and digestive functions, and that they influence traits like longevity, for instance (Bachmanov and Beauchamp 2007). As stressed by Campa, et al. (2012), the emerging picture is that taste receptors can behave as pleiotropic genes, whose products are used by various cells, or have signaling function on various targets not linked one to the other, and so they might play a broader role in human health.

It is today acknowledged that classic selective sweeps have been rare in recent human history, and so far the most prominent examples involve *loci* with large-effects alleles (see for reviews Fu and Akey (2013) or Wollstein and Stephan (2015)), including the diet-related lactose tolerance in adulthood (Tishkoff, et al. 2007), or the adaptation to low oxygen at high altitude (Alkorta-Aranburu, et al. 2012; Li, et al. 2014), among other. Still, it is being shown that selection in humans, particularly local selection, has been dominantly driven by adaptive responses implicating pathogens, which is understandable given that pathogens have represented, and still represent, one of the major causes of death for human populations (Fumagalli, et al. 2011; Karlsson, et al. 2014).

Recent research has identified novel roles of T2R bitter and T1R sweet taste receptors in the human airway, indicating that they contribute to regulate human upper respiratory innate immunity (Lee, et al. 2014). Previously Lee, et al. (2014) demonstrated that among the harmful compounds detected by T2Rs in the human airway, were included secreted bacterial products.

In view of these findings, as a future line of research we anticipate that in *TAS* genes harboring signatures of selection it should be explored their possible effects, probably pleiotropic/epistatic, in the interactions between humans and infectious agents, in order to obtain further insights into the still unclear narrative on the role of *TARs* genes in human adaptation.

3.3.5 Supplementary Material

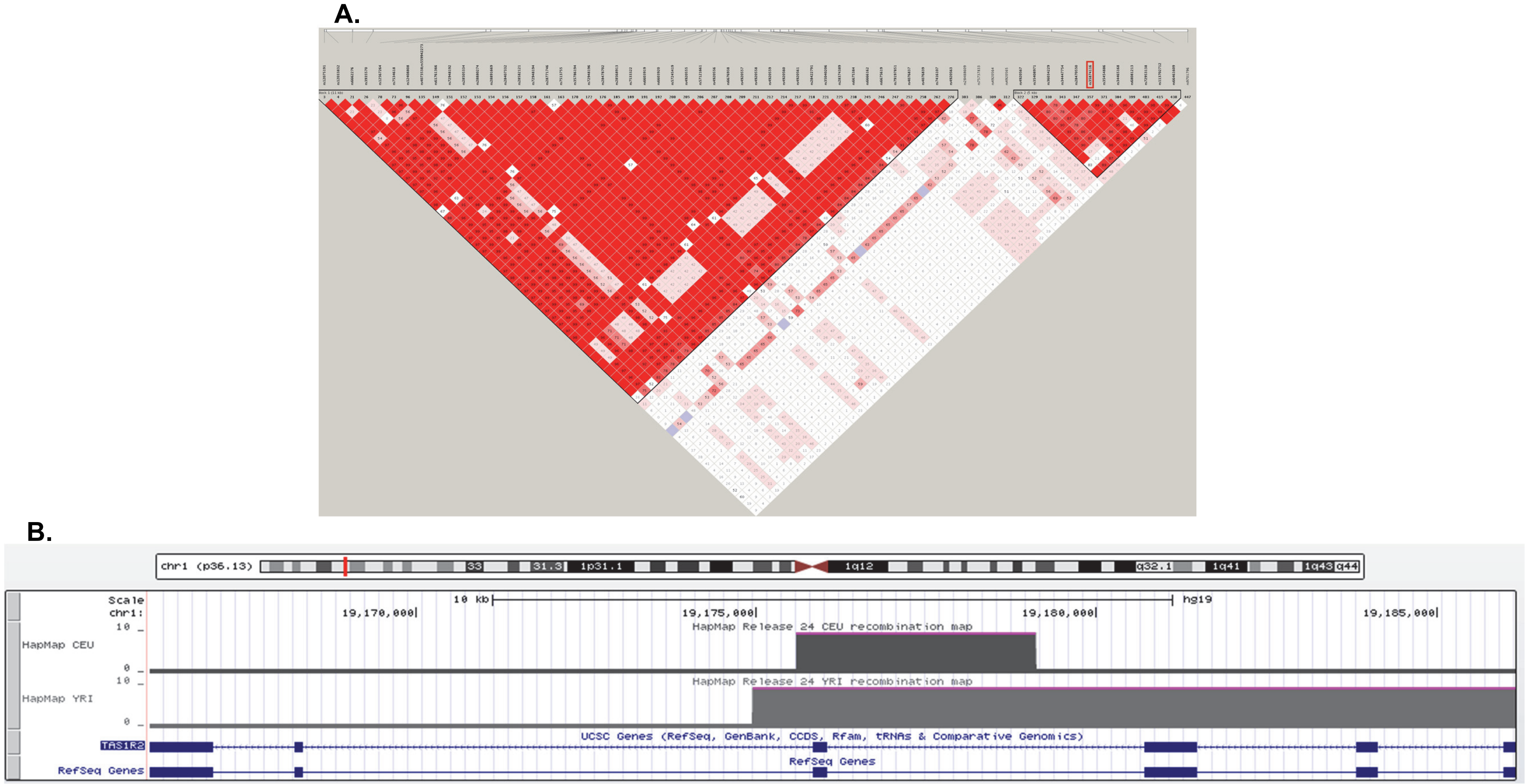


Figure S8. (A) LD patterns for *TAS1R2* exons 3 and 4 in the Yoruba from 1000 Genomes Database and (B) deCODE Recombination map for the *TAS1R2* complete gene in Yoruba and CEU populations.

The studied SNP is indicated by a red rectangle. In black triangles are represented the LD blocks. The degree of LD between pairs of markers is indicated by the |D'| statistic (|D'| = 1, red; |D'|<1, shades of red);

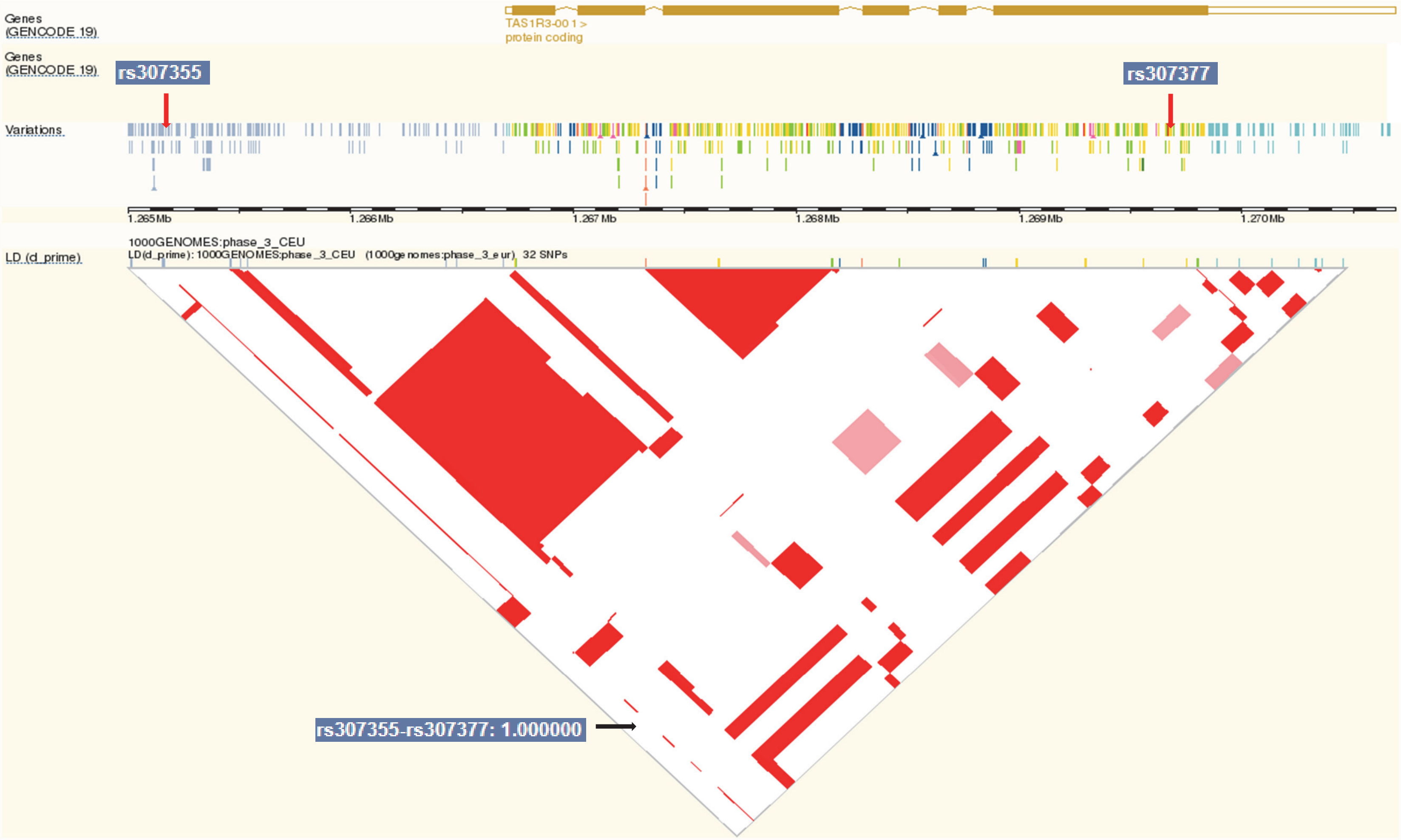


Figure S9 LD patterns for *TAS1R3* in CEU population from 1000 Genomes Database.
The two SNPs that are in LD are indicated by a red arrow: rs307355 (-T1572C) and rs307377 (C2269T). The degree of LD between pairs of markers is indicated by the |D'| statistic (|D'| = 1, red; |D'|<1, shades of red).

CHAPTER 4. FINAL REMARKS

As humans moved out of Africa to settle on the remaining parts of the world, they were exposed to new environments, diets, pathogens and other inexperienced pressures that compelled them to adapt, leading in some extent to the great diversity we observe today. This starting migration was accompanied by a series of demographic events, like consecutive bottlenecks followed by populations expansions, that also accounted much for the degree of structure that nowadays characterizes human populations. Uncovering the mechanisms underlying current patterns of diversity in human populations has been a long-standing interest in evolutionary genetics. As a matter of fact, it represents a very complex issue because is not easy to discern between genomic signatures due to selection from those due to demographic events, which often mimic each other.

Knowing the broad picture of the dispersion history of humans, it makes sense the accumulated evidence indicating that many adaptations in non-African populations have occurred in the recent past, while selection was a less important determinant in various African populations (reviewed in Wollstein and Stephan (2015)).

Different approaches applied to perform genome scans for selection have shown that the most consistent signs of human adaptations are related with i) pathogen resistance (Balaresque, et al. 2007; Barreiro and Quintana-Murci 2010; Daub, et al. 2013), ii) environmental variables, as amount of light, temperature, or oxygen levels (Balaresque, et al. 2007; Hancock, et al. 2008; Coop, et al. 2009; Coop, et al. 2010; Hancock, et al. 2011; Alkorta-Aranburu, et al. 2012; de Magalhaes and Matsuda 2012; Li, et al. 2014), and iii) diet/agriculture (Hancock, et al. 2010; Luca, et al. 2010; Scheinfeldt and Tishkoff 2013).

Two dietary adaptations are among the most paradigmatic and consensually accepted examples of human genetic adaptations, namely the ability to digest lactose from milk products in adulthood (a clear signature for a selective sweep around the *LCT* gene was replicated in several studies) (Tishkoff, et al. 2007; Ranciaro, et al. 2014), and another that involves a copy number variation of the gene for amylase, which was demonstrated to be associated with diets containing different amounts of starch (Perry, et al. 2007).

On these grounds, we sought to apply a candidate gene approach to investigate whether differences in diets, roughly inferred from lifestyles of populations, could also have shaped diversity at other genes that more or less directly play a role in the metabolism or in the taste perception of substances (including xenobiotics) that gain entry into the organism through food.

So, we began this work (section 3.1) searching for signals of selection in five genes involved in metabolic pathways that can influence food processing: *AGXT*, *CYP3A5, MTRR*, *NAT2* and *PLRP2*. Viewing that, we first screened 5 African groups with distinct lifestyles and the Portuguese to obtain a background of population data as enriched as possible to conduct the analyses.

With respect to *AGXT* and *MTRR,* no indications were found that their global distributions departed from the expectations under neutral evolution or have been influenced by lifestyle of populations. It was hypothesized that a common variation at *AGXT* was positively selected in populations with meat-rich diets (Caldwell, et al. 2004; Danpure 2006), while variations at MTRR could have been adaptive to a dietary specialization in folate-poor food (Hancock, et al. 2010).

Here, however, we failed to replicate these previous trends pointing to a role of selection modeling diversity at both *loci*. Since our study was based in a more comprehensive set of populations, able thus to provide more reliable conclusions, it seems likely that the before reported signs were only false positive associations resulting from unintentional bias in the number and/or kind of populations analyzed.

Concerning the genes *NAT2* and *PLRP2,* the results here obtained revealed some signs that both might represent instances of past dietary adaptations.

Diversity at *PLRP2* was found to be associated with the mode of subsistence in populations, as reflected in the clinal decrease in frequency of a candidate variant from farmers, next pastoralists towards hunter-gatherers. As a whole, our observations fitted an adaptive model holding that cereal content (or other *PLRP2* substrate, or amount of substrate) differing in hunter-gatherers, herders and farmers worked as a selective pressure contributing to shape the current pattern of diversity at *PLRP2* in human populations.

For *NAT2*, significant differences between hunter-gatherers and food-producer populations were detected in the widespread slow acetylator variant, which further revealed an unusual low level of geographical structure across Africa and Eurasia, indicating that it was subjected to strong drifting constrains. Our study adds support to an adaptive evolution of *NAT2* that assumedly could have been related with the diminished availability of folates in diet brought with the shift from economies relying in hunting and gathering to those based on farming and herding of domesticates.

A strong, unequivocal signature of selection was identified at *CYP3A5*, which displayed a level of inter-population differentiation dramatically surmounting even the most conservative neutral expectations. However, the amount of salt ingested across main dietary habits - which we hypothesized to be the biological link between *CYP3A5* and dietary adaptation, did not accounted for the distribution of *CYP3A5*, which instead was highly determined by the geographical location of populations in the North or in the South of the Tropic of Cancer. This finding supports an evolutionary history for *CYP3A5*3* under a selective pressure determined by an environmental factor correlated with latitude (Thompson, et al. 2004), which apparently does not depend on differences in eating habits between populations.

Having found out that at least for *PLRP2* and *NAT2* genes, the Neolithic transition to subsistence economies based on agricultural and pastoral resources likely represented a turning stage for new selective pressures triggered by changes in diets, raised the question on whether other genes implied in functions associated with food consumption could also have been adaptive during human evolution.

Since the sense of taste is crucial for guiding dietary habits and many aspects of human health' balance, we focused attention in genes codifying taste receptors for sweet, bitter and *umami*, because these taste sensations are considered the most important for food acceptance (Temussi 2009).

We first selected individual variations in 4 genes from the *TAS1R* (*TAS1R1* and *TAS1R3)* and *TAS2R* (*TAS2R16* and *TAS2R38)* families reported to have phenotypic effects in terms of taste, which were then screened in the same 6 population groups before mentioned (5 Africans plus the Portuguese), having in mind to apply a strategy similar to that used to address selection in the genes involved in metabolic pathways. However, it was only possible to perform some introductory analyses (section 3.2) due to the critical lack of population data for the *TAS* genes.

Even so, for three of the taste genes - *TAS1R1*, *TAS1R3* (gene and promoter regions) and *TAS2R38*, no minor signals came out that the subsistence economies, or, indirectly, diet-related variables in populations could have influenced their patterns of frequency distribution. Concerning *TAS2R16*, which codifies a bitter receptor that recognizes β-glucopyranosides compounds, strong structure was detected in Africa, and intriguingly the frequency of an allele associated with diminished sensibility to cyanogenic glycosides peaked in farmer groups, raising the question on whether in Africa the detection of compounds containing cyanogenic glycosides was less critical in agriculturalists than in hunter-gatherers.

Analysis of genetic distances across Africa, Europe and Asia further revealed unusual high levels of differentiation at *TAS2R16* and at the promoter of *TAS1R3*, patterns that *per se* fit the expectations of population models incorporating the action of positive selection, while contrarily at *TAS2R38* differentiation was too low comparatively to the average genome level, which in terms of worldwide population structure is predicted for *loci* evolving under balancing selection.

These preliminary results, in particular the observation of patterns of geographical distribution in some of the *TAS* genes difficult to reconcile with pure demographic explanations, prompted us to use an alternative approach to address selection at the taste genes.

Thus, taking advantage on the wealth of genomic data turned available by The 1000 Genomes Project Consortium (www.1000genomes.org), we have dissected the patterns of sequence diversity at *TAS1R1*, *TAS1R2*, *TAS1R3, TAS2R16* and *TAS2R38* in worldwide populations, to obtain insights on their evolution and on whether they were targets of selection in the recent human history (section 3.3).

One of our main findings was that the two genes from the *TAS2R* family, *TAS2R16* and *TAS2R38*, have evolved quite dissimilarly from the two genes from the other family we could fully analyze, *TAS1R1* and *TAS1R3*.

Clear signatures of balancing selection were detected at *TAS2R38* as captured in the low level of differentiation across continental populations, in the positive Tajima's *D* values often statistically significant under the best fit demographical models, and in the deep coalescence time of its sequences. *TAS2R16* revealed a more puzzling evolutionary history, still suggesting that inside Africa some selective pressure must have acted towards maintaining a large spectrum of phenotypic variability associated of this bitter taste receptor.

Within the *TAS1R* family, the two genes codifying for the heterodimer TAS1R1-TAS1R3, which mediates *umami* taste perception, evidenced many features expected for regions that have been under positive selection: haplotypes networks showing clear "star-like" structures; sequences coalescing in very recent times and systematic negative Tajima's *Ds* values. Though, whereas the signature detected in *TAS1R3* pinpointed to a recent selective sweep, the signature in *TAS1R1* rather suggested that two partial sweeps have occurred leading to the expansion of two different haplotypes.

These patterns of molecular evolution observed at different TAS genes raise the question on the forces that have acted modelling such trajectories.

Up to recently, proteins from the TAS1R and TAS2R families were essentially known due to their roles in the taste receptors located in the membrane of cells from the taste buds. However, new research is showing that different *TAS1R/TAS2R* genes besides expressed in organs of the gastrointestinal system are also expressed in several organs and tissues not implicated in food intake or digestion, such as the lung, the airway smooth muscle, the nose, the testis and including the hypothalamus (Campa, et al. 2012; Welcome, et al. 2015) Furthermore, there is growing evidence that taste genes play a broader role in human health, as is being inferred from the increasing reports on associations between allelic variants in taste receptor genes and body mass index, complex diseases such as cancer, alcohol consumption, smoking and nicotine dependence, longevity, for instance (Campa, et al. 2012).

In the future, with the availability of new information on the mechanism and process in which TAS1R/TAS2R are implied, hopefully we should be more able to determine the biological bases for the selective pressures that have acted over the encoding genes, leading also to further evolutionary explorations into the molecular nature of the adaptations.

In the past years, extensive research has been done aimed at identifying genetic adaptations in humans. As long as it is known today, classical selective sweeps were rare in humans, and the few that were identified occurred around *loci* with alleles associated with large phenotypic effects (Hernandez, et al. 2011; Fu and Akey 2013).

The recent progress in sequencing and genotyping methods, allows now to produce massive genome-wide data in humans populations, offering a key resource to explore selection in the genome. The general awareness and belief now, is that there is an urgent need for more theoretical modeling and better statistical methods to analyze selection, able to provide stronger resolution to identify adaptive signatures especially in polygenic phenotypes, for which account many alleles with small phenotypic effects (Wollstein and Stephan 2015). Another major challenge for researches will be to develop functional methods and approaches to demonstrate that putative selective variants have phenotypic effects that are import for organismal fitness, which ultimately presents the validation of the identified genomic signatures of selection (Vitti, et al. 2013).

Eventually then, it will be possible to obtain better insights on to the two-fold challenge in human population genetics which is to explain how population history accounts to current patterns of genetic diversity, and at the same time to understand the genetic basis of phenotypic adaptations in humans (Balaresque, et al. 2007), as well as the extent to which they were triggered by the changes that accompanied the dispersion of humans out of Africa, or latter in the Neolithic, the transition to new subsistence economies.

CHAPTER 5. REFERENCES

Achaz G. 2008. Testing for neutrality in samples with sequencing errors. Genetics 179:1409-1424.

Akey JM. 2009. Constructing genomic maps of positive selection in humans: where do we go from here? Genome Res 19:711-722.

Akey JM, Zhang G, Zhang K, Jin L, Shriver MD. 2002. Interrogating a high-density SNP map for signatures of natural selection. Genome Res 12:1805-1814.

Albrechtsen A, Moltke I, Nielsen R. 2010. Natural selection and the distribution of identity-by-descent in the human genome. Genetics 186:295-308.

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. PLoS Genet 8:e1003110.

Alves C, Gusmao L, Amorim A. 2001. STR data (AmpFlSTR Profiler Plus and GenePrint CTTv) from Mozambique. Forensic Sci Int 119:131-133.

Arbiza L, Dopazo J, Dopazo H. 2006. Positive selection, relaxation, and acceleration in the evolution of the human and chimp genome. PLoS Comput Biol 2:e38.

Ardlie KG, Kruglyak L, Seielstad M. 2002. Patterns of linkage disequilibrium in the human genome. Nat Rev Genet 3:299-309.

Attah-Poku A. 1998. African Ethnicity: History, Conflict Management, Resolution and Prevention: University Press of America.

Bachmanov AA, Beauchamp GK. 2007. Taste receptor genes. Annu Rev Nutr 27:389-414.

Bachmanov AA, Bosak NP, Floriano WB, Inoue M, Li X, Lin C, Murovets VO, Reed DR, Zolotarev VA, Beauchamp GK. 2011. Genetics of sweet taste preferences. 26:286-294.

Bachmanov AA, Bosak NP, Lin C, Matsumoto I, Ohmoto M, Reed DR, Nelson TM. 2014. Genetics of taste receptors. Curr Pharm Des 20:2669-2683.

Bachmanov AA, Reed DR, Li X, Beauchamp GK. 2002. Genetics of sweet taste preferences. Pure Appl Chem 74:1135-1140.

Bae J-W, Lee H-J, Oh S-K, Kim S-Y, Kim U-K. 2010. Genetic variation of umami taste genes in Koreans. Genes & Genomics 32:111-113.

Balaresque PL, Ballereau SJ, Jobling MA. 2007. Challenges in human genetic diversity: demographic history and adaptation. Hum Mol Genet 16 Spec No. 2:R134-139.

Ballinger SW, Schurr TG, Torroni A, Gan YY, Hodge JA, Hassan K, Chen KH, Wallace DC. 1992. Southeast Asian mitochondrial DNA analysis reveals genetic continuity of ancient mongoloid migrations. Genetics 130:139-152.

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

Barbieri C, Vicente M, Oliveira S, Bostoen K, Rocha J, Stoneking M, Pakendorf B. 2014. Migration and Interaction in a Contact Zone: mtDNA Variation among Bantu-Speakers in Southern Africa. PLoS ONE 9:e99117.

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.

Barreiro LB, Quintana-Murci L. 2010. From evolutionary genetics to human immunology: how selection shapes host defence genes. Nat Rev Genet 11:17-30.

Barrett JC, Fry B, Maller J, Daly MJ. 2005. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 21:263-265.

Batini C, Lopes J, Behar DM, Calafell F, Jorde LB, van der Veen L, Quintana-Murci L, Spedini G, Destro-Bisol G, Comas D. 2011. Insights into the demographic history of African Pygmies from complete mitochondrial genomes. Mol Biol Evol 28:1099-1110.

Beaumont MA, Nichols RA. 1996. Evaluating loci for use in the genetic analysis of population structure. P Roy Soc B-Biol Sci 263:1619-1626.

Behrens M, Gunn HC, Ramos PC, Meyerhof W, Wooding SP. 2013. Genetic, functional, and phenotypic diversity in TAS2R38-mediated bitter taste perception. Chem Senses 38:475-484.

Beleza S, Alves C, Reis F, Amorim A, Carracedo A, Gusmao L. 2004. 17 STR data (AmpF/STR Identifiler and Powerplex 16 System) from Cabinda (Angola). Forensic Sci Int 141:193-196.

Bersaglieri T, Sabeti PC, Patterson N, Vanderploeg T, Schaffner SF, Drake JA, Rhodes M, Reich DE, Hirschhorn JN. 2004. Genetic signatures of strong recent positive selection at the lactase gene. Am J Hum Genet 74:1111-1120.

Berton A, Sebban-Kreuzer C, Crenon I. 2007. Role of the structural domains in the functional properties of pancreatic lipase-related protein 2. FEBS J 274:6011-6023.

Bhatia G, Patterson N, Sankararaman S, Price AL. 2013. Estimating and interpreting FST: the impact of rare variants. Genome Res 23:1514-1521.

Biswas S, Akey JM. 2006. Genomic insights into positive selection. Trends Genet 22:437-446.

Blakeslee AF. 1932. Genetics of Sensory Thresholds: Taste for Phenyl Thio Carbamide. Proc Natl Acad Sci U S A 18:120-130.

Blum MG, Jakobsson M. 2011. Deep divergences of human gene trees and models of human origins. Mol Biol Evol 28:889-898.

Boateng EA. 1978. A Political Geography of Africa: Cambridge University Press.

Bowcock AM, Ruiz-Linares A, Tomfohrde J, Minch E, Kidd JR, Cavalli-Sforza LL. 1994. High resolution of human evolutionary trees with polymorphic microsatellites. Nature 368:455-457.

Boyko AR, Williamson SH, Indap AR, Degenhardt JD, Hernandez RD, Lohmueller KE, Adams MD, Schmidt S, Sninsky JJ, Sunyaev SR, et al. 2008. Assessing the evolutionary impact of amino acid mutations in the human genome. PLoS Genet 4:e1000083.

Breslin PA. 2013. An evolutionary perspective on food and human taste. Curr Biol 23:R409-418.

Breton G, Schlebusch CM, Lombard M, Sjodin P, Soodyall H, Jakobsson M. 2014. Lactase persistence alleles reveal partial East african ancestry of southern african Khoe pastoralists. Curr Biol 24:852-858.

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

Bufe B, Breslin PA, Kuhn C, Reed DR, Tharp CD, Slack JP, Kim UK, Drayna D, Meyerhof W. 2005. The molecular basis of individual differences in phenylthiocarbamide and propylthiouracil bitterness perception. Curr Biol 15:322-327.

Bufe B, Hofmann T, Krautwurst D, Raguse JD, Meyerhof W. 2002. The human TAS2R16 receptor mediates bitter taste in response to beta-glucopyranosides. Nat Genet 32:397-401.

Bustamante CD, Fledel-Alon A, Williamson S, Nielsen R, Hubisz MT, Glanowski S, Tanenbaum DM, White TJ, Sninsky JJ, Hernandez RD, et al. 2005. Natural selection on protein-coding genes in the human genome. Nature 437:1153-1157.

Caldwell EF, Mayor LR, Thomas MG, Danpure CJ. 2004. Diet and the frequency of the alanine:glyoxylate aminotransferase Pro11Leu polymorphism in different human populations. Hum Genet 115:504-509.

Calo C, Padiglia A, Zonza A, Corrias L, Contu P, Tepper BJ, Barbarossa IT. 2011. Polymorphisms in TAS2R38 and the taste bud trophic factor, gustin gene co-operate in modulating PROP taste phenotype. Physiol Behav 104:1065-1071.

Campa D, De Rango F, Carrai M, Crocco P, Montesanto A, Canzian F, Rose G, Rizzato C, Passarino G, Barale R. 2012. Bitter taste receptor polymorphisms and human aging. PLoS One 7:e45232.

Campbell MC, Ranciaro A, Froment A, Hirbo J, Omar S, Bodo JM, Nyambo T, Lema G, Zinshteyn D, Drayna D, et al. 2012. Evolution of functionally diverse alleles associated with PTC bitter taste sensitivity in Africa. Mol Biol Evol 29:1141-1153.

Campbell MC, Ranciaro A, Zinshteyn D, Rawlings-Goss R, Hirbo J, Thompson S, Woldemeskel D, Froment A, Rucker JB, Omar SA, et al. 2013. Origin and Differential Selection of Allelic Variation at TAS2R16 Associated with Salicin Bitter Taste Sensitivity in Africa. Mol Biol Evol.

Campbell MC, Tishkoff SA. 2010. The evolution of human genetic and phenotypic variation in Africa. Curr Biol 20:R166-173.

Chandrashekar J, Hoon MA, Ryba NJ, Zuker CS. 2006. The receptors and cells for mammalian taste. Nature 444:288-294.

Chaudhari N, Roper SD. 2010. The cell biology of taste. J Cell Biol 190:285-296.

Chen QY, Alarcon S, Tharp A, Ahmed OM, Estrella NL, Greene TA, Rucker J, Breslin PA. 2009. Perceptual variation in umami taste and polymorphisms in TAS1R taste receptor genes. Am J Clin Nutr 90:770S-779S.

Chu JY, Huang W, Kuang SQ, Wang JM, Xu JJ, Chu ZT, Yang ZQ, Lin KQ, Li P, Wu M, et al. 1998. Genetic relationship of populations in China. Proc Natl Acad Sci U S A 95:11763-11768.

Clark AG, Glanowski S, Nielsen R, Thomas PD, Kejariwal A, Todd MA, Tanenbaum DM, Civello D, Lu F, Murphy B, et al. 2003. Inferring nonneutral evolution from human-chimp-mouse orthologous gene trios. Science 302:1960-1963.

Coop G, Griffiths RC. 2004. Ancestral inference on gene trees under selection. Theor Popul Biol 66:219-232.

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.

Coop G, Witonsky D, Di Rienzo A, Pritchard JK. 2010. Using environmental correlations to identify loci underlying local adaptation. Genetics 185:1411-1423.

Cordain L, Eaton SB, Miller JB, Mann N, Hill K. 2002. The paradoxical nature of hunter-gatherer diets: meat-based, yet non-atherogenic. Eur J Clin Nutr 56 Suppl 1:S42-52.

Danpure CJ. 2006. Primary hyperoxaluria type 1: AGT mistargeting highlights the fundamental differences between the peroxisomal and mitochondrial protein import pathways. Biochim Biophys Acta 1763:1776-1784.

Danpure CJ. 1997. Variable peroxisomal and mitochondrial targeting of alanine: Glyoxylate aminotransferase in mammalian evolution and disease. Bioessays 19:317-326.

Daub JT, Hofer T, Cutivet E, Dupanloup I, Quintana-Murci L, Robinson-Rechavi M, Excoffier L. 2013. Evidence for polygenic adaptation to pathogens in the human genome. Mol Biol Evol 30:1544-1558.

De Caro J, Eydoux C, Cherif S, Lebrun R, Gargouri Y, Carriere F, De Caro A. 2008. Occurrence of pancreatic lipase-related protein-2 in various species and its relationship with herbivore diet. Comp Biochem Physiol B Biochem Mol Biol 150:1-9.

de Filippo C, Bostoen K, Stoneking M, Pakendorf B. 2012. Bringing together linguistic and genetic evidence to test the Bantu expansion. Proc Biol Sci 279:3256-3263.

de Magalhaes JP, Matsuda A. 2012. Genome-wide patterns of genetic distances reveal candidate Loci contributing to human population-specific traits. Ann Hum Genet 76:142-158.

Delaneau O, Marchini J, Zagury JF. 2012. A linear complexity phasing method for thousands of genomes. Nat Methods 9:179-181.

Diamond J. 2002. Evolution, consequences and future of plant and animal domestication. Nature 418:700-707.

Diamond J. 1997. GUNS, GERMS AND STEELS: THE FATES OF HUMAN SOCIETIES. New York, London: W. W. Norton & Company.

Doty RL. 2015. Handbook of Olfaction and Gustation.

Eap CB, Bochud M, Elston RC, Bovet P, Maillard MP, Nussberger J, Schild L, Shamlaye C, Burnier M. 2007. CYP3A5 and ABCB1 Genes Influence Blood Pressure and Response to Treatment, and Their Effect Is Modified by Salt. Hypertension 49:1007-1014.

Eaton SB, Eaton SB, 3rd. 2000. Paleolithic vs. modern diets--selected pathophysiological implications. Eur J Nutr 39:67-70.

Elhaik E. 2012. Empirical distributions of F(ST) from large-scale human polymorphism data. PLoS One 7:e49837.

Elhassan N, Gebremeskel EI, Elnour MA, Isabirye D, Okello J, Hussien A, Kwiatksowski D, Hirbo J, Tishkoff S, Ibrahim ME. 2014. The episode of genetic drift defining the migration of humans out of Africa is derived from a large east African population size. PLoS One 9:e97674.

Enard D, Messer PW, Petrov DA. 2014. Genome-wide signals of positive selection in human evolution. Genome Res 24:885-895.

Eny KM, Wolever TM, Corey PN, El-Sohemy A. 2010. Genetic variation in TAS1R2 (Ile191Val) is associated with consumption of sugars in overweight and obese individuals in 2 distinct populations. Am J Clin Nutr 92:1501-1510.

Epstein EL, Kole R. 1998. The Language of African Literature: Africa World Press.

Excoffier L, Hofer T, Foll M. 2009. Detecting loci under selection in a hierarchically structured population. Heredity (Edinb) 103:285-298.

Excoffier L, Lischer HE. 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol Ecol Resour 10:564-567.

Excoffier LH, T.; Foll, M. 2009. Detecting loci under selection in a hierarchically structured population. Heredity 103.

Fagny M, Patin E, Enard D, Barreiro LB, Quintana-Murci L, Laval G. 2014. Exploring the Occurrence of Classic Selective Sweeps in Humans Using Whole-genome Sequencing Datasets. Mol Biol Evol.

Fay JC, Wu CI. 2000. Hitchhiking under positive Darwinian selection. Genetics 155:1405-1413.

Fernstrom JD, Munger SD, Sclafani A, de Araujo IE, Roberts A, Molinary S. 2012. Mechanisms for sweetness. J Nutr 142:1134S-1141S.

Fox AL. 1932. The Relationship between Chemical Constitution and Taste. Proc Natl Acad Sci U S A 18:115-120.

Fregel R, Cabrera V, Larruga JM, Abu-Amero KK, Gonzalez AM. 2015. Carriers of Mitochondrial DNA Macrohaplogroup N Lineages Reached Australia around 50,000 Years Ago following a Northern Asian Route. PLoS One 10:e0129839.

Fu W, Akey JM. 2013. Selection and adaptation in the human genome. Annu Rev Genomics Hum Genet 14:467-489.

Fu YX, Li WH. 1993. Statistical tests of neutrality of mutations. Genetics 133:693-709.

Fumagalli M, Sironi M, Pozzoli U, Ferrer-Admetlla A, Pattini L, Nielsen R. 2011. Signatures of environmental genetic adaptation pinpoint pathogens as the main selective pressure through human evolution. PLoS Genet 7:e1002355.

Fushan AA, Simons CT, Slack JP, Manichaikul A, Drayna D. 2009. Allelic polymorphism within the TAS1R3 promoter is associated with human taste sensitivity to sucrose. Curr Biol 19:1288-1293.

Gann LH, Duignan P. 1972. Africa and the World: An Introduction to the History of Sub-Saharan Africa from Antiquity to 1840: University Press of America.

Garcia-Bailo B, Toguri C, Eny KM, El-Sohemy A. 2009. Genetic variation in taste and its influence on food selection. OMICS 13:69-80.

Gattepaille LM, Jakobsson M. 2012. Combining markers into haplotypes can improve population structure inference. Genetics 190:159-174.

Gaya-Vidal M, Alba MM. 2014. Uncovering adaptive evolution in the human lineage. BMC Genomics 15:599.

Gianola D, Simianer H, Qanbari S. 2010. A two-step method for detecting selection signatures using genetic markers. Genet Res (Camb) 92:141-155.

Gibbons A. 2011. Human evolution. African data bolster new view of modern human origins. Science 334:167.

Gleibermann L. 1973. Blood pressure and dietary salt in human populations. Ecology of Food and Nutrition 2:143-156.

Gomes V, Sanchez-Diz P, Alves C, Gomes I, Amorim A, Carracedo A, Gusmao L. 2009. Population data defined by 15 autosomal STR loci in Karamoja population (Uganda) using AmpF/STR Identifiler kit. Forensic Sci Int Genet 3:e55-58.

Gomes V, Sanchez-Diz P, Amorim A, Carracedo A, Gusmao L. 2010. Digging deeper into East African human Y chromosome lineages. Hum Genet 127:603-613.

Gorovic N, Afzal S, Tjonneland A, Overvad K, Vogel U, Albrechtsen C, Poulsen HE. 2011. Genetic variation in the hTAS2R38 taste receptor and brassica vegetable intake. Scand J Clin Lab Invest 71:274-279.

Gravel S, Henn BM, Gutenkunst RN, Indap AR, Marth GT, Clark AG, Yu F, Gibbs RA, Genomes P, Bustamante CD. 2011. Demographic history and rare allele sharing among human populations. Proc Natl Acad Sci U S A 108:11983-11988.

Greenberg JH, Denning KM, Kemmer S. 1990. On Language: Selected Writings of Joseph H. Greenberg: Stanford University Press.

Griffiths RC, Tavare S. 1994. Sampling theory for neutral alleles in a varying environment. Philos Trans R Soc Lond B Biol Sci 344:403-410.

Grossman SR, Andersen KG, Shlyakhter I, Tabrizi S, Winnicki S, Yen A, Park DJ, Griesemer D, Karlsson EK, Wong SH, et al. 2013. Identifying recent adaptations in large-scale genomic data. Cell 152:703-713.

Gunz P, Bookstein FL, Mitteroecker P, Stadlmayr A, Seidler H, Weber GW. 2009. Early modern human diversity suggests subdivided population structure and a complex out-of-Africa scenario. Proc Natl Acad Sci U S A 106:6094-6098.

Han E, Sinsheimer JS, Novembre J. 2014. Characterizing bias in population genetic inferences from low-coverage sequencing data. Mol Biol Evol 31:723-735.

Hancock AM, Witonsky DB, Alkorta-Aranburu G, Beall CM, Gebremedhin A, Sukernik R, Utermann G, Pritchard JK, Coop G, Di Rienzo A. 2011. Adaptations to climate-mediated selective pressures in humans. PLoS Genet 7:e1001375.

Hancock AM, Witonsky DB, Ehler E, Alkorta-Aranburu G, Beall C, Gebremedhin A, Sukernik R, Utermann G, Pritchard J, Coop G, et al. 2010. Colloquium paper: human adaptations to diet, subsistence, and ecoregion are due to subtle shifts in allele frequency. Proc Natl Acad Sci U S A 107 Suppl 2:8924-8930.

Hancock AM, Witonsky DB, Gordon AS, Eshel G, Pritchard JK, Coop G, Di Rienzo A. 2008. Adaptations to climate in candidate genes for common metabolic disorders. PLoS Genet 4:e32.

Hanotte O, Bradley DG, Ochieng JW, Verjee Y, Hill EW, Rege JE. 2002. African pastoralism: genetic imprints of origins and migrations. Science 296:336-339.

Harihara S, Saitou N, Hirai M, Gojobori T, Park KS, Misawa S, Ellepola SB, Ishida T, Omoto K. 1988. Mitochondrial DNA polymorphism among five Asian populations. Am J Hum Genet 43:134-143.

Haviland W, Prins H, McBride B, Walrath D. 2010. Cultural Anthropology: The Human Challenge: Cengage Learning.

Hein DW, Doll MA. 2012. Accuracy of various human NAT2 SNP genotyping panels to infer rapid, intermediate and slow acetylator phenotypes. Pharmacogenomics 13:31-41.

Heine B, Nurse D. 2000. African Languages: An Introduction: Cambridge University Press.

Henn BM, Botigue LR, Gravel S, Wang W, Brisbin A, Byrnes JK, Fadhlaoui-Zid K, Zalloua PA, Moreno-Estrada A, Bertranpetit J, et al. 2012. Genomic ancestry of North Africans supports back-to-Africa migrations. PLoS Genet 8:e1002397.

Henn BM, Gignoux CR, Jobin M, Granka JM, Macpherson JM, Kidd JM, Rodriguez-Botigue L, Ramachandran S, Hon L, Brisbin A, et al. 2011. Hunter-gatherer genomic diversity suggests a southern African origin for modern humans. Proc Natl Acad Sci U S A 108:5154-5162.

Hernandez RD, Kelley JL, Elyashiv E, Melton SC, Auton A, McVean G, Genomes P, Sella G, Przeworski M. 2011. Classic selective sweeps were rare in recent human evolution. Science 331:920-924.

Heyer E, Quintana-Murci L. 2009. Evolutionary genetics as a tool to target genes involved in phenotypes of medical relevance. Evol Appl 2:71-80.

Hindorff LA, Junkins HA, Hall PN, Mehta JP, Manolio TA. 2010. A catalog of published genome-wide association studies.http://www.genome.gov/ gwastudies.

Hinrichs AL, Wang JC, Bufe B, Kwon JM, Budde J, Allen R, Bertelsen S, Evans W, Dick D, Rice J, et al. 2006. Functional variant in a bitter-taste receptor (hTAS2R16) influences risk of alcohol dependence. Am J Hum Genet 78:103-111.

Hofer T, Foll M, Excoffier L. 2012. Evolutionary forces shaping genomic islands of population differentiation in humans. BMC Genomics 13:107.

Holsinger KE, Lewis PO, Dey DK. 2002. A Bayesian approach to inferring population structure from dominant markers. Mol Ecol 11:1157-1164.

Hombert JM, Hyman LM. 1999. Bantu Historical Linguistics: Theoretical and Empirical Perspectives: CSLI Publications.

Hudson RR. 2002. Generating samples under a Wright-Fisher neutral model of genetic variation. Bioinformatics 18:337-338.

Hwang LD, Zhu G, Breslin PA, Reed DR, Martin NG, Wright MJ. 2015. A Common Genetic Influence on Human Intensity Ratings of Sugars and High-Potency Sweeteners. Twin Res Hum Genet:1-7.

Ikeda K. 2002. New seasonings. Chem Senses 27:847-849.

Imai H, Suzuki N, Ishimaru Y, Sakurai T, Yin L, Pan W, Abe K, Misaka T, Hirai H. 2012. Functional diversity of bitter taste receptor TAS2R16 in primates. Biol Lett 8:652-656.

International HapMap C, Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, Gibbs RA, Belmont JW, Boudreau A, Hardenbol P, et al. 2007. A second generation human haplotype map of over 3.1 million SNPs. Nature 449:851-861.

Jin L, Baskett ML, Cavalli-Sforza LL, Zhivotovsky LA, Feldman MW, Rosenberg NA. 2000. Microsatellite evolution in modern humans: a comparison of two data sets from the same populations. Ann Hum Genet 64:117-134.

Jin L, Su B. 2000. Natives or immigrants: modern human origin in east Asia. Nat Rev Genet 1:126-133.

Jin L, Underhill PA, Doctor V, Davis RW, Shen P, Cavalli-Sforza LL, Oefner PJ. 1999. Distribution of haplotypes from a chromosome 21 region distinguishes multiple prehistoric human migrations. Proc Natl Acad Sci U S A 96:3796-3800.

John Haley CR, Rebecca Stefoff, Joseph Ziegler. 2002. Africa: An Encyclopedia for Students. In: Middleton J, editor. Africa: An Encyclopedia for Students: Charles Scribner's Sons.

Johnson MJ, Wallace DC, Ferris SD, Rattazzi MC, Cavalli-Sforza LL. 1983. Radiation of human mitochondria DNA types analyzed by restriction endonuclease cleavage patterns. J Mol Evol 19:255-271.

Jones BL, Raga TO, Liebert A, Zmarz P, Bekele E, Danielsen ET, Olsen AKg, Bradman N, Troelsen JT, Swallow DM. 2013. Diversity of Lactase Persistence Alleles in Ethiopia: Signature of a Soft Selective Sweep. American Journal of Human Genetics 93:538-544.

Kare MRB, G.K. 1984. Taste, smell, and hearing. In: Swenson MJ, editor. Dukes' physiology of domestic animals. Ithaca (NY): Cornell University Press. p. 742–760.

Karlsson EK, Kwiatkowski DP, Sabeti PC. 2014. Natural selection and infectious disease in human populations. Nat Rev Genet 15:379-393.

Ke Y, Su B, Song X, Lu D, Chen L, Li H, Qi C, Marzuki S, Deka R, Underhill P, et al. 2001. African origin of modern humans in East Asia: a tale of 12,000 Y chromosomes. Science 292:1151-1153.

Keese A. 2010. Ethnicity and the Long-term Perspective: The African Experience: Peter Lang.

Khan N, Pande V, Das A. 2013. NAT2 sequence polymorphisms and acetylation profiles in Indians. Pharmacogenomics 14:289-303.

Khan R, Khan BS. 2010. Diet, disease and pigment variation in humans. Med Hypotheses 75:363-367.

Kim UK, Jorgenson E, Coon H, Leppert M, Risch N, Drayna D. 2003. Positional cloning of the human quantitative trait locus underlying taste sensitivity to phenylthiocarbamide. Science 299:1221-1225.

Kim UK, Wooding S, Riaz N, Jorde LB, Drayna D. 2006. Variation in the human TAS1R taste receptor genes. Chem Senses 31:599-611.

Knight A, Underhill PA, Mortensen HM, Zhivotovsky LA, Lin AA, Henn BM, Louis D, Ruhlen M, Mountain JL. 2003. African Y chromosome and mtDNA divergence provides insight into the history of click languages. Curr Biol 13:464-473.

Ko WY, Kaercher KA, Giombini E, Marcatili P, Froment A, Ibrahim M, Lema G, Nyambo TB, Omar SA, Wambebe C, et al. 2011. Effects of natural selection and gene conversion on the evolution of human glycophorins coding for MNS blood polymorphisms in malaria-endemic African populations. Am J Hum Genet 88:741-754.

Kong A, Gudbjartsson DF, Sainz J, Jonsdottir GM, Gudjonsson SA, Richardsson B, Sigurdardottir S, Barnard J, Hallbeck B, Masson G, et al. 2002. A high-resolution recombination map of the human genome. Nat Genet 31:241-247.

Kosuge K, Chuang AI, Uematsu S, Tan KP, Ohashi K, Ko BC, Ito S. 2007. Discovery of Osmo-sensitive Transcriptional Regulation of Human Cytochrome P450 3As (CYP3As) by the Tonicity-Responsive Enhancer Binding Protein (TonEBP/ NFAT5). Molecular Pharmacology.

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. 1994. The Multiregional Model of Modern Human Origins - a Reassessment of Its Morphological Basis. J Hum Evol 26:23-56.

Lamba JK, Lin YS, Schuetz EG, Thummel KE. 2012. Genetic contribution to variable human CYP3A-mediated metabolism. Adv Drug Deliv Rev.

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.

Lee RJ, Cohen NA. 2015. Role of the bitter taste receptor T2R38 in upper respiratory infection and chronic rhinosinusitis. Curr Opin Allergy Clin Immunol 15:14-20.

Lee RJ, Kofonow JM, Rosen PL, Siebert AP, Chen B, Doghramji L, Xiong G, Adappa ND, Palmer JN, Kennedy DW, et al. 2014. Bitter and sweet taste receptors regulate human upper respiratory innate immunity. J Clin Invest 124:1393-1405.

Lewontin KC, Kojima K. 1960. The evolutionary dynamics of complex polymorphisms. Evolution 14:458-472.

Lewontin RC. 1964. The Interaction of Selection and Linkage. I. General Considerations; Heterotic Models. Genetics 49:49-67.

Lewontin RC, Krakauer J. 1973. Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. Genetics 74:175-195.

Li H, Pakstis AJ, Kidd JR, Kidd KK. 2011. Selection on the human bitter taste gene, TAS2R16, in Eurasian populations. Hum Biol 83:363-377.

Li X, Staszewski L, Xu H, Durick K, Zoller M, Adler E. 2002. Human receptors for sweet and umami taste. Proc Natl Acad Sci U S A 99:4692-4696.

Li Y, Wu DD, Boyko AR, Wang GD, Wu SF, Irwin DM, Zhang YP. 2014. Population Variation Revealed High-Altitude Adaptation of Tibetan Mastiffs. Mol Biol Evol.

Lischer HE, Excoffier L. 2012. PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs. Bioinformatics 28:298-299.

Liu X, Ong RT-H, Pillai EN, Elzein AM, Small KS, Clark TG, Kwiatkowski DP, Teo Y-Y. 2013. Detecting and Characterizing Genomic Signatures of Positive Selection in Global Populations. American Journal of Human Genetics 92:866-881.

Lowe ME. 2002. The triglyceride lipases of the pancreas. The Journal of Lipid Research 43:2007-2016.

Luca F, Bubba G, Basile M, Brdicka R, Michalodimitrakis E, Richards O, Vershubsky G, Quintana-Murci L, Kozlov AI, Novelleto A. 2008. Multiple Advantageous Amino Acid Variants in the NAT2 Gene in Human Populations. PLoS One 3.

Luca F, Perry GH, Di Rienzo A. 2010. Evolutionary adaptations to dietary changes. Annu Rev Nutr 30:291-314.

Lucotte G, Gerard N, Krishnamoorthy R, David F, Aouizerate A, Galzot P. 1994. Reduced variability in Y-chromosome-specific haplotypes for some Central African populations. Hum Biol 66:519-526.

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.

Macholdt E, Lede V, Barbieri C, Mpoloka SW, Chen H, Slatkin M, Pakendorf B, Stoneking M. 2014. Tracing pastoralist migrations to southern Africa with lactase persistence alleles. Curr Biol 24:875-879.

Macholdt E, Slatkin M, Pakendorf B, Stoneking M. 2015. New insights into the history of the C-14010 lactase persistence variant in Eastern and Southern Africa. Am J Phys Anthropol 156:661-664.

Magalon H, Patin E, Austerlitz F, Hegay T, Aldashev A, Quintana-Murci L, Heyer E. 2008. Population genetic diversity of the NAT2 gene supports a role of acetylation in human adaptation to farming in Central Asia. Eur J Hum Genet 16:243-251.

Marth GT, Czabarka E, Murvai J, Sherry ST. 2004. The allele frequency spectrum in genome-wide human variation data reveals signals of differential demographic history in three large world populations. Genetics 166:351-372.

Maruki T, Lynch M. 2014. Genome-Wide Estimation of Linkage Disequilibrium from Population-Level High-Throughput Sequencing Data. Genetics.

McVean GA, Myers SR, Hunt S, Deloukas P, Bentley DR, Donnelly P. 2004. The fine-scale structure of recombination rate variation in the human genome. Science 304:581-584.

Meirmans PG, Hedrick PW. 2011. Assessing population structure: F(ST) and related measures. Mol Ecol Resour 11:5-18.

Mellars P. 2006. Why did modern human populations disperse from Africa ca. 60,000 years ago? A new model. Proc Natl Acad Sci U S A 103:9381-9386.

Mengel MA, Delrieu I, Heyerdahl L, Gessner BD. 2014. Cholera outbreaks in Africa. Curr Top Microbiol Immunol 379:117-144.

Metspalu M, Kivisild T, Metspalu E, Parik J, Hudjashov G, Kaldma K, Serk P, Karmin M, Behar DM, Gilbert MT, et al. 2004. Most of the extant mtDNA boundaries in south and southwest Asia were likely shaped during the initial settlement of Eurasia by anatomically modern humans. BMC Genet 5:26.

Neel JV. 1962. Diabetes mellitus: a "thrifty" genotype rendered detrimental by "progress"? Am J Hum Genet 14:353-362.

Nei M, Li WH. 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc Natl Acad Sci U S A 76:5269-5273.

Nelson G, Chandrashekar J, Hoon MA, Feng L, Zhao G, Ryba NJ, Zuker CS. 2002. An amino-acid taste receptor. Nature 416:199-202.

Nielsen R, Bustamante C, Clark AG, Glanowski S, Sackton TB, Hubisz MJ, Fledel-Alon A, Tanenbaum DM, Civello D, White TJ, et al. 2005. A scan for positively selected genes in the genomes of humans and chimpanzees. PLoS Biol 3:e170.

Oota H, Pakendorf B, Weiss G, von Haeseler A, Pookajorn S, Settheetham-Ishida W, Tiwawech D, Ishida T, Stoneking M. 2005. Recent origin and cultural reversion of a hunter-gatherer group. PLoS Biol 3:e71.

Oppenheimer S. 2012. Out-of-Africa, the peopling of continents and islands: tracing uniparental gene trees across the map. Philos Trans R Soc Lond B Biol Sci 367:770-784.

Paabo S. 1995. The Y chromosome and the origin of all of us (men). Science 268:1141-1142.

Pagani L, Kivisild T, Tarekegn A, Ekong R, Plaster C, Gallego Romero I, Ayub Q, Mehdi SQ, Thomas MG, Luiselli D, et al. 2012. Ethiopian genetic diversity reveals linguistic stratification and complex influences on the Ethiopian gene pool. Am J Hum Genet 91:83-96.

Pakendorf B, Bostoen, K. and de Filippo, C. 2011. Molecular Perspectives on the Bantu Expansion: A Synthesis. In: Brill, editor. Language Dynamics and Change. p. 50-88.

Panter-Brick C, Layton R, Rowley-Conwy P. 2001. Hunter-Gatherers: An Interdisciplinary Perspective: Cambridge University Press.

Papiha SS, Mastana SS, Roberts DF, Onyemelukwe GC, Bhattacharya SS. 1991. Population variation in molecular polymorphisms of the short arm of the human X chromosome. Am J Phys Anthropol 85:329-334.

Patillon B, Luisi P, Poloni ES, Boukouvala S, Darlu P, Genin E, Sabbagh A. 2014. A Homogenizing Process of Selection Has Maintained an "Ultra-Slow" Acetylation NAT2 Variant in Humans. Hum Biol 86:185-214.

Patin E, Barreiro LB, Sabeti PC, Austerlitz F, Luca F, Sajantila A, Behar DM, Semino O, Sakuntabhai A, Guiso N, et al. 2006. Deciphering the Ancient and Complex Evolutionary History of Human Arylamine N-Acetyltransferase Genes. American Journal of Human Genetics 78.

Patin E, Quintana-Murci L. 2008. Demeter's legacy: rapid changes to our genome imposed by diet. Trends Ecol Evol 23:56-59.

Patterson N, Richter DJ, Gnerre S, Lander ES, Reich D. 2006. Genetic evidence for complex speciation of humans and chimpanzees. Nature 441:1103-1108.

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.

Perry GH, Kistler L, Kelaita MA, Sams AJ. 2015. Insights into hominin phenotypic and dietary evolution from ancient DNA sequence data. J Hum Evol 79:55-63.

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.

Programme UNHS, Africa UNECf. 2008. The State of African Cities 2008: A Framework for Addressing Urban Challenges in Africa: UN-HABITAT.

Programme UNHS, Programme UNE. 2010. The State of African Cities 2010: Governance, Inequality and Urban Land Markets: UN-HABITAT.

Przeworski M, Coop G, Wall JD. 2005. The signature of positive selection on standing genetic variation. Evolution 59:2312-2323.

Przeworski M, Hudson RR, Di Rienzo A. 2000. Adjusting the focus on human variation. Trends Genet 16:296-302.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, et al. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81:559-575.

Quintana-Murci L, Clark AG. 2013. Population genetic tools for dissecting innate immunity in humans. Nat Rev Immunol 13:280-293.

Quintana-Murci L, Semino O, Bandelt HJ, 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.

Rajeevan H, Soundararajan U, Kidd JR, Pakstis AJ, Kidd KK. 2012. ALFRED: an allele frequency resource for research and teaching. Nucleic Acids Res 40:D1010-D1015.

Raliou M, Grauso M, Hoffmann B, Schlegel-Le-Poupon C, Nespoulous C, Debat H, Belloir C, Wiencis A, Sigoillot M, Bano SP, et al. 2011. Human genetic polymorphisms in T1R1 and T1R3 taste receptor subunits affect their function. Chem Senses 36:527-537.

Raliou M, Wiencis A, Pillias AM, Planchais A, Eloit C, Boucher Y, Trotier D, Montmayeur JP, Faurion A. 2009. Nonsynonymous single nucleotide polymorphisms in human tas1r1, tas1r3, and mGluR1 and individual taste sensitivity to glutamate. Am J Clin Nutr 90:789S-799S.

Ralph P, Coop G. 2013. The geography of recent genetic ancestry across Europe. PLoS Biol 11:e1001555.

Ranciaro A, Campbell MC, Hirbo JB, Ko WY, Froment A, Anagnostou P, Kotze MJ, Ibrahim M, Nyambo T, Omar SA, et al. 2014. Genetic origins of lactase persistence and the spread of pastoralism in Africa. Am J Hum Genet 94:496-510.

Rawal S, Hayes JE, Wallace MR, Bartoshuk LM, Duffy VB. 2013. Do polymorphisms in the TAS1R1 gene contribute to broader differences in human taste intensity? Chem Senses 38:719-728.

Reyes-Centeno H, Ghirotto S, Detroit F, Grimaud-Herve D, Barbujani G, Harvati K. 2014. Genomic and cranial phenotype data support multiple modern human dispersals from Africa and a southern route into Asia. Proc Natl Acad Sci U S A 111:7248-7253.

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.

Rozas J. 2009. DNA sequence polymorphism analysis using DnaSP. Methods Mol Biol 537:337-350.

Russell T, Silva F, Steele J. 2014. Modelling the spread of farming in the Bantu-speaking regions of Africa: an archaeology-based phylogeography. PLoS One 9:e87854.

Sabbagh A, Darlu P, Crouau-Roy B, Poloni ES. 2011. Arylamine N-acetyltransferase 2 (NAT2) genetic diversity and traditional subsistence: a worldwide population survey. PLoS One 6:e18507.

Sabbagh A, Langaney A, Darlu P, Gerard N, Krishnamoorthy R, Poloni ES. 2008. Worldwide distribution of NAT2 diversity: implications for NAT2 evolutionary history. BMC Genet 9:21.

Sabeti PC, Reich DE, Higgins JM, Levine HZ, 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, 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.

Sambrook J. 1989. Molecular cloning : a laboratory manual NY: Cold Spring Harbor, N.Y. : Cold Spring Harbor Laboratory Press, 1989.

Sanchez JJ, Phillips C, Børsting C, Balogh K, Bogus M, Fondevila M, Harrison CD, Musgrave-Brown E, Salas A, Syndercombe-Court D, et al. 2006. A multiplex assay with 52 single nucleotide polymorphisms for human identification. ELECTROPHORESIS 27:1713-1724.

Scheinfeldt LB, Soi S, Tishkoff SA. 2010. Colloquium paper: working toward a synthesis of archaeological, linguistic, and genetic data for inferring African population history. Proc Natl Acad Sci U S A 107 Suppl 2:8931-8938.

Scheinfeldt LB, Tishkoff SA. 2013. Recent human adaptation: genomic approaches, interpretation and insights. Nat Rev Genet 14:692-702.

Schlebusch CM, Lombard M, Soodyall H. 2013. MtDNA control region variation affirms diversity and deep sub-structure in populations from southern Africa. BMC Evol Biol 13:56.

Schlebusch CM, Skoglund P, Sjodin P, Gattepaille LM, Hernandez D, Jay F, Li S, De Jongh M, Singleton A, Blum MG, et al. 2012. Genomic variation in seven Khoe-San groups reveals adaptation and complex African history. Science 338:374-379.

Schoeninger MJ. 2001. Composition of tubers used by Hadza foragers of Tanzania. J Food Comp Anal 14:15–25.

Segurel 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.

Shabalina SA, Ogurtsov AY, Spiridonov AN, Novichkov PS, Spiridonov NA, Koonin EV. 2010. Distinct patterns of expression and evolution of intronless and intron-containing mammalian genes. Mol Biol Evol 27:1745-1749.

Shah AS, Ben-Shahar Y, Moninger TO, Kline JN, Welsh MJ. 2009. Motile cilia of human airway epithelia are chemosensory. Science 325:1131-1134.

Sharma K, Kaur GK. 2014. PTC bitter taste genetic polymorphism, food choices, physical growth in body height and body fat related traits among adolescent girls from Kangra Valley, Himachal Pradesh (India). Annals of Human Biology 41:29-39.

Shaw GM, Lu W, Zhu H, Yang W, Briggs FB, Carmichael SL, Barcellos LF, Lammer EJ, Finnell RH. 2009. 118 SNPs of folate-related genes and risks of spina bifida and conotruncal heart defects. BMC Med Genet 10:49.

Shi H, Dong YL, Wen B, Xiao CJ, Underhill PA, Shen PD, Chakraborty R, Jin L, Su B. 2005. Y-chromosome evidence of southern origin of the East Asian-specific haplogroup O3-M122. Am J Hum Genet 77:408-419.

Shi H, Zhong H, Peng Y, Dong YL, Qi XB, Zhang F, Liu LF, Tan SJ, Ma RZ, Xiao CJ, et al. 2008. Y chromosome evidence of earliest modern human settlement in East Asia and multiple origins of Tibetan and Japanese populations. BMC Biol 6:45.

Shigemura N, Shirosaki S, Sanematsu K, Yoshida R, Ninomiya Y. 2009. Genetic and molecular basis of individual differences in human umami taste perception. PLoS One 4:e6717.

Shoup JA. 2011. Ethnic Groups of Africa and the Middle East: An Encyclopedia: ABC-CLIO.

Shriver MD, Kennedy GC, Parra EJ, Lawson HA, Sonpar V, Huang J, Akey JM, Jones KW. 2004. The genomic distribution of population substructure in four populations using 8,525 autosomal SNPs. Hum Genomics 1:274-286.

Sias B. 2004. Human Pancreatic Lipase-Related Protein 2 Is a Galactolipase. Biochemistry 43:10138-10148.

Sim E, Lack N, Wang CJ, Long H, Westwood I, Fullam E, Kawamura A. 2008. Arylamine N-acetyltransferases: structural and functional implications of polymorphisms. Toxicology 254:170-183.

Skelton MM, Kampira EE, Wonkam AA, Mhandire KK, Kumwenda JJ, Duri KK, Dandara CC. 2014. Frequency Variation Among Sub-Saharan Populations in Virus Restriction Gene, BST-2 Proximal Promoter Polymorphisms: Implications for HIV-1 Prevalence Differences Among African Countries. OMICS.

Slatkin M. 2008. Linkage disequilibrium--understanding the evolutionary past and mapping the medical future. Nat Rev Genet 9:477-485.

Smith JM, Haigh J. 1974. The hitch-hiking effect of a favourable gene. Genet Res 23:23-35.

Soares I, Moleirinho A, Oliveira GN, Amorim A. 2015. DivStat: a user-friendly tool for single nucleotide polymorphism analysis of genomic diversity. PLoS One 10:e0119851.

Soranzo N, Bufe B, Sabeti PC, Wilson JF, Weale ME, Marguerie R, Meyerhof W, Goldstein DB. 2005. Positive selection on a high-sensitivity allele of the human bitter-taste receptor TAS2R16. Curr Biol 15:1257-1265.

SPSS. 2013. BM SPSS Statistics for Windows, Version 22.0. Armonk, NY: IBM Corp: IBM Corp.

Stanyon R, Sazzini M, Luiselli D. 2009. Timing the first human migration into eastern Asia. J Biol 8:18.

Su B, Jin L, Underhill P, Martinson J, Saha N, McGarvey ST, Shriver MD, Chu J, Oefner P, Chakraborty R, et al. 2000. Polynesian origins: insights from the Y chromosome. Proc Natl Acad Sci U S A 97:8225-8228.

Su B, Xiao J, Underhill P, Deka R, Zhang W, Akey J, Huang W, Shen D, Lu D, Luo J, et al. 1999. Y-Chromosome evidence for a northward migration of modern humans into Eastern Asia during the last Ice Age. Am J Hum Genet 65:1718-1724.

Tajima F. 1983. Evolutionary relationship of DNA sequences in finite populations. Genetics 105:437-460.

Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123:585-595.

Talbot J, Magno LA, Santana CV, Sousa SM, Melo PR, Correa RX, Di Pietro G, Rios-Santos F. 2010. Interethnic diversity of NAT2 polymorphisms in Brazilian admixed populations. BMC Genet 11:87.

Temussi PA. 2009. Sweet, bitter and umami receptors: a complex relationship. Trends Biochem Sci 34:296-302.

Thomas MG, Weale ME, Jones AL, Richards M, Smith A, Redhead N, Torroni A, Scozzari R, Gratrix F, Tarekegn A, et al. 2002. Founding mothers of Jewish communities: geographically separated Jewish groups were independently founded by very few female ancestors. Am J Hum Genet 70:1411-1420.

Thompson EE, Kuttab-Boulos H, Witonsky D, Yang L, Roe BA, Di Rienzo A. 2004. CYP3A variation and the evolution of salt-sensitivity variants. Am J Hum Genet 75:1059-1069.

Thompson EE, Kuttab-Boulos H, Yang L, Roe BA, Di Rienzo A. 2006. Sequence diversity and haplotype structure at the human CYP3A cluster. Pharmacogenomics J 6:105-114.

Tishkoff SA, Dietzsch E, Speed W, Pakstis AJ, Kidd JR, Cheung K, Bonne-Tamir B, Santachiara-Benerecetti AS, Moral P, Krings M. 1996. Global patterns of linkage disequilibrium at the CD4 locus and modern human origins. Science 271:1380-1387.

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, Verrelli BC. 2003. Patterns of human genetic diversity: implications for human evolutionary history and disease. Annu Rev Genomics Hum Genet 4:293-340.

Tishkoff SA, Williams SM. 2002. Genetic analysis of African populations: human evolution and complex disease. Nat Rev Genet 3:611-621.

Toda Y, Nakagita T, Hayakawa T, Okada S, Narukawa M, Imai H, Ishimaru Y, Misaka T. 2013. Two Distinct Determinants of Ligand Specificity in T1R1/T1R3 (the Umami Taste Receptor). J Biol Chem 288:36863-36877.

Valente C, Alvarez L, Marks SJ, Lopez-Parra AM, Parson W, Oosthuizen O, Oosthuizen E, Amorim A, Capelli C, Arroyo-Pardo E, et al. 2015. Exploring the relationship between lifestyles, diets and genetic adaptations in humans. BMC Genet.

Vallone PM, Butler JM. 2004. AutoDimer: a screening tool for primer-dimer and hairpin structures. Biotechniques 37:226-231.

Ventura AK, Worobey J. 2013. Early influences on the development of food preferences. Curr Biol 23:R401-408.

Vigilant L, Stoneking M, Harpending H, Hawkes K, Wilson AC. 1991. African populations and the evolution of human mitochondrial DNA. Science 253:1503-1507.

Vitalis R, Gautier M, Dawson KJ, Beaumont MA. 2014. Detecting and measuring selection from gene frequency data. Genetics 196:799-817.

Vitti JJ, Grossman SR, Sabeti PC. 2013. Detecting natural selection in genomic data. Annu Rev Genet 47:97-120.

Voight BF, Adams AM, Frisse LA, Qian Y, Hudson RR, Di Rienzo A. 2005. Interrogating multiple aspects of variation in a full resequencing data set to infer human population size changes. Proc Natl Acad Sci U S A 102:18508-18513.

Voight BF, Kudaravalli S, Wen X, Pritchard JK. 2006. A map of recent positive selection in the human genome. PLoS Biol 4:e72.

Watterson GA. 1975. On the number of segregating sites in genetical models without recombination. Theor Popul Biol 7:256-276.

Weale ME, Yepiskoposyan L, Jager RF, Hovhannisyan N, Khudoyan A, Burbage-Hall O, Bradman N, Thomas MG. 2001. Armenian Y chromosome haplotypes reveal strong regional structure within a single ethno-national group. Hum Genet 109:659-674.

Welcome MO, Mastorakis NE, Pereverzev VA. 2015. Sweet taste receptor signaling network: possible implication for cognitive functioning. Neurol Res Int 2015:606479.

Wen B, Li H, Gao S, Mao X, Gao Y, Li F, Zhang F, He Y, Dong Y, Zhang Y, et al. 2005. Genetic structure of Hmong-Mien speaking populations in East Asia as revealed by mtDNA lineages. Mol Biol Evol 22:725-734.

Wen B, Xie X, Gao S, Li H, Shi H, Song X, Qian T, Xiao C, Jin J, Su B, et al. 2004. Analyses of genetic structure of Tibeto-Burman populations reveals sex-biased admixture in southern Tibeto-Burmans. Am J Hum Genet 74:856-865.

Williamson SH, Hernandez R, Fledel-Alon A, Zhu L, Nielsen R, Bustamante CD. 2005. Simultaneous inference of selection and population growth from patterns of variation in the human genome. Proc Natl Acad Sci U S A 102:7882-7887.

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.

Wollstein A, Stephan W. 2015. Inferring positive selection in humans from genomic data. Investig Genet 6:5.

Wooding S. 2005. Evolution: a study in bad taste? Curr Biol 15:R805-807.

Wooding S, Kim UK, Bamshad MJ, Larsen J, Jorde LB, Drayna D. 2004. Natural selection and molecular evolution in PTC, a bitter-taste receptor gene. Am J Hum Genet 74:637-646.

Xiao X, Mukherjee A, Ross LE, Lowe ME. 2011. Pancreatic Lipase-related Protein-2 (PLRP2) Can Contribute to Dietary Fat Digestion in Human Newborns. Journal of Biological Chemistry 286:26353-26363.

Yao YG, Kong QP, Bandelt HJ, Kivisild T, Zhang YP. 2002. Phylogeographic differentiation of mitochondrial DNA in Han Chinese. Am J Hum Genet 70:635-651.

Yao YG, Watkins WS, Zhang YP. 2000. Evolutionary history of the mtDNA 9-bp deletion in Chinese populations and its relevance to the peopling of east and southeast Asia. Hum Genet 107:504-512.

Yarmolinsky DA, Zuker CS, Ryba NJ. 2009. Common sense about taste: from mammals to insects. Cell 139:234-244.

Ye K, Gu Z. 2011. Recent advances in understanding the role of nutrition in human genome evolution. Adv Nutr 2:486-496.

Young JH, Chang YP, Kim JD, Chretien JP, Klag MJ, Levine MA, Ruff CB, Wang NY, Chakravarti A. 2005. Differential susceptibility to hypertension is due to selection during the out-of-Africa expansion. PLoS Genet 1:e82.

Yu C, Deng M, Zheng L, He RL, Yang J, Yau SS. 2014. DFA7, a new method to distinguish between intron-containing and intronless genes. PLoS One 9:e101363.

Zeng K, Fu YX, Shi S, Wu CI. 2006. Statistical tests for detecting positive selection by utilizing high-frequency variants. Genetics 174:1431-1439.

Zhang L, Miyaki K, Wang W, Muramatsu M. 2010. CYP3A5 polymorphism and sensitivity of blood pressure to dietary salt in Japanese men. J Hum Hypertens 24:345-350.